CONSEQUENCES OF BIOTECH INNOVATION IN CROP SEED

A dissertation presented

by

Irena Vasileva Tsvetkova

to
The Department of Economics

In partial fulfillment of the requirements for the degree of
Doctor of Philosophy

in the field of
Economics

Northeastern University
Boston, Massachusetts
April, 2014
CONSEQUENCES OF BIOTECH INNOVATION IN CROP SEED

by

Irena Vasileva Tsvetkova

ABSTRACT OF DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics in the College of Social Sciences and Humanities of Northeastern University
April, 2014
Abstract

My dissertation analyses the effect of technology on consumers and producers in the U.S. cottonseed industry. The industry has experienced a sharp innovation spur since the commercialization of genetically-modified seed in 1996. Product differentiation has increased many-fold due to these new technologies giving rise to concerns about market power and to a much larger menu of choices for farmers. The biotechnology and seed-producer sectors have adapted to these changes by engaging in rapid consolidation (Fuglie et al., 2011). In the meantime, farmers have almost fully adopted genetically-modified seed in the 17 years since their introduction.

The first chapter estimates a structural discrete-choice demand model and analyzes product substitution patterns. I deal with the problem of dimensionality with the aid of a great dataset and by using a demand system that reduces the dimensionality problem yet does not restrict substitution patterns. I use several sources of exogenous variation to obtain consistent estimates. I find that genetically-modified seed are preferred to conventional seed and that preferences vary by individual.

The second chapter looks at the economic effects from new product introductions. In the context of industry transformations of the past 17 years biotech companies have been accused of using innovation to reap surplus from farmers. I address this concern by looking at one of the more controversial genetic innovations in cottonseed. Using demand estimates from the first
chapter and under assumptions about the nature of competition, I simulate prices that would clear the market without the innovation. This allows me to calculate changes in farmer surplus and in seed firm variable profits. I find that both farmers and seed firms benefited from this biotech innovation in cottonseed.

The third chapter measures the effect of technology and market factors on product turnover rate. While the introduction of new production technologies has definitely increased the number of products offered, the isolated effect of industry consolidation on product availability is unclear. I test five hypotheses on product turnover and limit my analysis to products introduced and dropped out of the market within the observed period to estimate the effects of technology and concentration on survival. I find that genetically-modified seed survives longer than conventional seed and that products sold by bigger vertically-integrated seed firms have a longer product life. This could negatively affect the investment decisions of smaller local seed firms and decrease the availability of seed specifically adapted to meet local climate and soil conditions.
Acknowledgements

I would like to express my gratitude to Professor James Dana, my research supervisor, for his enthusiastic encouragement and useful critiques. I would also like to thank Professor John Kwoka, for helping me keep the practical focus of my work, and Dr.Gustavo Vicentini, who suggested “the minivan paper” and was always available to provide feedback. Special thanks also go to Professor Kyle Stiegert and the University of Wisconsin - Madison for enabling me to visit their campus and for supporting a collaborative effort with their data. My thanks are also extended to Assistant Professor Tiago Pires from the University of North Carolina and to Dr.Vardges Hovhannisyan for support in implementing the mixed-logit random-coefficients demand estimator.

Finally, I wish to thank my parents and friends for their support and encouragement throughout my study.
# TABLE OF CONTENTS

Abstract.................................................................................................................................3

Acknowledgements..............................................................................................................5

Table of Contents................................................................................................................6

Background .............................................................................................................................8

Chapter 1. Mixed-logit Random-coefficients Demand for Cottonseed in the U.S. .................9
  1.1. Introduction ...................................................................................................................9
  1.2. Demand Model ..........................................................................................................12
  1.3. Data ............................................................................................................................21
  1.4. Results .......................................................................................................................23
  1.5. Conclusion ..................................................................................................................32

Chapter 2: The Effects of Biotech Innovation in the Market for Crop Seed .........................34
  2.1. Introduction ................................................................................................................34
  2.2. Supply Model, Producer Profits, and Consumer Surplus ..........................................37
  2.3. Data ............................................................................................................................40
  2.4. Results .......................................................................................................................41
  2.5. Conclusion ..................................................................................................................44

  3.1. Introduction ................................................................................................................46
  3.2. Survival Model ..........................................................................................................49
  3.3. Data ............................................................................................................................51
  3.4. Results .......................................................................................................................54
  3.5. Conclusion ..................................................................................................................58
Appendix .....................................................................................................................................................60

References..................................................................................................................................................62
Background

Scientists and farmers had been altering plant DNA through traditional hybridization methods as far back as the 1920s, but the first genetically-modified (GM) plants produced in a lab first came out in 1982. Genetically-engineered seed was not available for sale until 1996, which renders the industry very young and particularly interesting for study. GM seed are characterized by increasing product differentiation and are available for purchase through multiple vertical channels. In the United States there are hundreds of GM varieties for corn, cotton, and soybeans. All of these varieties can be categorized in three main groups based on their genetic characteristics: (1) herbicide-tolerant (HT) plants have a new gene, which makes them resistant to chemicals sprayed by farmers; (2) insect-resistant (IR) plants stand stronger against pest infestation; and finally, (3) stacked seed are both resistant to insects and tolerant to herbicides. In addition, seed companies further differentiate their seed by adapting them to better fit local weather and soil conditions.
Chapter 1. Mixed-logit Random-coefficients Demand for Cottonseed in the U.S.

1.1. Introduction

Genetically-modified seed is characterized by high prices and aggressive product introductions. In spite of the fact that prices are between two and six times more expensive than conventional seed, farmers have almost fully adopted the new technology. Figure 1.1 shows that the adoption of biotech seed for the three main GM crops reached about 90% in 2011. With this high uptake it becomes increasingly more important to understand demand for biotech seed. Knowing how much farmers value different seed characteristics helps analyze substitution patterns, which is useful to both private stakeholders and regulators. Shapiro (1996) describes how cross-price elasticity is used in constructing diversion ratios and evaluating unilateral merger effects in mergers with differentiated products (e.g. the merger between Interstate Bakeries Corporation and Continental Baking Company; also, the acquisition by Kimberly-Clark of the Scott Paper Company). Demand estimation is useful when addressing a variety of issues including measuring market power (Nevo, 2001), analyzing horizontal mergers (Nevo, 2000), evaluating welfare gains due to new products (Petrin, 2002), and construction of price indices accounting for quality changes and product introductions (Nevo, 2003).

In this chapter I estimate farmers’ preference parameters for cottonseed and calculate product substitution patterns. I do this using several sources of exogenous variation and an estimator
introduced by Berry, Levinsohn, and Pakes (1995) (henceforth BLP) to obtain consistent estimates.

Biotech seed like most real-world markets involve product differentiation with similar but not identical products. The traditional way to estimate demand in this context would be to specify a system of demand equations – an equation for each product’s demand as a function of its own price, the prices of other products, and other variables. However, estimating such systems poses the so called dimensionality problem, which means that with $J$ products, the econometrician has to estimate at least $J \times J$ parameters.

Figure 1.1  Adoption of Biotech Seed: Percent Acreage Planted with GM Seed (2000 – 2011)

Source: USDA NASS
One way to bypass the dimensionality problem is to use a discrete-choice demand model, which categorizes products based on their characteristics. The number of estimated coefficients is no longer $J \times J$, but becomes the size of the characteristics space, which is much more manageable. The logit demand model was first put forward by McFadden (1974), but while successful in dealing with the dimensionality problem, it produces unrealistic substitution patterns (Berry, 1994). Extensions of the basic logit model, which produce more realistic substitution patterns, include the nested logit model, the principles of differentiation generalized extreme value (PD GEV) model, and the mixed-logit model, all of which can be estimated using both market and individual level data. The mixed-logit random-coefficients model provides the most flexible and realistic substitution patterns, but is more computationally cumbersome.

In the agricultural literature discrete-choice demand models have been used to describe consumer choice with only two options. Gracia et al. (2008) estimate the uptake of organic versus non-organic foods using a bivariate probit model. Other papers model both bivariate choice and store choice sequentially, using an ordered probit model (Huang, 1996; Thompson et al., 1998; Verhoef, 2005; Kuhar et al., 2005).

With the release of the farmer survey on biotech seed use in the U.S. by market research firm dmrkynetec (DMR), multi-product discrete choice demand estimation for GM seed became possible. In some yet unpublished work Ma (forthcoming) estimates BLP demand for corn.
Zhang (forthcoming) also estimates random-coefficients discrete-choice demand for corn allowing farmers to choose more than one products. In this chapter I estimate BLP demand for cottonseed similar to Xingliang Ma’s study of corn. I find that farmers prefer herbicide-tolerant and stacked seed to insect-resistant and conventional seed. The results presented here show that taste parameters vary by individual and that farmers are less price-sensitive towards cottonseed relative to corn seed. This resonates with the fact that cottonseed is characterized by less product differentiation and is produced by way fewer firms than corn seed, which makes it less substitutable.

The rest of the chapter proceeds as follows. Section 1.2 outlines the model, discusses some econometrical concerns, introduces the GMM estimator, and summarizes the optimization design. Section 1.3 describes the data. Section 1.4 presents the results, and Section 1.5 concludes.

1.2. Demand Model

I estimate a slightly modified version of the BLP mixed-logit random-coefficients demand model. Farmers are the consumers of biotech seed and seed firms are the producers. Farmer $i$'s utility from product $j$ in market $m$ is:

$$U_{ijm} = x_{jm} \beta_i - \alpha_i p_{jm} + \xi_{jm} + \epsilon_{ijm} = V_{ijm} + \epsilon_{ijm},$$

$$i = 1, \ldots, I, \ j = 1, \ldots, J, \ m = 1, \ldots, M$$
where \( x_{jm} \) are observed product attributes varying with market \( m \), while \( \xi_{jm} \) are unobserved product attributes.\(^1\) The term \( \epsilon_{ijm} \) is an error component that varies by consumer, product, and market. Preference parameters are allowed to vary by individual in each market:

\[
(\alpha_{im}, \beta_{im}) = (\alpha, \beta) + PD_{im} + \Sigma v_{im} \quad v_{im} \sim N(0,1).
\]

Consumers are described through observed demographic characteristics, \( D_{im} \), and unobserved idiosyncratic taste preferences, \( v_{i} \), which are assumed to have a normal distribution. \( \Pi \) and \( \Sigma \) are the random coefficients. Combining (1.1) and (1.2) utility becomes

\[
u_{ijm} = (x_{jm}\beta - \alpha p_{jm} + \xi_{jm}) + [-p_{jm}, x_{jm}](\Pi D_{im} + \Sigma v_{im}) + \epsilon_{ijm}
\]

\[
\delta_{jm} \quad \theta_1 = \{\alpha, \beta\} \quad \mu_{jm}
\]

\[
\theta_2 = \{\Pi, \Sigma\}
\]

where \( \alpha \) and \( \beta \) are the mean valuations for the product attributes – these preference parameters are common to all consumers and are jointly referred to as \( \theta_1 \). Notice that \( \Pi \) and \( \Sigma \) capture the effect on utility from the interaction of product attributes and consumer characteristics. The mean utility common to all consumers is \( \delta_{jm} \), and the term \( \mu_{ijm} + \epsilon_{ijm} \) captures individual deviations from the mean utility. Cross-price substitution will be driven by this random shock \( (\mu_{ijm} + \epsilon_{ijm})^2 \).

---

\(^1\) Farmers and seed firms, however, observe \( \xi_{jm} \), which gives rise to endogeneity concerns.

\(^2\) The outside good terms \( \xi_{om}, \pi_0, \text{ and } \sigma_0 \) are not separately identified, and therefore, it is standard practice to set them equal to zero.
When the utility from product $j$ exceeds the utility from all other products in the market, a consumer would purchase product $j$ and thus increase its market share. Each consumer is characterized by a set of characteristics $v_i, D_i$, and $\epsilon_{ij}$, and all consumers who prefer product $j$ in market $m$ are grouped in set $A_{jm}$:

$$A_{jm}(X_m, P_m, \delta_m; \theta_2) = \{(D_i, v_i, \epsilon_{i0m}, \ldots, \epsilon_{ijm})|u_{ijm} \geq u_{ilm} \text{ for } \forall l = 1, \ldots, J\}.$$ (1.4)

Product market shares for product $j$ in market $m$ are simulated as the probability that consumer $i$ purchases product $j$ integrated over all consumers in set $A_{jm}$

$$s_{jm}(X_m, P_m, \delta_m; \theta_2) = \int_{A_{jm}} dP^*(D, v, \epsilon) = \int_{A_{jm}} dP_{\epsilon}^*(\epsilon) dP_{\nu}^*(\nu) dP_{D}^*(D),$$ (1.5)

where $P_{\epsilon}^*(\epsilon)$, $P_{\nu}^*(\nu)$ and $P_{D}^*(D)$ are the distributions of $\epsilon, \nu$ and $D$. The mixed-logit model uses an empirical distribution for both $\nu$ and $D$ and a parametric i.i.d Type I extreme value distribution for $\epsilon^3$. Finally, preference parameters are estimated by looking for the values that equate simulated market shares $s_{jm}$ with observed market shares $S_{jm}$:

$$s_{jm}(X_m, P_m, \delta_m; \theta_2) = S_{jm}$$ (1.6)

The logit demand model is a special case of the mixed-logit model presented so far. If consumer heterogeneities are assumed to enter the model only through $\epsilon_{ijm}$, then $\theta_2$ becomes zero.

---

3 Market shares depend only on differences in utilities, hence, the estimation algorithm ends up subtracting the utility of the outside good from the utility of the inside good and estimating a model where the outside alternative is “normalized” to zero. This implies that there is a random coefficient on the constant term in the utility function of the inside good (BLP, p.849).
resulting model is easy to estimate, but has very restrictive substitution patterns. Market shares have the logit distribution

\[ s_{jm} = \frac{e^{\delta_{jm}}}{\sum_{k=0}^{l} e^{\delta_{km}}}, \]  

(1.7)

and the demand parameters can be estimated using linear instrumental variable techniques from:

\[ \ln(s_{jm}) - \ln(s_{0m}) = \delta_{j} = X_{jm}\beta - \alpha P_{jm} + \xi_{jm}. \]  

(1.8)

In spite of its limitations, the logit model is a good tool to get a general feel for the data and I present some logit results below.

**Econometric Concerns**

Product quality is unobserved by the econometrician and therefore price is likely to be endogenous. This is due both to simultaneous determination and to omitted variables. On the one hand, producers know the values of the unobserved characteristics and set product price accordingly. In addition, consumers also know the values of unobserved quality when making their purchasing decisions, which means that there is an omitted variables problem. The direction of bias can often be determined logically. For example if desirable unobserved attributes are positively correlated with price, the estimation without correcting for endogeneity will result in a price coefficient that is smaller in magnitude. Since higher prices are

\[ \text{An additional source of endogeneity is the fact that different consumers face different prices for the same product. In this case using average transaction prices will lead to measurement error bias, which is a third reason why prices might be correlated with the error term (Nevo, 2000, p.518).} \]
associated with desirable attributes, consumers would dislike prices less than if the higher prices occurred after controlling for the unobserved desirable attributes. The size of the difference in coefficients indicates the importance of accounting for endogeneity (Train, 2009, p.316, 345).

Obtaining consistent estimates is contingent on correcting for endogeneity. Three methods for dealing with endogeneity are popular in discrete choice modeling. One of them is the solution proposed by BLP, another one is called the control function approach, and the third solution is maximum likelihood (a simultaneous version of the control function approach). The way BLP deal with endogeneity is by including product-market fixed effects to pick up unobserved product attributes and then implementing a contraction mapping along with linear IV estimation\(^5\) to recover average and consumer-specific parameters. BLP use a GMM estimator. They propose the estimation of the following moment conditions:

\[
E[\xi_{jm} | Z_{jm}, X_{jm}] = 0, 
\]

along with the non-linear search in equation (1.6). \(Z_{jm}\) are price instruments and \(\xi_{jm}\) and \(X_{jm}\) are defined as in Section 1.2.

The control function (CF) approach controls for unobserved product characteristics in a computationally simpler way. The approach involves two stages. First, the endogenous explanatory variable is regressed on all exogenous variables and instruments. Then, the

---

\(^5\) Berry(1994) notes that the IV procedure can deal with measurement error only if the variable measured with error only enters the mean utility term common to all consumers (Nevo, 2000, p.518).
residuals measuring unobserved factors are recovered from the estimation, and a function of these residuals (the control function) is included in the choice model and estimated using maximum likelihood (Petrin et al., 2010; Train, 2009; chapter 13). The problem with the control function approach is that it needs a correct and complete specification of the control function, which is not easy to justify given the non-linear nature of choice models. For this reason, the CF solution to endogeneity is not very popular in the field of economics, although widely used in marketing (Petrin et al., 2010).

I correct for endogeneity bias using a modified version of the BLP solution to endogeneity – the Generalized Method of Simulated Moments, i.e. the GMM (or GMSM) estimator. BLP include product fixed effects to capture any unobservables in the error term $\xi_j$. I do not include product-specific shocks as the number of products in my analysis is very large relative to the number of markets. This has a number of implications. First, in the absence of product shocks I need to be more careful when choosing instruments because they need to take into account brand specificities. I need to make sure that my instruments satisfy the orthogonality conditions. Second, I do not need to use the minimum-distance procedure implemented by BLP to recover taste parameters. This procedure is only necessary in the presence of product dummies, which absorb the preference parameters. In this sense, the model used here is slightly different from the BLP model.

---

6 This can be done with standard software packages like Stata and SAS.

7 Here I follow the recommendation in Nevo (2000) on what he calls brand dummies. I tried including product fixed effect in the estimation, but the non-linear search fails to run at all.
A common problem when inverting the simulated market shares obtained from (1.5) to obtain the preference parameters in (1.6) is that some products have zero market shares in spite of the fact that they are offered to consumers. The estimation of the BLP model proceeds by ignoring those observations for which market shares are zero. Gandhi et al. (2014) show that this introduces bias in the estimation, which leads to underestimating the demand coefficients, which in turn overestimates markups. This is a limitation to keep in mind when interpreting the results.

The GMM Estimator

The GMM estimates are obtained through a nested iterated procedure called the Nested Fixed Point (NFP) algorithm. The algorithm is comprised of an inner loop and an outer loop.\(^8\) For a given \(\theta\) the inner loop simulates product market shares using the smooth simulator and 20 individual draws from each market.\(^9\) Then, it uses the simulated shares to obtain mean utilities that minimize the distance between simulated and observed market shares. These are the moments described by equation (1.6). Then, the structural error term is calculated from the mean utilities as

\[
\omega(\theta) = \delta - x_t \theta_1. \tag{1.10}
\]

\(^8\) Dube et al. (2011) proposes an alternative to the NFP algorithm, which has the same mathematical properties, but is much faster. He suggests performing constrained optimization with the market share equations as constraints to the objective function instead of as an inner loop to unconstrained minimization.

\(^9\) The individual draws are from a standard normal distribution. They are made once and kept constant during the estimation.
The outer loop looks for the values of $\theta$ that minimize the moment conditions in (1.9) with a weight matrix $\Phi^{-1}$. The objective function is

$$(1.11) \quad \min_{\theta} \omega(\theta)'Z\Phi^{-1}Z'\omega(\theta).$$

More precisely, the non-linear search is performed on the first order condition of the objective function, which has the desirable property that it is only non-linear in $\theta_2$ but linear in $\theta_1$. Once an estimate of $\theta$ is obtained, the whole process is repeated until certain convergence criteria are met.

**Optimization Design**

Estimation of the NFP algorithm requires specification of starting values, algorithm, and stopping criteria. The starting values for the mean utility in the contraction mapping are the fitted values from logit demand estimation. The starting values for the $\theta_2$ vector are obtained by first estimating two parameters at a time, and then using these estimates as starting values to estimate all elements of $\theta_2$.\(^{10}\)

Built-in MATLAB algorithms from two different algorithm classes are tested. One of them is the derivative-based Quasi-Newton algorithm and the other is the deterministic direct search Simplex algorithm. The difference is that derivative-based algorithms use information about the steepness and curvature of the objective function, while direct search algorithms are based

---

\(^{10}\) This methodology is recommended by Knittel et al. (forthcoming) as well as by Aviv Nevo in personal communication. I use 0.01 as a starting value when breaking the problem into smaller parts.
on function evaluations. In theory, the minimum occurs when the gradient vector is zero, but in practice the calculated gradient vector is never exactly zero. Therefore, different stopping criteria are utilized to specify how close to the minimum the algorithm has to be in order to stop the iterative process. Convergence is often based on the change in the objective function and the parameter vector between two consecutive iterations of an algorithm. Alternative stopping rules are a maximum number of iterations or function evaluations.

A number of stopping rules are utilized. I use a tolerance of $10^{-06}$ for changes in the objective function and the parameter vector. Nevo (2000) uses 0.1, while Knittel et al. (forthcoming) experiment with tolerance levels of $10^{-03}$ and $10^{-06}$.

A stopping criterion is also required for the inner loop contraction mapping. I use a tight tolerance following Knittel et al. (forthcoming). The tight inner loop stopping criterion is fixed at $10^{-16}$ imposing an upper bound of 5,000 iterations. In contrast, the loose tolerance is initially set to $10^{-06}$, and once the parameter estimates, $\theta_2$, are within 0.01 in subsequent iterations, the tolerance is tightened to $10^{-09}$. This is the approach used in Nevo (2000).

*Instruments*

I use three selection criteria when looking for instrumental variables. In order to ensure that my instruments are relevant, I use a first-stage F-test for excluded instruments. If the F-statistic is greater than 10 (Staiger et al., 1997), my instruments are relevant. I test for instrument
exogeneity using the Hansen J test,\textsuperscript{11} and finally, I use intuition to confirm that the candidate instruments are not directly correlated with the dependent variable.

A large menu of instruments for price was considered as suggested by the literature. Following BLP I use the exogenous characteristics of product $j$, the sum of the exogenous characteristic for the same firm’s products excluding product $j$, and the sum of the exogenous characteristic of rival firms’ products in the same market to instrument for price.\textsuperscript{12} The intuition is that products that face good substitutes will usually have low markups, whereas products without good substitutes will have higher markups and prices. In addition, because markups respond differently to own and others’ products, the instruments will distinguish between the characteristics of products produced by the same firm versus the characteristics of products produced by rival firms (BLP, p.855, 861).

1.3. Data

Two main data sources are used in this paper. One is proprietary and collected by the market research firm DMR. The second dataset is publicly available through the United States Department of Agriculture (USDA). DMR has undertaken an annual farmer survey on genetically-modified crop seed use in the United States since 2000. There are three DMR

\textsuperscript{11} The Hansen J test of over-identification is preferred over the Sargan test when using robust standard errors.

\textsuperscript{12} I also experimented with other candidate instruments including dummy variables for some GM brands, the number of similar products in a market, the Herfindahl-Hirschman Index (HHI) in a market excluding the share of the firm selling the product in question, and Hausman price instruments. All of these failed one or more of the tests for good instruments with the exception of the Hausman price instruments, which decreased the number of observations and hence were not able to produce significant coefficients.
datasets on GM seed purchases – one each for corn, cotton, and soybeans. I am using the cottonseed data for the years 2000 through 2007 (excluding 2001), which is available for the 12 states where cotton is a major crop. The 12 states can be subdivided into 24 crop reporting districts (CRD’s) and the survey contains 377 cottonseed varieties total. A product is defined at the variety level so there are 377 products. An observation is defined as a variety-region-year combination, and so 377 varieties, 12 states, and 7 years yield 2,488 observations. The DMR cottonseed dataset contains variation in transaction prices, sales, discounts, acres planted, farm size, modes of purchase, and variety characteristics including variety name, trait content, and brand.

The second dataset is the USDA National Agricultural Statistical Service’s (NASS) acreage information. I obtain the number of acres planted with all principal crops, including corn, cotton, soybeans, and 14 other crops in each market (a region-year combination). I use the acreage planted with all principal crops as the market size when calculating product shares for cottonseed varieties. The outside good is seed for all principal crops but cotton. As a result, inside good shares are very small (between 1.08E-06 and 0.193) whereas outside good shares are fairly large (between 0.620 and 0.997), which is standard. Table 1.1 shows these and other descriptive statistics. Most of the observations are for stacked cottonseed (40%), while only 5%...

---

13 The survey is not a panel since different farmers are surveyed each year. However, once aggregated to the product level, the dataset becomes a panel.
14 Since cottonseed varieties are introduced and exit over this period, the resulting panel is unbalanced.
15 The survey also provides projection factors, which project the information from the local markets to the national level and allow for aggregation.
fall under insect-resistant products. This is not surprising given the fact that stacked seed contain insect-resistant traits and in light of the evidence of subadditive bundle pricing presented in Shi et al. (2011). The demographics I am using pertain to the size of the farm as measured by the number of acres, which vary between 3 and 22,000 acres.

Table 1.1 also contains descriptive statistics on the price instruments. As pointed out earlier, the instruments I use are the exogenous characteristics of product $j$, the sum of the exogenous characteristic for the same firm’s products excluding product $j$, and the sum of the exogenous characteristic of rival firms’ products in the same market. I observe two exogenous characteristics – trait content (the dummies HT, IR, HT_IR, CONV) and product age. This gives me 15 instruments. I drop CONV and the sum of the ages for other firms’ products in the same market which are collinear, leaving me with 13 instruments for price.

1.4. Results

*Logit Demand Parameter Estimates*

The logit demand model produces unrealistic substitution pattern as mentioned earlier. However, due to its computational simplicity, I present some results here in order to get a feel for the data. This allows me to examine (1) the importance of instruments for price, and (2) the effects of different price instruments. Table 1.2 presents some of the logit demand results.

---

16 Subadditive bundle pricing means that the price of the stacked seed, which bundle different traits together, is less than the sum of the prices of components of the bundle (i.e. the single traits).

17 VIF test of collinearity. This test is not based on regression results but can be run independently.
Table 1.1. Descriptive Statistics for the Panel Dataset

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Market Shares:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Market Size</td>
<td>Thousands of acres planted with all principal crops per market</td>
<td>84</td>
<td>10,109.21</td>
<td>8549.193</td>
<td>674</td>
<td>24545</td>
</tr>
<tr>
<td>Cottonseed Share</td>
<td>Acre share of cottonseed varieties in a market</td>
<td>2448</td>
<td>0.004</td>
<td>0.011</td>
<td>1.08E-06</td>
<td>0.193</td>
</tr>
<tr>
<td>Outside Good Share</td>
<td>Acre share of other principal crops in a market</td>
<td>2448</td>
<td>0.854</td>
<td>0.078</td>
<td>0.620</td>
<td>0.997</td>
</tr>
<tr>
<td><strong>Product Attributes:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HT</td>
<td>Dummy for herbicide-tolerance</td>
<td>2448</td>
<td>0.311</td>
<td>0.463</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>IR</td>
<td>Dummy for insecticide-resistance</td>
<td>2448</td>
<td>0.046</td>
<td>0.209</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>HT_IR</td>
<td>Dummy for stacking</td>
<td>2448</td>
<td>0.401</td>
<td>0.490</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>CONV</td>
<td>Dummy for conventional seed</td>
<td>2448</td>
<td>0.243</td>
<td>0.429</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Price</td>
<td>Retail price minus discounts plus royalties in $/bag</td>
<td>2448</td>
<td>28.629</td>
<td>18.225</td>
<td>1.123</td>
<td>132.874</td>
</tr>
<tr>
<td><strong>Consumer Demographics:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Farm Size</td>
<td>Acres per Farm</td>
<td>13,997</td>
<td>1,026.616</td>
<td>1,369.597</td>
<td>3</td>
<td>22,000</td>
</tr>
<tr>
<td><strong>Instruments for Price:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exogenous Characteristics I</td>
<td>Sum of each trait dummy and age for the same firm's products in this market (5 instruments)</td>
<td>2448</td>
<td>28.89</td>
<td>30.84</td>
<td>0</td>
<td>134</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>2448</td>
<td>4.37</td>
<td>4.15</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>HT</td>
<td>2448</td>
<td>0.85</td>
<td>1.40</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>IR</td>
<td>2448</td>
<td>4.85</td>
<td>3.66</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>HT_IR</td>
<td>2448</td>
<td>3.33</td>
<td>4.25</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>CONV</td>
<td>2448</td>
<td>70.68</td>
<td>87.68</td>
<td>4</td>
<td>350</td>
</tr>
<tr>
<td></td>
<td>HT</td>
<td>2448</td>
<td>9.71</td>
<td>11.26</td>
<td>0</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>IR</td>
<td>2448</td>
<td>1.23</td>
<td>1.64</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>HT_IR</td>
<td>2448</td>
<td>10.66</td>
<td>11.63</td>
<td>0</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>CONV</td>
<td>2448</td>
<td>9.85</td>
<td>12.22</td>
<td>0</td>
<td>47</td>
</tr>
</tbody>
</table>

Note - There are 377 cottonseed varieties, 7 years, and 12 regions in my dataset. A product is a cottonseed variety and products change over time. A market is a region/year combination, and an observation is a product in a given market.
Table 1.2. Logit Demand Estimation Results

<table>
<thead>
<tr>
<th>Variables</th>
<th>OLS</th>
<th>2SLS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Price</td>
<td>-0.003 (-0.004)</td>
<td>-0.036*** (0.009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Benchmark is conventional)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HT</td>
<td>0.347** (0.128)</td>
<td>0.931*** (0.201)</td>
</tr>
<tr>
<td>IR</td>
<td>0.428 (0.358)</td>
<td>1.269*** (0.287)</td>
</tr>
<tr>
<td>HT_IR</td>
<td>0.686*** (0.166)</td>
<td>1.673*** (0.278)</td>
</tr>
<tr>
<td>Constant</td>
<td>-7.810*** (0.140)</td>
<td>-6.313*** (0.212)</td>
</tr>
<tr>
<td>Instruments</td>
<td>n/a</td>
<td>Exogenous</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Characteristics</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td>Region, Company</td>
<td>Region</td>
</tr>
<tr>
<td>Obs</td>
<td>2,448</td>
<td>2,448</td>
</tr>
<tr>
<td>R²</td>
<td>0.059</td>
<td>0.131</td>
</tr>
</tbody>
</table>

Instrument Tests:
- First-st F: 18.14 | 402.2 | 84.77 | 11.16 | 179.0 | 37.62
- Hansens J: 56.60 | n/a | 46.50 | 13.85 | n/a | 14.65
- Hansens J(p): 1.59e-08 | n/a | 1.16e-06 | 0.180 | n/a | 0.145

Implied Elasticities
- Mean: -0.0716 | -1.035 | -0.120 | -0.224 | -0.867 | 0.107 | -0.001
- Standard Deviation: 0.0455 | 0.658 | 0.0765 | 0.142 | 0.552 | 0.0678 | 0.001

Robust and clustered standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1
The dependent variable is the difference in the logs of the shares of the inside good and the outside good as seen in equation (7). All specifications include geographic market fixed effects, and columns (5) through (7) also include firm fixed effects. Column (1) shows the OLS results without controlling for price endogeneity. The price coefficient is insignificant and very small. As discussed in Section 4 price endogeneity is partially a result of an omitted variables problem – the positive attribute quality is not observed and becomes part of the error term. For this reason, the coefficient on price is biased upwards in the OLS regression as people dislike price less without controlling for quality.

The instruments in specifications (2) through (7) are all relevant as reflected by the F-statistic which is above 10. Columns (2) through (4) compare the candidate instruments without controlling for firm fixed effects. Columns (5) through (7) include firm dummies as controls. We observe that the Hansen J p-statistic is higher when only functions of the exogenous characteristics are used as instruments. This is desirable because a higher p-statistic means we cannot reject the null hypothesis of the Hansen J test (i.e. instrument exogeneity). The preferred specification is the one in column (5) with functions of the exogenous product characteristics as price instruments and geographic market and firm fixed effects as covariates. The Hansen J p-value is 0.180, which means that we cannot reject instrument exogeneity.

Full Model Demand-side Parameter Estimates

The estimates of the full model are based on equation (1.3) and predicted market shares are computed using equation (1.5). The latter are based on the empirical distribution of
demographics (farm size as sampled from the DMR survey), independent normal distributions (for \( \nu \)), and a Type I extreme value distribution (for \( \varepsilon \)). The instruments include exogenous characteristics of product \( j \), the sum of the exogenous characteristic for the same firm’s products excluding product \( j \), and the sum of the exogenous characteristic of rival firms’ products in the same market. The results from the full model are presented in Table 1.3.

<table>
<thead>
<tr>
<th>Mean Interaction with Consumer Unobservables (Standard Deviation)</th>
<th>Interactions with Demographic Variables (Farm Size)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta )</td>
<td>S.E.’s</td>
</tr>
<tr>
<td>Constant</td>
<td>-7.969***</td>
</tr>
<tr>
<td>Price (Benchmark is conventional)</td>
<td>-0.125***</td>
</tr>
<tr>
<td>HT</td>
<td>1.496**</td>
</tr>
<tr>
<td>IR</td>
<td>-0.014</td>
</tr>
<tr>
<td>HT_IR</td>
<td>1.201**</td>
</tr>
</tbody>
</table>

GMM objective: 0.00000004

Note - Based on 2,488 observations. Parameters are GMM estimates. Estimation includes region and company fixed effects. Standard errors are asymptotically robust.

As expected, the estimates reflect distaste for price and a preference for GM seed relative to conventional seed. Almost all random coefficients are statistically significant indicating that the preferences for the product attributes vary over consumers and that the random-coefficients specification is appropriate. Taking all effects of price together, price enters utility with a coefficient: -0.125 + 0.025* \( \eta \) – 0.139*FarmSize, where \( \eta \) is a standard normal random term. The price coefficient has a constant component, a part that varies with the size of the farm and a part that varies randomly over households. Bigger farms tend to be more price-sensitive.
possibly reflecting some bargaining power. The random coefficients on the genetic attributes HT and HT_IR are also significant, which produces the realistic substitution patterns reported in the next section.

**Elasticities**

The coefficients reported in Table 1.3 represent the effect of different product attributes on farmer utility. The signs are indicative of whether a farmer likes or dislikes a particular attribute, but the magnitudes have no meaningful interpretation. For this reason, the coefficients need to be converted to elasticities. In this section I report the substitution patterns from the random-coefficients model and compare them with elasticities from the logit model.

The effect of a change in the price of product $k$ on the share of product $j$ in market $m$ for the logit model is calculated as:

$$
\varepsilon_{jk} = \frac{\partial s_j p_k}{\partial p_k s_j} = \begin{cases} 
-\alpha p_j(1 - s_j) & \text{if } j = k \\
\alpha p_k s_k & \text{if } j \neq k
\end{cases}.
$$

The formulas for own- and cross-price elasticities implied by the mixed-logit model are:

$$
\varepsilon_{jk} = \frac{\partial s_j p_k}{\partial p_k s_j} = \begin{cases} 
-\frac{p_j}{s_j} \int \alpha_i s_{ij} (1 - s_{ij}) dP_D^*(D) dP_v^*(\nu) & \text{if } j = k \\
\frac{p_k}{s_j} \int \alpha_i s_{ij} s_{ik} dP_D^*(D) dP_v^*(\nu) & \text{if } j \neq k,
\end{cases}
$$

where $P_D(D)$ and $P_v(\nu)$ are the distributions of the demographic and idiosyncratic consumer characteristics\(^{18}\) (Nevo, 2000). As mentioned earlier, while the logit model surpasses the dimensionality problem intrinsic in demand estimation with many products, it produces

\(^{18}\) The smooth simulator is used to simulate these integrals.
unrealistic substitution patterns. Table 1.4 presents the average cross- and own-price elasticities by trait as predicted by the logit model along with average shares.

<table>
<thead>
<tr>
<th>% Δ in pjm</th>
<th>CONV</th>
<th>HT</th>
<th>IR</th>
<th>HT IR</th>
<th>Own-price Elasticities</th>
<th>Number Products</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONV</td>
<td>0.002</td>
<td>0.005</td>
<td>0.022</td>
<td>0.013</td>
<td>-0.607</td>
<td>154</td>
</tr>
<tr>
<td>HT</td>
<td>0.002</td>
<td>0.004</td>
<td>0.023</td>
<td>0.012</td>
<td>-1.380</td>
<td>106</td>
</tr>
<tr>
<td>IR</td>
<td>0.001</td>
<td>0.005</td>
<td>0.023</td>
<td>0.013</td>
<td>-1.700</td>
<td>19</td>
</tr>
<tr>
<td>HT IR</td>
<td>0.001</td>
<td>0.004</td>
<td>0.023</td>
<td>0.012</td>
<td>-2.071</td>
<td>98</td>
</tr>
<tr>
<td>Share</td>
<td>0.002</td>
<td>0.003</td>
<td>0.007</td>
<td>0.006</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The elasticities are to be read across each row heading, and so according to the logit model, when the price of a conventional seed increases, farmers are the most likely to substitute away to IR seed, then to stacked and HT seed, and least likely to substitute towards other conventional seed. The same is true for the predicted substitution patterns for all trait groups. This is clearly unrealistic and follows directly from equation (1.11), which predicts that
substitution will occur towards the “most popular” product in the market – i.e. the product with the largest share.

The mixed-logit model improves on this property and produces realistic substitution patterns by allowing consumer preferences for product attributes to vary with consumer characteristics through the random coefficients. Table 1.5 reports the mixed-logit substitution patterns. Substitution from conventional seed will be more likely towards other conventional seed than any of the GM groups. If the price of HT seed increases, farmers will more likely substitute towards other HT seed; from stacked seed – to stacked seed, and from insect-resistant seed – to insect-resistant seed. This is the behavior we expect in reality.

Similar patterns can be seen if we break the trait groups into the respective subgroups as seen in table 1.6, which reports the elasticities from the mixed-logit model for one out of the 84 markets (Arizona in 2006). Similar substitution patterns hold for other markets as well and this information is available upon request. Focusing on a single market allows me to present the information more concisely.

Farmers purchasing cottonseed from a particular subgroup are much more likely to substitute to products with the exact same or similar genetic content. In contrast, the logit model predicts that farmers would always switch away to IR3 seed, which have the largest market share in the market (see Appendix table A1 for details). Varieties from nine out of the 15 trait groups were sold in this particular market, which is why some trait groups are missing.
Table 1.6. Average Elasticities by Trait for the Mixed-logit Random-coefficients Model for Arizona in 2006

<table>
<thead>
<tr>
<th>% Δ in ( s_{jm} )</th>
<th>CONV</th>
<th>HT3</th>
<th>HT1</th>
<th>IR2-HT3</th>
<th>IR2-HT2</th>
<th>IR2-HT1</th>
<th>IR1-HT1</th>
<th>IR3</th>
<th>IR1</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONV</td>
<td><strong>0.001</strong></td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0003</td>
<td>0.0005</td>
<td>0.0002</td>
<td>0.0004</td>
<td>0.0003</td>
<td>0.0002</td>
</tr>
<tr>
<td>HT3</td>
<td>0.008</td>
<td><strong>0.022</strong></td>
<td><strong>0.025</strong></td>
<td>0.008</td>
<td>0.007</td>
<td>0.009</td>
<td>0.007</td>
<td>0.003</td>
<td>0.004</td>
</tr>
<tr>
<td>HT1</td>
<td>0.004</td>
<td>0.010</td>
<td><strong>0.012</strong></td>
<td>0.004</td>
<td>0.004</td>
<td>0.004</td>
<td>0.004</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>IR2-HT3</td>
<td>0.043</td>
<td>0.014</td>
<td>0.016</td>
<td>0.047</td>
<td><strong>0.062</strong></td>
<td>0.046</td>
<td><strong>0.060</strong></td>
<td>0.028</td>
<td>0.030</td>
</tr>
<tr>
<td>IR2-HT2</td>
<td>0.008</td>
<td>0.001</td>
<td>0.002</td>
<td>0.006</td>
<td>.</td>
<td>0.005</td>
<td><strong>0.009</strong></td>
<td>0.004</td>
<td>0.003</td>
</tr>
<tr>
<td>IR2-HT1</td>
<td>0.093</td>
<td>0.033</td>
<td>0.037</td>
<td><strong>0.107</strong></td>
<td><strong>0.136</strong></td>
<td><strong>0.096</strong></td>
<td><strong>0.131</strong></td>
<td>0.063</td>
<td>0.066</td>
</tr>
<tr>
<td>IR1-HT1</td>
<td>0.003</td>
<td>0.000</td>
<td>0.001</td>
<td><strong>0.003</strong></td>
<td><strong>0.004</strong></td>
<td>0.002</td>
<td>.</td>
<td>0.002</td>
<td>0.001</td>
</tr>
<tr>
<td>IR3</td>
<td>0.898</td>
<td>0.089</td>
<td>0.122</td>
<td>0.489</td>
<td>0.616</td>
<td>0.461</td>
<td>0.594</td>
<td>.</td>
<td><strong>1.046</strong></td>
</tr>
<tr>
<td>IR1</td>
<td>0.002</td>
<td>0.001</td>
<td>0.001</td>
<td>0.003</td>
<td>0.002</td>
<td>0.003</td>
<td>0.003</td>
<td><strong>0.006</strong></td>
<td>.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>% Δ in ( P_{km} )</th>
<th>CONV</th>
<th>HT3</th>
<th>HT1</th>
<th>IR2-HT3</th>
<th>IR2-HT2</th>
<th>IR2-HT1</th>
<th>IR1-HT1</th>
<th>IR3</th>
<th>IR1</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONV</td>
<td><strong>0.001</strong></td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0003</td>
<td>0.0005</td>
<td>0.0002</td>
<td>0.0004</td>
<td>0.0003</td>
<td>0.0002</td>
</tr>
<tr>
<td>HT3</td>
<td>0.008</td>
<td><strong>0.022</strong></td>
<td><strong>0.025</strong></td>
<td>0.008</td>
<td>0.007</td>
<td>0.009</td>
<td>0.007</td>
<td>0.003</td>
<td>0.004</td>
</tr>
<tr>
<td>HT1</td>
<td>0.004</td>
<td>0.010</td>
<td><strong>0.012</strong></td>
<td>0.004</td>
<td>0.004</td>
<td>0.004</td>
<td>0.004</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>IR2-HT3</td>
<td>0.043</td>
<td>0.014</td>
<td>0.016</td>
<td>0.047</td>
<td><strong>0.062</strong></td>
<td>0.046</td>
<td><strong>0.060</strong></td>
<td>0.028</td>
<td>0.030</td>
</tr>
<tr>
<td>IR2-HT2</td>
<td>0.008</td>
<td>0.001</td>
<td>0.002</td>
<td>0.006</td>
<td>.</td>
<td>0.005</td>
<td><strong>0.009</strong></td>
<td>0.004</td>
<td>0.003</td>
</tr>
<tr>
<td>IR2-HT1</td>
<td>0.093</td>
<td>0.033</td>
<td>0.037</td>
<td><strong>0.107</strong></td>
<td><strong>0.136</strong></td>
<td><strong>0.096</strong></td>
<td><strong>0.131</strong></td>
<td>0.063</td>
<td>0.066</td>
</tr>
<tr>
<td>IR1-HT1</td>
<td>0.003</td>
<td>0.000</td>
<td>0.001</td>
<td><strong>0.003</strong></td>
<td><strong>0.004</strong></td>
<td>0.002</td>
<td>.</td>
<td>0.002</td>
<td>0.001</td>
</tr>
<tr>
<td>IR3</td>
<td>0.898</td>
<td>0.089</td>
<td>0.122</td>
<td>0.489</td>
<td>0.616</td>
<td>0.461</td>
<td>0.594</td>
<td>.</td>
<td><strong>1.046</strong></td>
</tr>
<tr>
<td>IR1</td>
<td>0.002</td>
<td>0.001</td>
<td>0.001</td>
<td>0.003</td>
<td>0.002</td>
<td>0.003</td>
<td>0.003</td>
<td><strong>0.006</strong></td>
<td>.</td>
</tr>
</tbody>
</table>

Note - These elasticities are the averages for a given market (Arizona in 2006), but the substitution patterns hold for all markets. The “.” in the diagonal corresponds to the fact that there is only one variety of a particular trait group in this market, which makes calculating cross-price elasticities within the group impossible.
The average own-price elasticities at the national level across all seven years are \(-0.766\) for conventional seed, \(-1.961\) for HT seed, \(-1.865\) for stacked seed, and \(-1.523\) for IR seed. Farmers are less price-sensitive toward conventional seed, which could be considered necessity goods relative to GM seed. Price sensitivity is highest for HT and stacked seed reflecting the largest number of available substitute varieties. In fact, own elasticities for these two seed types increases with time as adoption and number of products increases.

### 1.5. Conclusion

In this chapter I estimated a mixed-logit random-coefficients demand model for cottonseed in the U.S. for the period 2000-2007. This model allowed me to recover farmers’ preference parameters for different product attributes in a differentiated-products setting with close to 400 products while maintaining realistic substitution patterns. Traditional demand estimation methods would not work in the present context since they would require estimating close to 160,000 parameters. The BLP demand analysis presented here reduces the number of estimated coefficients to the size of the product characteristics space. In particular, I only estimate 45 parameters including various product attributes, their interactions with consumer characteristics, and various fixed effects.

An important limitation of my work is the fact that I do not account for zero market shares. Gandhi et al. (2014) shows that this can bias the demand estimates towards zero and underestimate price elasticities. Their paper proposes a Bayesian method for dealing with this problem.
I find that farmers dislike price and prefer genetically modified seed to conventional seed, which is consistent with common sense and industry trends. I also find that these preferences are not the same across farmers and vary by individual. The model produces realistic substitution patterns and although computationally cumbersome, it is better than the discrete-choice logit demand model. I also find that own-price elasticities are lower than previous estimates for the corn industry. In particular, Ma (forthcoming) reports own average elasticities, which are all above 4.5 in absolute value. This confirms that farmers are more price-sensitive towards cottonseed for which there are fewer available substitute products relative to the corn seed market.
Chapter 2: The Effects of Biotech Innovation in the Market for Crop Seed

2.1. Introduction

The seed production sector is made up of an upstream market for genetic material and a downstream market for seed varieties with or without genetic modifications. In the U.S. there are three biotech firms producing genetic material for cottonseed. The number of cottonseed firms downstream decreased from 17 in 2000 to 11 in 2007. Thirteen of the seed firms sell genetically-modified seed in addition to conventional seed. Biotech companies sell the traits they have patented either through seed subsidiaries if the biotech firm is vertically integrated or through licensing to independent seed firms. Innovation takes place both in the upstream and in the downstream sectors.

The introduction of new traits upstream spurts even more product introductions downstream. Figure 2.1 shows the number of new traits introduced each year between 1996 and 2010 for cotton and corn seed. New traits vary between 0 and 4 per year for cottonseed and between 0 and 8 for corn seed. Figure 2.2 shows downstream seed innovation for cottonseed varieties. The number of new hybrids per year is between 10 and 63.\(^\text{19}\) This aggressive product introduction begs the question “Are firms innovating too much in search of first mover advantage and the temporary rents that come with an innovation?” Farmer groups have been complaining that there is too much innovation on crop seed. At the same time, they are compelled to buy the new seed in order to keep up with competition. Does this mean that

\(^{19}\) Hybrid entries for corn seed increase from 1,200 per year to 2,500 per year between 2000 and 2007.
consumer surplus increases as a result of GM innovations? Similarly, what is the effect from
innovation on seed firms’ profits?

In the past 17 years biotech companies have been accused of using innovation to reap surplus
from farmers. I address this concern by looking at a specific genetic innovation in cottonseed. I
estimate demand for cottonseed\(^\text{20}\) and quantify the economic effects from the introduction of
the HT1 gene. This is the first instance of using the discrete-choice methodology to analyze
innovation in cottonseed.

I estimate farmers’ demand for cottonseed sold by seed firms in the downstream seed market,
and then calculate changes in farmers’ surplus and in seed firms’ variable profits from the
innovation. I do this using a panel dataset with regional variation and changing products over

\[ \text{Figure 2.1. Trait Innovations} \]
\[ \text{New Traits for Corn and Cotton} \]

\[ \text{Source: Josh et al.(2008), Howell et al.(2009), Magnier et al.(2007)} \]

---

\(^{20}\) Demand estimation alone is valuable to firms as they devise their pricing strategies and to regulators as they
consider market power and merger effects. Shapiro (1996) describes how cross-price elasticity is used in
constructing diversion ratios and evaluating unilateral merger effects in mergers with differentiated products (e.g.
the merger between Interstate Bakeries Corporation and Continental Baking Company; also, the acquisition by
Kimberly-Clark of the Scott Paper Company).
time. I do not analyze the upstream biotech sector and do not address profits to the innovating biotech company which, no doubt, are positive.

There is a large early literature on the distribution of benefits from the introduction of GM seed. These papers focus on a specific type of seed (e.g. herbicide-tolerant corn) and model it as a homogeneous good (Frisvold et al., 2000; Falck-Zepeda, 1999; Falck-Zepeda et al., 2000). Moschini et al. (2000) estimate that 44% of the benefits from the introduction of Roundup Ready (RR) soybeans (a genetically-modified herbicide-tolerant variety) go to the innovators, 16% to the farmers, and 40% to the final consumers. The problem with these studies is that the seed products within each GM group are not homogeneous. There are significant differences in prices and gene content.

---

21 The small share of benefits allocated to farmers has aroused concern among regulators as reflected in a report by the American Anti-trust Institute called “Transgenic Seed Platforms: Competition between a Rock and a Hard Place?” (Moss, 2009).
In the context of differentiated products, welfare estimation from new products has received much attention in the discrete choice literature. Petrin (2002) calculates consumer and producer surplus changes following the introduction of the minivan. Goolsbee et al. (2004) quantify the consumer gains from the entry of direct broadcast satellites. Other goods under study have been computers (Greenstein, 1994), health care technology (Trajtenberg, 1989), breakfast cereals (Hausman, 1994), telecommunication services (Hausman, 1997), and cellular phones (Hausman, 1999). My methodology is similar to Petrin’s study of the minivan (Petrin, 2002), and this is the first application of the discrete-choice framework to evaluate innovation effects in cottonseed.

2.2. Supply Model, Producer Profits, and Consumer Surplus

I make assumptions about the nature of competition and use my estimates of demand to construct counterfactuals without the introduction one cottonseed genetic modification. Marginal costs are calculated under multi-product Bertrand competition and under the observed prices. Next, I solve for the new equilibrium prices for a counterfactual scenario. This allows me to compute the change in variable industry profits as well as the change in farmer surplus.

Each firm $f$ offers a set $K$ of products. Under multi-product Bertrand pricing the firm’s maximization problem is

$$\max_{p_j, \forall j \in K} \pi_f = \max_{p_j, \forall j \in K} \sum_{j \in K} (p_j - mc_j)[s_j(p, X; \theta) \times M] - F_f.$$
Product $j$’s market share $s_j$ is a function of all other products’ prices $p$, product characteristics $X$, and the demand parameters $\theta$. Sales for product $j$ equal $s_j$ times market size $M$. Firm-specific fixed cost is denoted by $F_f$. Given this maximization problem, the K first order conditions are

$$\frac{d\pi_f}{dp_j} = 0 \text{ or}$$

$$(2.2) \quad s_j(p, X; \theta) + \sum_{k \in K}(p_k - m c_k) \frac{\partial s_k(p, X; \theta)}{\partial p_j} = 0.$$  

Each firm solves K of these first order conditions. Markups $(p_k - m c_k)$ are related to products’ market shares and how many products are made by the same parent firm. Firms with more products will have higher estimated markups, whereas firms with few products will have smaller estimated markups (BLP, p.874). In matrix notation the above first order conditions can be expressed as

$$\begin{equation}
(2.3) \quad S + D(P - MC) = 0,
\end{equation}$$

where $S$, $P$, and $MC$ are K x 1 matrices of product shares, prices, and marginal costs, and $D$ is a K x K matrix of own- and cross-price elasticities of demand. This identity can be rearranged to yield marginal costs as:

$$\begin{equation}
(2.4) \quad MC = P + D^{-1} \times S.
\end{equation}$$

Counterfactual

Once marginal costs are calculated, I construct the counterfactual scenario by removing the HT1 innovation from all markets. This reduces the number of products by 40 and the number of
observations by 250, which in turn reduces the number of first order conditions faced by each firm. The system of equations (2.4) is rearranged to solve for the new equilibrium price vectors that clear the market absent HT1 varieties:

\[ P = MC - D^{-1} \times S. \]

Marginal cost is assumed to be unaffected by the counterfactual. While prices are treated as endogenous, upstream gene offerings and downstream product variety offerings are assumed to be exogenous. This means that product offerings by biotech and seed firms are assumed to be unaffected by the innovation. Shares are also endogenous and are solved for as functions of the counterfactual prices through an iterative process that yields the prices and shares which minimize the firms’ first order conditions\(^{22}\).

The new prices and shares are used to compute the change in variable profits under the counterfactual for each firm \( f \):

\[ \pi_f(P_{act}, mc; \theta) - \pi_f(P_{cf}, mc; \theta), \]

where \( P_{act} \) is the initial vector of prices and \( P_{cf} \) are the counterfactual prices. Summing over all firms gives the total change in producer profits:

\[ \sum_f [\pi_f(P_{act}, mc; \theta) - \pi_f(P_{cf}, mc; \theta)]. \]

Compensating variation (CV) is used to measure changes in consumer welfare from the introduction of HT1 cottonseed. The compensating variation is the change in consumers’

\(^{22}\) MATLAB’s fsolve command is used.
income that equates utility in a particular situation to some chosen benchmark utility. Here, the benchmark will be utility without HT1 cottonseed, which is constructed through simulation. Compensating variation is the dollar amount a consumer would need to be compensated in order to be indifferent between the equilibrium with HT1 seed and the equilibrium without HT1 seed.

Using the initial vector of prices and the vector of counterfactual prices, I calculate the CV for individual $i$ as

$$
\Delta E [CS_i] = CV_i = \frac{\ln[\sum_{j \in J_{1, \text{act}}} \exp(V_{ij}^{\text{act}})] - \ln[\sum_{j \in J_{1, \text{cf}}} \exp(V_{ij}^{\text{cf}})]}{\bar{e}_i},
$$

where $V_{ij}$ is defined as in equation (1.1). $V_{ij}^{\text{act}}$ is consumer $i$’s utility from product $j$ in the baseline scenario, i.e. with HT1 cottonseed available as a choice, while $V_{ij}^{\text{cf}}$ is consumer $i$’s utility in the counterfactual scenario without HT1 products. Therefore, a positive CV would mean that consumers gained from the introduction whereas a negative CV would mean consumers lost surplus. Total compensating variation is obtained by aggregating over consumers.

2.3. Data

The data used for this analysis is the DMR farmer survey with descriptive statistics as reported in Table 1.1. Table 2.1 presents summary sales by trait as a percent of all cottonseed sales for 2000-2007. Seed containing HT3, HT4, BG, and BG-HT3 were in the market before 2000. Their sales go down with the introduction of seed containing new traits and some are completely taken out of the market (HT4 and its derivative HT4-Bt). The IR2 trait was first introduced in
2003; HT2 first hit the market in 2004; and IR1 and HT1 came out in 2005. Sales of varieties with
the HT1 trait climbed from less than one percent in 2005 to 13.5 percent in 2007. Together with
its derivatives, IR2 – HT1 and IR1 – HT1, it had captured 42% percent of all cottonseed sales in
2007. I explore counterfactuals without the introduction of HT1 and its derivatives IR1 – HT1
and IR2 – HT1, which have had the highest economic significance, as reflected by the trait

Table 2.1. National Trait Shares (Seed Sales by Trait as Percent of Total Cottonseed Sales)

<table>
<thead>
<tr>
<th>Year</th>
<th>HT Traits</th>
<th>IR Traits</th>
<th>Stacked</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HT1</td>
<td>HT2</td>
<td>HT3</td>
</tr>
<tr>
<td>2000</td>
<td>.</td>
<td>36.02</td>
<td>8.39</td>
</tr>
<tr>
<td>2002</td>
<td>.</td>
<td>45.57</td>
<td>3.11</td>
</tr>
<tr>
<td>2003</td>
<td>.</td>
<td>37.96</td>
<td>0.24</td>
</tr>
<tr>
<td>2004</td>
<td>.</td>
<td>1.57</td>
<td>31.55</td>
</tr>
<tr>
<td>2005</td>
<td>0.04</td>
<td>3.25</td>
<td>32.44</td>
</tr>
<tr>
<td>2006</td>
<td>3.42</td>
<td>2.34</td>
<td>20.98</td>
</tr>
<tr>
<td>2007</td>
<td>12.2</td>
<td>1.75</td>
<td>8.81</td>
</tr>
</tbody>
</table>

Note – Conventional seed are not shown.

shares, out of all new trait introductions. Logit demand estimates result in unrealistic
substitution patterns and are therefore inadequate in modeling industry behavior.

2.4. Results

I construct a counterfactual scenario without the introduction of the HT1 technology. I do this
by removing all observations with HT1 seed types. A total of 73 varieties containing the HT1
gene (single or the stacked IR1 - HT1 and IR2 – HT1) are removed. This decreases the menu of
choices faced by farmers as well as the number of first order conditions solved by the firms.
Table 2.2 summarizes the changes in equilibrium prices for different seed types that occur as a result of the entry of HT1 cottonseed in 2007.

The average prices by trait group decreased for most seed types by as little as 1% to as much as 19%. Average prices increased for three of the trait groups. Two of these groups are produced and sold by only one company and this could be the reason why we are not seeing the

Table 2.2
Average Prices with and without HT1, 2007 [2000 CPI-Adjusted Dollars]

<table>
<thead>
<tr>
<th></th>
<th>Without RR Flex (Simulated Prices)</th>
<th>With RR Flex (Observed Prices)</th>
<th>Δ Price</th>
<th>% Δ Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>HT3</td>
<td>42.11</td>
<td>33.92</td>
<td>-8.19</td>
<td>-19</td>
</tr>
<tr>
<td>Conventional</td>
<td>14.52</td>
<td>11.72</td>
<td>-2.80</td>
<td>-19</td>
</tr>
<tr>
<td>IR2 - HT3</td>
<td>47.89</td>
<td>42.76</td>
<td>-5.13</td>
<td>-11</td>
</tr>
<tr>
<td>IR3</td>
<td>18.91</td>
<td>17.14</td>
<td>-1.77</td>
<td>-9</td>
</tr>
<tr>
<td>IR2</td>
<td>26.66</td>
<td>25.59</td>
<td>-1.08</td>
<td>-4</td>
</tr>
<tr>
<td>IR1</td>
<td>53.18</td>
<td>52.01</td>
<td>-1.17</td>
<td>-2</td>
</tr>
<tr>
<td>IR1 - HT3</td>
<td>58.66</td>
<td>58.32</td>
<td>-0.34</td>
<td>-1</td>
</tr>
<tr>
<td>IR2 - HT2</td>
<td>35.29</td>
<td>37.00</td>
<td>1.71</td>
<td>5</td>
</tr>
<tr>
<td>IR3 - HT3</td>
<td>51.58</td>
<td>55.32</td>
<td>3.74</td>
<td>7</td>
</tr>
<tr>
<td>HT2</td>
<td>20.24</td>
<td>22.14</td>
<td>1.90</td>
<td>9</td>
</tr>
</tbody>
</table>

Note - Equilibrium prices without RR Flex seed are estimated using the model with random coefficients and Bertrand-Nash first order conditions.

Table 2.3 reports the national share changes induced by the innovation in 2007. The first column presents the simulated counterfactual acre shares by trait. These are obtained from the simulated product market shares by recovering the counterfactual number of acres per category and dividing by the total number of acres for that year. The second column presents actual observed national shares, and the last two columns present the competitive effect there.
share changes induced by the innovation. The results from the simulation show that the introduction of the HT1 gene resulted in farmers planting more stacked cottonseed and less conventional, HT, and IR cottonseed.

Table 2.4 summarizes changes in consumer and producer welfare across the industry for the years 2006 - 2007 after the introduction of HT1. Farmer’s compensating variation increased by $207 million due to the innovation. Seed firms’ variable profits also increased in spite of the fact that incumbent products’ prices fell.
2.5. Conclusion

In this chapter I present a static supply model of firm behavior and analyze the economic effects from innovation in the cottonseed industry. In particular, this paper quantifies the farmer and seed firm gains from the introduction of HT1 seed types in 2005. I find that farmer compensating variation increased by $207 million in 2006 and 2007 combined. This is 2% of the total value of cotton production of $10.6 billion for that period. This finding is informative in light of the controversy surrounding the HT1 technology. I also find that in spite of increased product competition and lower prices for incumbent products in the seed market, the gains from the HT1 technology outweigh the losses and the net effect on seed firms’ variable profits is positive. In particular, seed firms gain $189 million in profits from the introduction.

The analysis in this paper could be enriched in a number of ways. First, the problem of zero market shares could be addressed by using the estimator proposed by Gandhi et al. (2014). Without doing so, the demand estimates are likely biased towards zero and the markups are likely overestimated.

Moreover, product offerings by firms in the upstream and the downstream sectors could be modelled explicitly. Eizenberg (2013) does some of this in his analysis of laptop innovation with an upstream CPU market and a downstream personal computers market. His work provides a template for how an extension like this could be carried out. Crawford (2012) presents a survey of the literature that endogenizes all product characteristics (not just price) in empirical differentiated-products demand models. He documents the effects of endogenizing product
choice on welfare calculations are modest (Crawford, 2012, p.320). Finally, another way to extend the current work would be to incorporate dynamics in consumer and firm behavior. There are some papers that estimate a dynamic discrete-choice demand model (Goettler et al., 2011; Gowrisankaran et al., 2012; Sweeting, 2013), but none of them incorporate heterogeneous consumers and multi-product firms on the supply side.
Chapter 3: Product Turnover Rate: What Technical and Market Factors Influence the Survival of Cottonseed?

3.1. Introduction

The biotech seed sector has undergone significant structural change since 1996. Horizontal mergers between seed firms, vertical mergers and acquisitions, and entry by new firms have attracted the attention of policy-makers, various farm groups, and academics. Concerns are raised not only about prices, but also about the availability of seed. While the introduction of new production technologies has doubtlessly increased the number of products offered, the isolated effect of industry consolidation on product availability is unclear. Consolidation has led to an increase in the number of bigger vertically integrated firms, while the number of small local independent seed companies has gone down. If products sold by bigger integrated firms survive longer, this could affect the investment decisions of smaller local firms, and decrease the availability of seed specially adapted to meet local climate and soil conditions. This chapter measures the effect of technology and market factors on product turnover rate and addresses concerns about local product diversity.

Structural modeling in empirical survival analysis is uncommon due to identification problems, which are empirically intractable. There is some empirical literature on the turnover rate of firms, but few empirical papers discuss product survival\(^{23}\). Bayus (1998) and Greenstein et al. (1998) conduct product-level duration analysis for computers using a proportional hazard and a

---

\(^{23}\) Many firms conduct survival analysis for their products, but the results are only used internally and usually not published.
Weibull hazard models. Asplund et al. (1999) study the life cycle of Swedish beer with a Weibull model, while Requena-Silvente et al. (2009) analyze the survival of UK cars using a proportional hazard model.

In the biotech crop seed literature Magnier et al. (2010) use an Accelerated Failure Time (AFT) hazard model to analyze corn seed survival between 1997 and 2009. Ma et al. (2010) use a proportional hazard model to look at corn seed life cycles between 2000 and 2007. They enrich the analysis by evaluating the effect of time-varying covariates like market concentration on product duration. I study cottonseed survival using a method similar to the one used by Ma and Shi (2010).

With the introduction of GM technology there are several potentially conflicting forces into play. On the one hand, more new products could lead to faster product turnover and decrease survival. On the other hand, better technology would lead to higher quality seed, which could survive longer. I expect to find evidence of external and internal competition effects (Greenstein et al., 1998, Ma et al., 2010). These imply that the number of products offered by other firms and the number of products offered by the same firm in a market would both have a negative effect on product survival. The internal competition hypothesis is also called a “cannibalization effect”. Ma et al. (2010) find counterintuitive evidence against the internal and external competition effects in their analysis of corn seed. They report that the number of corn hybrids offered by same firm and the number of GM corn hybrids offered by other firms have a positive effect on product survival. This is explained through growing positive reputation effects. Farmers might buy more from firms that offer more variety, which is why firms selling
more products enjoy longer product life. In addition, they suspect that there are positive learning spillover effects among GM technologies as farmers pay attention to the adoption of GM seed and relate their own purchases to adoption trends. This would explain the evidence against external competition effects.

I find that after controlling for the flow of products in addition to the stock of products, the reputation effect remains, but the “trending” effect disappears. In other words, the number of similar products offered by other companies in the same market no longer has a positive effect on survival. Moreover, the number of new products introduced nationally each year decreases product survival which constitutes evidence of product competition.

The effect of vertical integration on product survival is of special interest. In my sample the vertically integrated firms tend to be larger multinational seed firms, whereas independent seed firms are smaller local companies. The latter contribute significantly to local product differentiation. I find that vertical integration increases product survival, which puts smaller independent seed firms at a disadvantage. This could affect their R&D investment decisions and decrease the availability of locally differentiated seed. This is consistent with evidence from Ma et al. (2010) who document a similar effect of vertical integration for corn seed.

The rest of this chapter proceeds as follows: Section 3.2 presents the model, Section 3.3 discusses the data and some trends, Section 3.4 reports the results, and Section 3.5 concludes.
3.2. Survival Model

I will use duration analysis to evaluate the effect of technology and market factors on product survival. The advantage of using this methodology as opposed to a binary model of exit is that it takes into account both the occurrence and the timing of exit while estimating the effects of explanatory variables. Duration models look at the probability of a certain event taking place (a failure) in a particular period conditional on survival up to that period.

In particular, I will model product exit using a Cox Proportional Hazard (PH) model. This model is based on the proportional hazard (PH) assumption, which states that when all covariates are set to zero, the baseline hazard rate only depends on time and is independent of when the covariates change their values. All proportional hazard models take the form:

\[ h(t, X)_j = h_0(t)k(X), \]

where \( h \) is the hazard rate for subject \( j \), \( h_0(t) \) is the baseline hazard, and \( X \) is a set of subject characteristics. The Cox PH model assumes that function \( k(\cdot) \) is exponential:

\[ h(t, X)_j = h_0(t)e^{\theta X} \]

and \( \theta \) are the coefficients to be estimated. The reduced-form estimation equation for subject \( j \) in analysis time \( t \) becomes

\[ \log(h(t, X))_j = \varphi_0 + \theta X. \]
Econometric Concerns

I address three econometric concerns in this model. One of them is that some of my covariates vary with time, which could violate the PH assumption, i.e. baseline hazard may depend on when the time-varying covariates change. I test the PH assumption using a goodness of fit test with Schoenfeld residuals. The null hypothesis is that proportional hazard holds and a high p-value indicates that the PH assumption cannot be rejected.

A second econometric concern is right- and left-censoring, which could bias my estimates. I take care of left-censoring by considering an entry only if it occurred after 2000\(^24\). I deal with right-censoring by defining exit only if it occurred before 2007\(^25\). Kaplan-Meier survival estimation and the Cox PH model rely on the independent censoring assumption (i.e. censoring is random within a covariate groups) to obtain valid estimates (Kleinbaum et al., 2005). I test this assumption by re-estimating the model on the uncensored sample and comparing the results.

A third econometric concern is endogeneity. The dependent variable, which describes product exit, may have a feedback effect on the market structure covariates. In particular, product exit may affect firm market share, and the number of own and competing products. I deal with endogeneity by using the introductory-year values of the market concentration variables for

\(^{24}\) Since 75\% of the subjects have survival times of up to three years (Kaplan-Meier survival estimates at the national level using only the within-sample subjects), ignoring 2000 and 2001 should take care of most of the left-censoring problem. Ma and Shi (2010) as well as Magnier et al. (2010) follow this approach.

\(^{25}\) I then use STATA’s –stcox- command, which accounts for right censoring.
each subject (following Ma and Shi, 2010) since exit is not likely to have an effect on concentration values at the year of entry.

3.3. Data

My analysis relies on the DMR dataset covering cottonseed purchases in the U.S. between 2000 and 2007. I observe that the U.S. cottonseed market is characterized by a large number of varieties as presented in Figure 3.1. The number of all cottonseed products varies between 146 and 153 and there is a 12% increase between 2002 and 2006. The number of GM varieties increases by 150% between 2000 and 2007. This is good news as it means that on average more varieties have become available to farmers with the introduction of GM technologies. What some people are concerned about is that the number of conventional varieties has been decreasing with the introduction of biotech seed. This can be seen on the graph as the shrinking gap between the total and GM number of products.

Figure 3.1. Total Number of Cotton Varieties in the U.S.
The dataset suffers both from left and right censoring. Some varieties observed in 2000 entered the market before the first survey year, and some products observed in 2007 are also available in subsequent years. A product is defined as entering the market when it is not observed in any previous year. A product is defined as exiting the market when it is not observed in any subsequent year. I focus my analysis on the sub-sample of “uncensored” subjects, which entered the market after 2000. I consider an exit only if it occurred before 2007 and account for ignorable censoring.

Table 3.1 summarizes the variables used in the survival analysis. There are 1,497 observations total, out of which 710 are not censored at all (entry occurs after 2000 and exit occurs before 2007). Out of the 710 uncensored observations 146 are conventional and 564 are genetically-modified26. Conventional seed survive 2.37 years on average and GM seed survive slightly longer. It is worth noting that 83% of the right-censored observations are genetically-modified, and so the average survival time for GM seed is probably biased down.

Figure 3.2 shows Kaplan-Meier survival estimates by seed trait content. These are to be interpreted as the probability of survival if the seed is one through five years old. Stacked and HT seed seem to survive longer than both IR and conventional cottonseed. While this is informative in terms of average behavior, the survival model will consider whether HT and stacked seed still survive longer after controlling for other factors.

---
26 This uncensored sample is made up of 417 subjects in 12 states over 5 years. A subject is defined at the state/variety level, so that the same variety appearing in different states is counted as different subjects.
### Table 3.1. Descriptive Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Obs</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>survivalCONV</td>
<td>Survival Conventional (years from introduction to exit)</td>
<td>146</td>
<td>2.370</td>
<td>1.339</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>survivalGM</td>
<td>Survival GM</td>
<td>564</td>
<td>2.433</td>
<td>1.277</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>year_id</td>
<td>Year</td>
<td>1497</td>
<td>2004.960</td>
<td>1.641</td>
<td>2002</td>
<td>2007</td>
</tr>
<tr>
<td>HT</td>
<td>HT dummy</td>
<td>1497</td>
<td>0.339</td>
<td>0.474</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>IR</td>
<td>IR dummy</td>
<td>1497</td>
<td>0.030</td>
<td>0.171</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>HT_IR</td>
<td>Stacked seed dummy</td>
<td>1497</td>
<td>0.491</td>
<td>0.500</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>CONV</td>
<td>Conventional dummy</td>
<td>1497</td>
<td>0.140</td>
<td>0.347</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>firm_share</td>
<td>Firm market share within each trait/state/year.</td>
<td>1497</td>
<td>0.385</td>
<td>0.336</td>
<td>4.32E-05</td>
<td>1</td>
</tr>
<tr>
<td>Dvi</td>
<td>Integration Dummy</td>
<td>1497</td>
<td>0.359</td>
<td>0.480</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>N_own</td>
<td>The number of similar products by the same firm in a state/year</td>
<td>1497</td>
<td>4.976</td>
<td>3.571</td>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>N_others</td>
<td>The number of similar products by other firms in a state/year</td>
<td>1497</td>
<td>13.399</td>
<td>12.930</td>
<td>0</td>
<td>51</td>
</tr>
<tr>
<td>N_entries</td>
<td>Number of new products introduced each year</td>
<td>1497</td>
<td>39.277</td>
<td>13.941</td>
<td>26</td>
<td>63</td>
</tr>
</tbody>
</table>

Note – A subject is defined at the state/variety level. There are 750 subjects. An observation is a subject-year combination and there are 1497 observations.

![Kaplan-Meier survival estimates](image-url)
3.4. Results

Table 3.2 presents the coefficient estimates from the Cox proportional model. The dependent variable is the log of the hazard rate (probability of exit conditional on survival until time $t$ from introduction). A negative coefficient is to be interpreted as lowering the probability of failure and increasing survival probability. Positive coefficients imply a negative effect on survival.

Table 3.2. Cox PH Model Coefficient Estimates

<table>
<thead>
<tr>
<th>Biotech Characteristics (benchmark is conventional)</th>
<th>(1)</th>
<th>Schoenfeld p-values</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HT</td>
<td>-0.478***</td>
<td>0.276</td>
<td>-0.565***</td>
<td>-0.586***</td>
</tr>
<tr>
<td></td>
<td>(0.116)</td>
<td></td>
<td>(0.114)</td>
<td>(0.116)</td>
</tr>
<tr>
<td>IR</td>
<td>-0.030</td>
<td>0.826</td>
<td>-0.184</td>
<td>-0.042</td>
</tr>
<tr>
<td></td>
<td>(0.251)</td>
<td></td>
<td>(0.231)</td>
<td>(0.242)</td>
</tr>
<tr>
<td>Stacked</td>
<td>-0.675***</td>
<td>0.246</td>
<td>-0.720***</td>
<td>-0.820***</td>
</tr>
<tr>
<td></td>
<td>(0.115)</td>
<td></td>
<td>(0.111)</td>
<td>(0.115)</td>
</tr>
<tr>
<td>Market Structure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integration</td>
<td>-0.269***</td>
<td>0.928</td>
<td>-0.300***</td>
<td>-0.282***</td>
</tr>
<tr>
<td></td>
<td>(0.091)</td>
<td></td>
<td>(0.096)</td>
<td>(0.096)</td>
</tr>
<tr>
<td>Firm Share</td>
<td>-0.643***</td>
<td>0.459</td>
<td>-0.467***</td>
<td>-0.499***</td>
</tr>
<tr>
<td></td>
<td>(0.170)</td>
<td></td>
<td>(0.170)</td>
<td>(0.169)</td>
</tr>
<tr>
<td>$N_{own}$</td>
<td>-0.027*</td>
<td>0.877</td>
<td>-0.054***</td>
<td>-0.033*</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td></td>
<td>(0.018)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>$N_{others}$</td>
<td>-0.020***</td>
<td>0.981</td>
<td>-0.017***</td>
<td>-0.006</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td></td>
<td>(0.004)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>$N_{entries}$</td>
<td>0.021***</td>
<td>0.571</td>
<td>0.011***</td>
<td>0.009***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
</tbody>
</table>

| Intro-year values | No | Yes | Yes |
| Fixed Effects     | -  | -   | State |
| Observations      | 1,497 | 1,497 | 1,497 |
| Log Likelihood    | -2501 | -2512 | -2505 |

Note - The dependent variable is log(hazard rate), i.e. log(probability of exit given survival up to time $t$). Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1
Specification (1) presents the coefficient estimates with the time-varying covariates and reports the results from the test of the PH assumption. The Schoenfeld Residuals p-values are all greater than 0.2 indicating that the null hypothesis (that the PH assumption holds) cannot be rejected at any reasonable level – i.e. the PH assumption is not violated. Specification (2) presents the results after controlling for endogeneity and specification (3) adds state fixed effects. All coefficient estimates show that HT and stacked cottonseed have a lower hazard on average and survive significantly longer than the baseline conventional seed. It must be the case that the new technology results into higher quality cottonseed, which increases product life. Insect-resistant cottonseed, on the other hand, survives about as long as conventional varieties. I suspect that IR seed are less popular as farmers prefer stacked seed, which include both HT and IR characteristics.

The coefficients on N_own and N_others are both negative and significant in the first two specifications. It appears that as the number of competing products increases, the hazard rate decreases and product survival increases. This is evidence against the internal and external competition effects and could be due to a positive reputation and word-of-mouth effect (Ma et al., 2010). After controlling for state fixed effects in column (3), N_others becomes insignificant and N_own loses most of its significance. Both of these covariates are defined at the state-trait-year level, and so after controlling for state-level unobservables, the effects are no longer as significant. The average product still benefit from reputation effects by similar products, but only if they are sold by the same firm. While I find no evidence of product competition effects

27 The current analysis includes subjects, which were only observed for one year. If I exclude those, the coefficient on IR becomes significant.
when considering the stock of similar products, the results in table 3.2 suggest that the number of new products introduced on the U.S. market is hazardous for existing products and decreases survival.

Table 3.2 also shows the effect of firm market share and vertical integration on product life. Products sold by firms with larger market share have a lower hazard rate and survive longer as indicated by the negative and significant coefficient on firm share. In addition, products sold by vertically integrated firms survive significantly longer than products sold by independent seed companies.

Marginal Effects

The estimates in Table 3.2 present the effect of different covariates on the log of the hazard rate (see Equation (3.3)). A more meaningful effect is given by the hazard ratio – the change in hazard rates resulting from an increase in the covariate from \( X^* \) to \( X^{**} \):

\[
(3.4) \quad HR = \frac{h(t,X^{**})}{h(t,X^*)} \equiv \frac{h_0(t)e^{\beta X^{**}}}{h_0(t)e^{\beta X^*}} = e^{\beta(X^{**}-X^*)} = \begin{cases} e^{\beta(1-0)} = e^0 \text{ if the covariate is discrete} \\ e^{\beta(X^{**}-X^*)} \text{ if the covariate is continuous} \end{cases}
\]

For discrete and factor covariates I report the change in hazard probability between the baseline (\( D=0 \)) and comparison situations (\( D=1 \)). For continuous variables, the hazard ratio is

the effect of a one standard deviation increase from the mean\(^{28}\).

Table 3.3 presents the marginal effects of the technology and market factors. The first column restates the coefficient estimates from Table 3.2. The second and third columns report the

\(^{28}\) I report average marginal effects as opposed to marginal effects at the averages.
Table 3.3. Marginal Effects Calculated Using Coefficient Estimates from the Model with State FE’s and Introductory-year Values for the Time-Varying Covariates

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficients</th>
<th>Baseline (D = 0, X = mean)</th>
<th>Comparison (D = 1, X = (mean + SD))</th>
<th>Hazard Ratio</th>
<th>Marginal Effect on Survival Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Biotech Characteristics (benchmark is conventional)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HT</td>
<td>-0.586***</td>
<td>(0.116)</td>
<td>Conventional</td>
<td>0.56***</td>
<td>44%***</td>
</tr>
<tr>
<td>IR</td>
<td>-0.042</td>
<td>(0.242)</td>
<td>Conventional</td>
<td>0.96</td>
<td>4%</td>
</tr>
<tr>
<td>Stacked</td>
<td>-0.820***</td>
<td>(0.115)</td>
<td>Conventional</td>
<td>0.44***</td>
<td>56%***</td>
</tr>
<tr>
<td><strong>Market Structure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integration</td>
<td>-0.282***</td>
<td>(0.096)</td>
<td>Not Integrated</td>
<td>0.75***</td>
<td>25%***</td>
</tr>
<tr>
<td>Firm Share</td>
<td>-0.499***</td>
<td>(0.169)</td>
<td>Vertically Integrated</td>
<td>0.85***</td>
<td>15%***</td>
</tr>
<tr>
<td>N_own</td>
<td>-0.033*</td>
<td>(0.019)</td>
<td></td>
<td>0.90*</td>
<td>10%*</td>
</tr>
<tr>
<td>N_others</td>
<td>-0.006</td>
<td>(0.007)</td>
<td></td>
<td>0.93</td>
<td>7%</td>
</tr>
<tr>
<td>N_entries</td>
<td>0.009***</td>
<td>(0.003)</td>
<td></td>
<td>1.12***</td>
<td>-12%***</td>
</tr>
</tbody>
</table>

Note - Marginal effects are evaluated by changing the value of dummy variables from 0 to 1 and by increasing the value of continuous variables one standard deviation above their means. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

comparison groups considered for each variable when calculating the marginal effects. Finally, the last two columns present the hazard ratios and marginal effects on the probability of survival. Hazard ratios are computed based on equation (3.4), and the effects on the survival probability are calculated by subtracting the HR from one. The hazard rate for herbicide tolerant seed is 56% the hazard rate for conventional seed. This means that survival probability for HT seed is 44% higher than for conventional seed after controlling for other factors. Survival
increases even more for stacked seed (56%) relative to conventional seed. Insect-resistant cottonseed survival does not differ from that of conventional seed.

Vertical integration increases product survival by 25%. An increase of one standard deviation above the mean firm share increases survival by 15% – i.e. products sold by a firm that holds 72% of the market compared to a firm that holds 39% of the market will experience 15% higher survival. If internal product competition increases from 5 products to 9 products, survival increases by 10%, which is the reputation effect mentioned above. If product introductions increase from 39 to 53 nationally, product survival decreases by 12%.29

3.5. Conclusion

This chapter analyzes the determinants of product life for cottonseed sold in the U.S. between 2002 and 2007. I establish that herbicide-tolerant and stacked seed survive longer than conventional and insect-resistant seed after controlling for other factors. This is in line with farmers’ preference estimates from chapter 1 (Table 6.2), where I find that farmers prefer HT and stacked seed to conventional seed, which is not the case for IR seed. Longer survival for stacked cottonseed is also consistent with the finding of subadditive bundle pricing, which (Shi et al., 2011), which states that it is cheaper to buy stacked seed than HT and IR seed separately.

Vertical integration prolongs product life. As mentioned earlier, this could have negative implications on seed availability. If local independent seed firms find themselves at a

29 These results are robust to alternative market definitions. In particular, if I define the market at the CRD level, the only significant difference is that the reputation effects from the own-firm products are no longer important, whereas general trends in biotech seed use captured by N_others become significant (see Table A.2 in the Appendix).
disadvantage, they might decide to invest less at local product differentiation, which would negatively affect the availability of locally-differentiated seed to farmers. The same concern can be extended when interpreting the effect of firm share. Larger firms’ products survive longer, putting smaller firms at a disadvantage and discouraging R&D in local product differentiation.

I find no evidence of internal and external competition effects as measured by the stock of similar products. In fact, my results show that the number of similar products a firm sells increases survival through positive firm reputation effects. Product competition as measured by the flow of products decreases product life. A possible explanation here is that survival is driven by demand factors as farmers are driven to try new products on the market forsaking older products.
Appendix

Table A.1. Average Elasticities by Trait for the Logit Model

<table>
<thead>
<tr>
<th>% Δ in $km</th>
<th>CONV</th>
<th>HT3</th>
<th>HT1</th>
<th>IR2-HT3</th>
<th>IR2-HT2</th>
<th>IR2-HT1</th>
<th>IR1-HT1</th>
<th>IR3</th>
<th>IR1</th>
<th>Own-price Elasticities</th>
<th>Number Products</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONV</td>
<td>0.0001</td>
<td>0.008</td>
<td>0.004</td>
<td>0.015</td>
<td>0.002</td>
<td>0.034</td>
<td>0.001</td>
<td>0.268</td>
<td>0.001</td>
<td>-1.094</td>
<td>2</td>
</tr>
<tr>
<td>HT3</td>
<td>0.0001</td>
<td>0.008</td>
<td>0.004</td>
<td>0.015</td>
<td>0.002</td>
<td>0.034</td>
<td>0.001</td>
<td>0.268</td>
<td>0.001</td>
<td>-1.207</td>
<td>6</td>
</tr>
<tr>
<td>HT1</td>
<td>0.0001</td>
<td>0.008</td>
<td>0.004</td>
<td>0.015</td>
<td>0.002</td>
<td>0.034</td>
<td>0.001</td>
<td>0.268</td>
<td>0.001</td>
<td>-1.576</td>
<td>2</td>
</tr>
<tr>
<td>IR2-HT3</td>
<td>0.0001</td>
<td>0.008</td>
<td>0.004</td>
<td>0.015</td>
<td>0.002</td>
<td>0.034</td>
<td>0.001</td>
<td>0.268</td>
<td>0.001</td>
<td>-3.358</td>
<td>8</td>
</tr>
<tr>
<td>IR2-HT2</td>
<td>0.0001</td>
<td>0.008</td>
<td>0.004</td>
<td>0.015</td>
<td>0.002</td>
<td>0.034</td>
<td>0.001</td>
<td>0.268</td>
<td>0.001</td>
<td>-3.527</td>
<td>1</td>
</tr>
<tr>
<td>IR2-HT1</td>
<td>0.0001</td>
<td>0.008</td>
<td>0.004</td>
<td>0.015</td>
<td>0.002</td>
<td>0.034</td>
<td>0.001</td>
<td>0.268</td>
<td>0.001</td>
<td>-3.268</td>
<td>6</td>
</tr>
<tr>
<td>IR1-HT1</td>
<td>0.0001</td>
<td>0.008</td>
<td>0.004</td>
<td>0.015</td>
<td>0.002</td>
<td>0.034</td>
<td>0.001</td>
<td>0.268</td>
<td>0.001</td>
<td>-3.72</td>
<td>1</td>
</tr>
<tr>
<td>IR3</td>
<td>0.0001</td>
<td>0.008</td>
<td>0.004</td>
<td>0.015</td>
<td>0.002</td>
<td>0.034</td>
<td>0.001</td>
<td>0.268</td>
<td>0.001</td>
<td>-1.127</td>
<td>1</td>
</tr>
<tr>
<td>IR1</td>
<td>0.0001</td>
<td>0.008</td>
<td>0.004</td>
<td>0.015</td>
<td>0.002</td>
<td>0.034</td>
<td>0.001</td>
<td>0.268</td>
<td>0.001</td>
<td>-2.861</td>
<td>1</td>
</tr>
<tr>
<td>Share</td>
<td>0.0002</td>
<td>0.004</td>
<td>0.002</td>
<td>0.004</td>
<td>0.001</td>
<td>0.01</td>
<td>0.0003</td>
<td>0.143</td>
<td>0.001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note - These elasticities are the averages for a given market (Arizona in 2006).

The logit demand elasticities suggest that farmers would always substitute towards the most popular products in the market ignoring similarities in product attributes. These unrealistic substitution patterns are reflected by the fact that cross elasticities are the same across each trait group and that they are largest w.r.t. IR3 seed for all trait groups.
Table. A.2. Marginal Effect on Survival Probability by Market Definition

<table>
<thead>
<tr>
<th>Variables</th>
<th>Benchmark (D = 0, X = mean)</th>
<th>Change (D = 1, X = (mean + SD))</th>
<th>CRD</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Biotech Characteristics (benchmark is conventional)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.traitcode</td>
<td>Conventional</td>
<td>HT</td>
<td>38%***</td>
<td>44%***</td>
</tr>
<tr>
<td>3.traitcode</td>
<td>Conventional</td>
<td>IR</td>
<td>-12%</td>
<td>4%</td>
</tr>
<tr>
<td>4.traitcode</td>
<td>Conventional</td>
<td>Stacked</td>
<td>49%***</td>
<td>56%***</td>
</tr>
<tr>
<td><strong>Market Structure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integration</td>
<td>Not Integrated</td>
<td>Vertically Integrated</td>
<td>23%***</td>
<td>25%***</td>
</tr>
<tr>
<td>Firm Share</td>
<td>Mean</td>
<td>Mean + SD</td>
<td>18%***</td>
<td>15%***</td>
</tr>
<tr>
<td>N_own</td>
<td>Mean</td>
<td>Mean + SD</td>
<td>4%</td>
<td>10%*</td>
</tr>
<tr>
<td>N_others</td>
<td>Mean</td>
<td>Mean + SD</td>
<td>12%***</td>
<td>7%</td>
</tr>
<tr>
<td>N_entries</td>
<td>Mean</td>
<td>Mean + SD</td>
<td>-13%***</td>
<td>-12%***</td>
</tr>
<tr>
<td>Intro-year Values</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
<td>CRD</td>
<td>State</td>
</tr>
<tr>
<td>Observations</td>
<td></td>
<td></td>
<td>2,486</td>
<td>1,497</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td></td>
<td></td>
<td>-4939</td>
<td>-2501</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses

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


