

An Analysis of Selectivity in the Productivity Evaluation of Biotechnology:

An Application to Corn

by

Guanming Shi

Jean-Paul Chavas

Joseph Lauer

Elizabeth Nolan¹

Abstract: This paper investigates selectivity bias issues arising in the productivity evaluation of biotech hybrids, with an application to GM corn yield. The analysis is applied to experimental data on Wisconsin corn yields covering the period 1990-2010. Relying on a “Heckman-like” factor that accounts for selectivity, the analysis uncovers evidence of selection bias. It indicates that some of the observed yield advantage associated with GM hybrids is contributed by their conventional genes. This stresses that, even after the widespread adoption of GM corn in the US, traditional breeding still plays an important role in productivity improvements for corn. We also document how rising market concentration of biotech firms has contributed to increasing selectivity bias in corn yield. Our results suggest that current patenting laws may not appropriately differentiate between GM genes and conventional genes.

Keywords: biotechnology, corn, genetic modification, productivity, selectivity bias.

JEL: D2, L1, O3, Q1

¹ Guanming Shi is Assistant Professor, Department of Agricultural and Applied Economics, University of Wisconsin, Madison, WI 53706; Tel: (608) 263-6250; Email: gshi@wisc.edu. Jean-Paul Chavas is Professor, Department of Agricultural and Applied Economics, University of Wisconsin, Madison, WI 53706; Email: jchavas@wisc.edu. Joe Lauer is Professor, Department of Agronomy, University of Wisconsin, Madison, WI 53706; Email: jglauer@wisc.edu. Elizabeth Nolan is Lecturer, Department of Agricultural and Resource Economics, University of Sydney, NSW 2006 Australia; Email: elizabeth.nolan@sydney.edu.au. This research was supported by a Hatch Grant #142-PRJ37WQ from the College of Agricultural and Life Sciences, University of Wisconsin, and by a USDA/NIFA grant #2010-04833.

1. Introduction

Increases in corn productivity have been very large over the last few decades. In the US, corn yield has increased from 28.9 bushels per acre in 1940 to 72.4 bushels per acre in 1970, and then to 152.8 bushels per acre in 2010 (USDA, 2011). Genetic improvements have been major contributing factors (Duvick, 1992, 2005). Two innovations in corn breeding have played a key role. The first was the introduction of hybrid corn in the 1930s. Corn hybrids benefited from heterosis (or hybrid vigor) that generated large gains in corn productivity (Griliches, 1957, 1960; Fernandez-Cornejo 2004; Springer and Stupar 2007). The widespread adoption of corn hybrids from the 1930s also led to a profitable corn seed industry, in which private firms dominate production and distribution of corn seed (Fernandez-Cornejo 2004).

The second innovation was the introduction, in the 1990s, of genetically modified (GM) traits embedded in seeds, following the development of gene transfer technology. Adoption of GM corn hybrids was rapid in the US. In 2011, 88% of the corn acreage in the US was planted in hybrids with at least one GM trait, and 49% of corn acreage was planted to hybrids with at least three GM traits (USDA Economics Research Service 2011). The GM hybrids are typically patented by biotech firms.¹ The development of biotechnology was also associated with mergers and acquisition activity as firms who owned GM traits acquired firms who owned basic hybrid germplasm and distribution networks. Mergers in the seed industry were particularly active towards the end of the 1990s and in the early 2000s (Alston and Venner 2002; Wright and Pardey 2006).

The trend to consolidation in the seed industry was driven by a number of factors: exploiting asset complementarities, mitigating contractual hazards, and/or seeking market power (Rausser, Scotchmer and Simon 1999; Graff, Rausser and Small 2003; Shi 2009). The presence

of the high fixed cost and low variable costs associated with biotechnology indicates a need to price above marginal cost to support high levels of investment (Wright and Pardey 2006). Consolidation has also been driven by the need for access to elite germplasm (Wright and Pardey 2006): New and valuable traits cannot generate value unless they are incorporated in commercial hybrid lines that are produced and sold to farmers. Since the marginal cost of incorporating a trait into a hybrid is small, the earnings from licensing the trait are almost entirely a function of market size, as a large marketing network complements a portfolio of traits (Rausser, Scotchmer and Simon 1999). This fact partly explains the expansion of the large firms into retailing through the purchase of regional seed companies. By 1995, a number of producers of crop protection chemicals had begun to integrate into the retail seed market (King and Schimmelpfennig 2005; Wright and Pardey 2006).

The US corn seed market has seen a rise in market concentration over the last few decades. While the four-firm concentration ratio (CR4) was 60 percent in 1983 and 69 per cent in 1997 (Fernandez-Cornejo 2004), the corn seed industry is now dominated by four large biotech firms, all with subsidiary seed companies (Fernandez-Cornejo 2004; Shi, Chavas and Stiegert 2010). They are: DuPont (Pioneer International), Monsanto, Syngenta and Dow AgroSciences.² The market share of the top 4 biotech firms in the US corn seed market has increased over the last 15 years and reached 86 percent in 2010.

Increased market concentration can raise concerns about adverse effects of imperfectly competitive pricing (Fulton and Giannakas 2001; Fernandez-Cornejo 2004; Shi, Chavas and Stiegert 2010). However, there may be other effects of increased concentration. While the technologies associated with GM hybrids are concentrated, the four main players have also acquired a large share of the elite basic inbred germplasm. Concentrated seed markets may be

associated with greater control over the quality of the germplasm, indicating that high concentrations in seed markets could possibly contribute to an increase in selectivity in the productivity evaluation of GM hybrids (as discussed below).

Both biotech genes and conventional genes change over time. The genetic selection of conventional genes has been a major source of agricultural productivity growth over the centuries (Duvick 1992, 2005; Alston and Pardey 1996). The use of gene-transfer technology offers good prospects for additional productivity growth in agriculture (Herdt 2006; Bouis 2007; Qaim 2009; James 2010; National Academies 2010; Ronald 2011). The biotech genes are often patented by a limited number of biotech firms and thus observable via trademarks of the biotech hybrids. In contrast, there are typically a very large number of conventional genes that evolve over time. This raises the question as to how one can assess the separate effects of GM genes versus the effects of the conventional germplasm on agricultural productivity. Not being able to observe the many conventional genes present in both conventional and GM hybrids indicates the possibility of facing selectivity issues in the productivity evaluation of GM traits.

Selection bias can arise when data are generated in a non-random fashion. Examples include self-selection, pre-screening and the general evaluation of a sub-sample that excludes some members of the population of interest. This has been found relevant in the analysis of labor, migration and marketing decisions (e.g., Heckman 1979; Goetz 1992).³ In these cases, selectivity arises when individual performance is evaluated based on a sub-sample that is not representative of the population. We argue below that selectivity can arise in the productivity analysis of biotech hybrids. For example, this would occur if companies were to incorporate GM genes only in high-quality germplasm.⁴ Under such circumstances, the high performance of biotech hybrids would be due in part to the high quality of their conventional genes. To our

knowledge, the role of selectivity in the productivity analysis of GM seeds has not been investigated before. Addressing this issue is the major objective of this paper.

In this study, we investigate the potential selectivity bias associated with biotech hybrids, with an application to GM corn yield. We use a “Heckman-like” factor to correct for selectivity bias. This provides a basis to test whether selection bias is present. We also examine factors contributing to selection bias in GM corn yield. In particular, we investigate whether the rising market dominance of biotech firms may have influenced this potential selection bias. The analysis is applied to experimental data on Wisconsin corn yields covering the period 1990-2010.

Our analysis of GM corn yield uncovers evidence of selection bias. Our findings indicate that some of the observed yield gain associated with GM hybrids comes from their conventional genes. This emphasizes that even after the widespread adoption of GM corn in the US, traditional breeding still plays an important role in productivity improvements. Our analysis also finds that selectivity effects vary with the GM genes. We document how rising market concentration of biotech firms has contributed to increasing selection bias in corn yield. Implications of our findings are discussed.

2. Evaluating selectivity bias in genetic changes

Agricultural production involves three sets of inputs: biological organisms and their genetics \mathbf{G} ; environmental inputs \mathbf{V} ; and managerial inputs \mathbf{M} . The technology is represented by the following production function

$$y = f(\mathbf{G}, \mathbf{V}, \mathbf{M}), \tag{1}$$

where y denotes agricultural output. The genetic inputs include two types of genes, $\mathbf{G} = (\mathbf{G}_b, \mathbf{G}_c)$, where \mathbf{G}_b are “biotech genes” introduced using gene-transfer biotechnology, and \mathbf{G}_c denotes “conventional genes” that are part of the traditional breeding germplasm. The environmental

effects, \mathbf{V} , include agro-climatic conditions, weather effects (e.g., rainfall, temperature) and pest population. Finally, the management effects, \mathbf{M} , include crop rotation, fertilizer/feed use, pest/disease management, etc. Note that (1) represents a generic agricultural production technology, allowing for possible interactions among genes, and between genotype, environmental conditions and management factors.

Equation (1) provides the information needed to evaluate productivity. For example, consider a genetic change from \mathbf{G}^t to $\mathbf{G}^{t'}$, between time t and time t' . Then, conditional on (\mathbf{V}, \mathbf{M}) , the associated change in productivity can be measured by

$$\Delta y(\mathbf{G}^{t'}, \mathbf{G}^t; \mathbf{V}, \mathbf{M}) = f(\mathbf{G}^{t'}, \mathbf{V}, \mathbf{M}) - f(\mathbf{G}^t, \mathbf{V}, \mathbf{M}). \quad (2)$$

Note that $\Delta y(\cdot)$ in (2) is the marginal productivity effect (measured in terms of agricultural output) associated with a genetic change from \mathbf{G}^t to $\mathbf{G}^{t'}$. Conditional on (\mathbf{V}, \mathbf{M}) , having $\Delta y(\cdot) > 0$ (< 0) in (2) means that switching from genes \mathbf{G}^t to genes $\mathbf{G}^{t'}$ increases (decreases) production by $\Delta y(\cdot)$ units of output. Such results apply to general changes in biotech genes \mathbf{G}_b as well as conventional genes \mathbf{G}_c .

The biotech genes \mathbf{G}_b are typically few in number and each can be identified by patents and trademarks held by the biotech firm that generated them. In contrast, conventional genes \mathbf{G}_c are numerous and evolve over time in complex ways depending on natural selection and/or genetic selection implemented by farmers and plant breeders. As a result, the majority of the vast number of \mathbf{G}_c are unobserved, which makes it more difficult to evaluate their effects on agricultural productivity. Yet, the evolving effects of \mathbf{G}_c are of significant interest. Indeed, since the beginning of agriculture some 10,000 years ago, conventional genes have been the subject of intense genetic selection by farmers and, more recently, by plant breeders. This has been a major

source of agricultural productivity growth over the centuries (Alston and Pardey 1996; Duvick 1992). This should be recognized in any study of the effects of the introduction of GM traits.

Conditional on \mathbf{G}_b and in a given region denote the population distribution of \mathbf{G}_c at time t by $\Gamma_t(\cdot | \mathbf{G}_b)$. Then from (1) and conditional on $(\mathbf{G}_b, \mathbf{V}, \mathbf{M})$, average production at time t is given by

$$g_t(\mathbf{G}_b, \mathbf{V}, \mathbf{M}, \boldsymbol{\beta}) \equiv E_t[f(\mathbf{G}_b, \mathbf{G}_c^t, \mathbf{V}, \mathbf{M})] = \int_{\mathbf{G}_c^t} f(\mathbf{G}_b, \mathbf{G}_c^t, \mathbf{V}, \mathbf{M}) d\Gamma_t(\mathbf{G}_c^t | \mathbf{G}_b), \quad (3)$$

where E_t is the expectation operator based on the conditional distribution function $\Gamma_t(\cdot | \mathbf{G}_b)$ and $\boldsymbol{\beta}$ is a vector of parameters capturing the effects of $(\mathbf{G}_b, \mathbf{V}, \mathbf{M})$ on mean output.

Since the development of biotechnology, we now face situations where both biotech genes \mathbf{G}_b and conventional genes \mathbf{G}_c change over time. As noted in the introduction, the use of gene-transfer technology offers good prospects for additional productivity growth in agriculture. However, this also raises the question: what is the relative contribution of biotech genes versus conventional genes to agricultural productivity improvements? When conventional genes are numerous and difficult to measure, the potential for selection bias arises. This could occur if the biotech seed companies add biotech genes, \mathbf{G}_b , only to “high quality” conventional genes, \mathbf{G}_c , in order to generate superior productivity of the biotech seeds.

To analyze possible selectivity in the productivity evaluation of biotech genes, we start with the average production given in (3). It follows from (3) that equation (1) can be written as

$$y = g(\mathbf{G}_b, \mathbf{V}, \mathbf{M}, \boldsymbol{\beta}) + e, \quad (4)$$

where e is an error term satisfying $E_t(e) = 0$. Equation (4) is a regression model. The error term e in (4) measures the deviation from mean production reflecting heterogeneity in unobservable conventional genes, \mathbf{G}_c . As such, the distribution of e provides an indirect measurement of the distribution of conventional genes, \mathbf{G}_c .

In principle, once specified, equation (4) can be estimated using observations on production output y , and $(\mathbf{G}_b, \mathbf{V}, \mathbf{M})$. For given (\mathbf{V}, \mathbf{M}) , this can give useful information on how alternative biotech genes \mathbf{G}_b affect productivity. Yet, one issue can arise in equation (4): is the distribution of e independent of \mathbf{G}_b ? If yes, then standard least-squares estimation of (4) would provide an unbiased estimate of the productivity effects of biotech genes, \mathbf{G}_b .

However, this condition (e being independent of \mathbf{G}_b) would fail to hold in the presence of genetic selectivity. Let $\mathbf{G}'_b = 0$ in the absence of biotech genes. Conditional on $(\mathbf{V}_t, \mathbf{M}_t)$, define $e_t(0) \equiv y_t - g(0, \mathbf{V}_t, \mathbf{M}_t, \boldsymbol{\beta})$, which captures the heterogeneity in basic germplasm. Denote the distribution of $e_t(0)$ by $H_t(a|0) = \text{Prob}[e_t(0) \leq a]$. Next, consider situations where biotech genes, \mathbf{G}_b , are added to the basic germplasm. This can be done in many ways. To address selectivity issues, we consider the case where biotech genes, \mathbf{G}_b , may be bundled together with “high quality” conventional genes in \mathbf{G}_c . Let $m \geq 0$ be a threshold of germplasm quality satisfying $H_t(m|0) \in (0, 1)$, where $e_t \geq m$ corresponds to “high quality” germplasm while $e_t < m$ corresponds to “low quality” germplasm. Let S_t be a scalar between 0 and 1, $S_t \in [0, 1]$. In the presence of biotech genes $\mathbf{G}'_b \neq 0$, assume that the distribution of $e_t(\mathbf{G}'_b)$ is given by

$$\begin{aligned} H_t(a|\mathbf{G}'_b) &= [1 - S_t]H_t(a|0) \text{ if } a < m, \\ &= -\alpha_t + [1 + \alpha_t]H_t(a|0) \text{ if } a \geq m. \end{aligned} \tag{5}$$

where $\alpha_t \equiv S_t H_t(m|0) / [1 - H_t(m|0)]$. The parameter S_t in (5) acts as a selectivity index.

When $S_t = 0$, it follows from (5) that $H_t(\cdot|\mathbf{G}'_b) = H_t(\cdot|0)$. Then, the distribution of basic germplasm quality is the same between conventional and biotech seeds. In this case, $E[e_t(\mathbf{G}'_b)] = 0$ and there is no selectivity bias in equation (4). However, when $S_t > 0$, there is selectivity bias as equation (5) implies that $H_t(a|0)$ tends to be larger than $H_t(a|\mathbf{G}'_b)$. This is illustrated in

Figure 1, where the distribution functions $H_t(a|0)$ and $H_t(a|\mathbf{G}'_b)$, and the associated probability density functions $h_t(a|0)$ and $h_t(a|\mathbf{G}'_b)$ are shown in (1a) and (1b) respectively. In Figure 1b, the effects of S_t on $h_t(a|\cdot)$ can be decomposed into two steps. In a first step, the line AC is shifted down as the probabilities of facing any event $a \in [-\infty, m]$ are rescaled by a factor $(1 - S_t) \in [0, 1]$. When $S_t > 0$, this means a reduction in the probability that biotech seeds would have “low quality” germplasm. In a second step, all probabilities are shifted up proportionally to keep the area below the line ($AC'C''D$) equal to 1 (so that area ACC' equals the area $CC''D$ and the probabilities sum up to 1). As illustrated in Figure 1a, having $S_t > 0$ means a reduction in the probability of facing “low quality” germplasm and a rightward shift in the distribution function from $H_t(a|0)$ to $H_t(a|\mathbf{G}'_b)$.

In the extreme case where $S_t = 1$, equation (5) would imply that $H_t(a|\mathbf{G}'_b) = 0$ when $a < m$, i.e. that the distribution function $H_t(a|\mathbf{G}'_b)$ becomes truncated below the threshold point m . This corresponds to a case of extreme selection where no low-quality germplasm is used in biotech seeds. In intermediate situations where $S_t \in (0, 1)$, equation (5) allows for “partial selectivity”, the intensity of selection increasing with S_t .

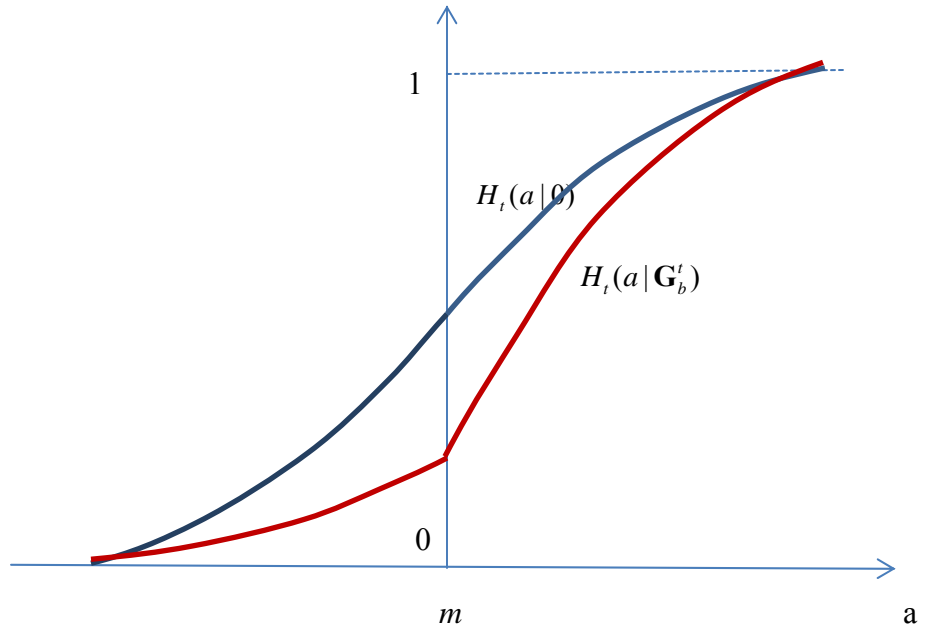
Given \mathbf{G}'_b , the expected value of output y_t at time t is

$$\begin{aligned}
E_t(y_t | \mathbf{G}'_b) &= g(\mathbf{G}'_b, \mathbf{V}_t, \mathbf{M}_t, \boldsymbol{\beta}) + E_t[e_t(\mathbf{G}'_b)] \\
&= g(\mathbf{G}'_b, \mathbf{V}_t, \mathbf{M}_t, \boldsymbol{\beta}) + \int_e e dH_t(e | \mathbf{G}'_b), \text{ or using (5),} \\
&= g(\mathbf{G}'_b, \mathbf{V}_t, \mathbf{M}_t, \boldsymbol{\beta}) + [1 - S_t] \int_{e < m} e dH_t(e | 0) + [1 - S_t + S_t + \alpha_t] \int_{e \geq m} e dH_t(e | 0), \\
&= g(\mathbf{G}'_b, \mathbf{V}_t, \mathbf{M}_t, \boldsymbol{\beta}) + K_t(m, S_t) \int_{e \geq m} e dH_t(e | 0), \text{ since } e_t(0) \text{ has mean zero,} \tag{6}
\end{aligned}$$

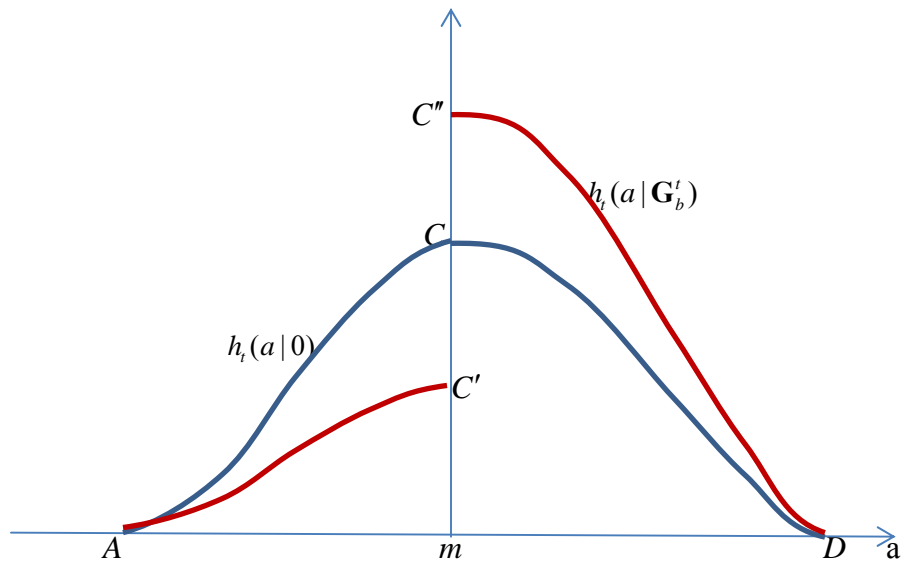
where

$$K_t(m, S_t) \equiv S_t + \alpha_t = S_t / [1 - H_t(m | 0)] \geq 0. \tag{7}$$

Figure 1. An illustration of selectivity effects.



(1a): Distribution functions $H_t(a|0)$ and $H_t(a|G'_b)$



(1b): Probability density functions $h_t(a|0)$ and $h_t(a|G'_b)$

Note that $K_t(m, S_t)$ is proportional to S_t and satisfies $K_t(m, 0) = 0$. Also, $[\int_{e \geq m} e dH_t(e|0)] > 0$. Then, equation (6) implies that $E(y_t) \geq g(\mathbf{G}'_b, \mathbf{V}_t, \mathbf{M}_t, \boldsymbol{\beta})$ in general, with $E(y_t) = g(\mathbf{G}'_b, \mathbf{V}_t, \mathbf{M}_t, \boldsymbol{\beta})$ when $S_t = 0$ and $E(y_t) > g(\mathbf{G}'_b, \mathbf{V}_t, \mathbf{M}_t, \boldsymbol{\beta})$ when $S_t \in (0, 1]$. This identifies

$$D_t \equiv K_t(m, S_t) \int_{e \geq m} e dH_t(e|0) \geq 0, \quad (8)$$

as the selectivity bias. Equation (8) gives the general and intuitive result that selecting high quality basic germplasm increases the average productivity of biotech seeds. This result applies under any distribution function $H_t(e|0)$.

In the special case where $e_t(0)$ has a normal distribution with mean zero and variance σ_t^2 , we can obtain more specific results. Let $\phi(\cdot)$ and $\Phi(\cdot)$ denote the density function and the distribution function, respectively, for a standard normal random variable $N(0, 1)$. Then, $\int_{e \geq m} e dH_t(e|0) = \{\phi(m/\sigma_t)/[1 - \Phi(m/\sigma_t)]\}\sigma_t$ (Johnson and Kotz, 1970, pp. 81-83), and the selectivity bias in (8) becomes:

$$D_t \equiv K_t(m, S_t) \{\phi(m/\sigma_t)/[1 - \Phi(m/\sigma_t)]\}\sigma_t. \quad (8')$$

Note that the term $\{\phi(m/\sigma_t)/[1 - \Phi(m/\sigma_t)]\}$ in (8') is the inverse Mills ratio, which is commonly used in the analysis of selectivity bias under normality (e.g., Heckman 1979). This shows that the selectivity bias is proportional to $K_t(m, S_t)$, to the inverse Mills ratio, and to the standard deviation σ_t .

Combining (6), (7) and (8') gives

$$y_t = g(0, \mathbf{V}_t, \mathbf{M}_t, \boldsymbol{\beta}) + e_t, \text{ when } \mathbf{G}'_b = 0, \quad (9a)$$

$$= g(\mathbf{G}'_b, \mathbf{V}_t, \mathbf{M}_t, \boldsymbol{\beta}) + S_t [1 - H_t(m|0)]^{-1} \{\phi(m/\sigma_t)/[1 - \Phi(m/\sigma_t)]\}\sigma_t + u_t, \text{ when } \mathbf{G}'_b \neq 0, \quad (9b)$$

where u_t is an error term with mean zero. Equation (9) provides a basis to evaluate empirically the presence of gene selectivity involving both traditional genes and biotech genes.

What factors would lead to genetic selectivity? As discussed above, selectivity would be absent if biotech genes are inserted at random in the current germplasm. Alternatively, selectivity would arise if only the higher quality germplasm is chosen for the production of GM seeds. It means that selectivity would be associated with genetic selection that imposes some level of control on the distribution of germplasm and its quality. This suggests that the selectivity parameter S_t in (9) can vary depending on the situation considered.

To investigate this issue, we consider the parameterization $S_t \equiv \mathbf{z}_t \boldsymbol{\gamma}$, where \mathbf{z}_t is a vector of variables presumed to affect selectivity and $\boldsymbol{\gamma}$ is a vector of parameters.⁵ Then, equations (9a)-(9b) provide a basis to estimate the parameters $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ and to test the null hypothesis of no selectivity bias (when $\boldsymbol{\gamma} = 0$). In the presence of selectivity bias (when $\boldsymbol{\gamma} \neq 0$), equation (9b) allows the investigation and measurement of the effects of genetic selectivity on productivity assessment. Note that asserting greater control over the quality of the germplasm may be associated with more concentrated seed markets. This suggests that increased concentrations in seed markets could possibly contribute to an increase in genetic selectivity. Equation (9) (with $S_t \equiv \mathbf{z}_t \boldsymbol{\gamma}$) will allow us to investigate whether increasing market dominance of biotech companies in the seed market have contributed to genetic selectivity.

3. Data

To investigate the potential for genetic selectivity, we use data on corn yield obtained from field experiments conducted over the period 1990 - 2010 at 12 University of Wisconsin Agricultural Research Stations. In these fields, management practices were typical of those utilized on Corn Belt farms practicing rainfed agriculture, including planting density, spring/fall tillage, fertilizer/herbicide/insecticide applications and irrigation. A total of 4748 hybrids have been tested in the past 21 years. Of these, 2653 are conventional hybrids and 2095 are GM

hybrids. Some hybrids are tested in multiple sites and/or for multiple years, yielding 31799 usable observations for the analysis.

There are two major groups of GM traits in the corn hybrid market: those related to insect resistance (IR) and those related to herbicide tolerance (HT). The IR traits are designed to control specific pest populations, thus reducing corn yield losses due to pest damages. Two IR traits are identified: those controlling the European Corn Borer (ECB), and those controlling corn rootworms (RW). The HT traits are designed to make it easier to control weed infestations. Two HT traits are identified: those related to glyphosate tolerance (GT), and those related to glufosinate, often labeled as Liberty Link (LL). Altogether these traits are embedded in a total of 12 different types of GM hybrids: four single-trait hybrids (ECB, GT, RW, and LL), four double stack hybrids (ECB/GT, ECB/RW, ECB/LL, and GT/RW), three triple stack hybrids (ECB/RW/GT, ECB/GT/LL, and ECB/RW/LL) and one quadruple stack hybrid (ECB/RW/GT/LL). In addition to detailed information on management practices and yield, the data also contain specific “event” information on the patented gene(s) included in the GM hybrids, as well as when the event was first introduced and commercialized.⁶

To investigate the potential selectivity bias associated with increasing market dominance of the biotech companies in the seed market, we use the aggregate seed market share of the four biotech companies involved in the corn seed market. These companies are: Monsanto, DuPont (via acquisition of Pioneer Hi-Bred Int’l in 1999), Dow AgroSciences, and Syngenta. We constructed the aggregate market share of these four companies for 2000 to 2007 using acreage data collected by **dmr**kynetec [hereafter DMR], St. Louis, MO. Prior to the commercialization of biotech corn hybrids in 1996, none of these four firms was involved in corn seed production. Thus, we set their market share at zero prior to 1997. For 1997 to 1999, we use market share

statistics reported in Fernandez-Cornejo (2004) for 1998 and assume that these numbers are the same for 1997 and 1998 (given that most acquisitions occur either in 1997 or after 1999). For the more recent years, we obtained the information from the media reports on these firms.

Figure 2 shows the evolution of market concentration as measured by the market share of the four biotech firms over the last 21 years. DuPont’s acquisition of Pioneer in 1999 increased the four firms’ aggregate share from about 20% to over 50% in 2000. By 2010, these four firms supplied about 86% of the US corn seed market. Much of the expansion was through mergers and acquisitions of local seed companies by these vertically integrated biotech firms.

Figure 2. Market share of the four biotech firms, 1990 – 2010.

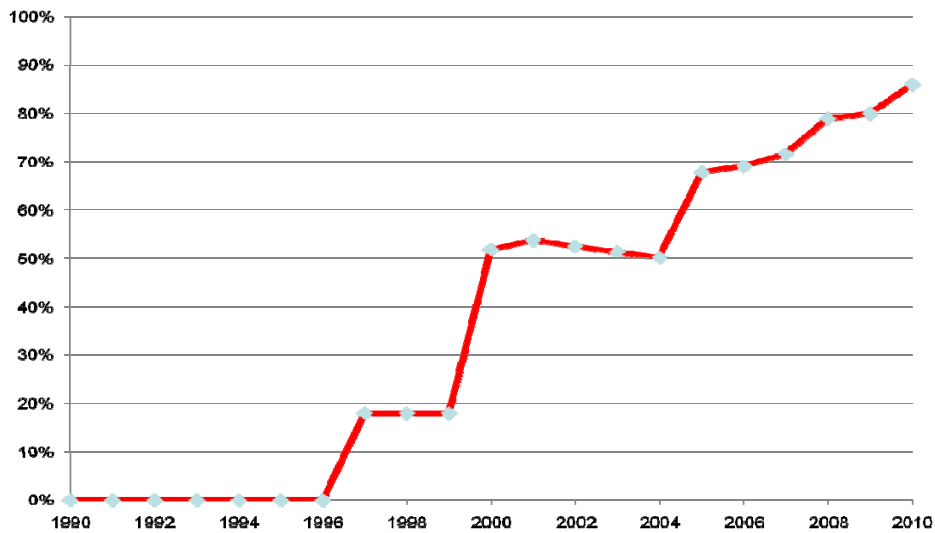


Table 1 presents summary statistics of the major variables used in the empirical analysis. On average, GM hybrids tend to have a higher yield than conventional hybrids, and most stack hybrids have a higher yield than the single-trait hybrids. However, as argued above, these productivity differences could be due in part to selectivity (if GM traits have come to be associated with “better” germplasm). For GM hybrids, the data also include the number of years since the corresponding event was first introduced, named as “event lag”. These variables are

Table 1. Summary statistics.

Variable	Number of observations	Mean	Standard Deviation	Min.	Max.
<i>Yield (bushels/acre)</i>					
Conventional	19652	166.6	37.9	21	285.2
Glyphosate Tolerant (GT) single	972	182	37.9	56.2	276
Liberty Link (LL) single	103	189.8	31.4	104.8	258
European Corn Borer (ECB) single	3484	197.6	36	45.7	287.8
Root worm (RW) single	36	185.1	28.5	125.3	252.4
ECB/RW double	85	210.5	26.8	130.3	264.3
ECB/GT double	1454	191.2	35.6	74.2	280.8
GT/RW double	166	204.6	35.6	109.9	268.7
ECB/LL double	998	197.6	39.9	65.7	285.5
ECB/RW/GT triple	3215	202.8	31	95	288
ECB/GT/LL triple	631	201.1	36.4	98.3	283.9
ECB/RW/LL triple	206	209.1	31.7	104.8	285.1
ECB/GT/RW/LL quad.	797	208.5	33.2	78.5	289.8
<i>Number of years since event introduction</i>					
Glyphosate Tolerant (GT) single	972	7.1	2.8	1	13
Liberty Link (LL) single	103	6.7	4.7	1	14
European Corn Borer (ECB) single	3484	7.2	2.1	1	14
Root worm (RW) single	36	1.8	0.7	1	4
ECB/RW double	85	2	0.9	1	4
ECB/GT double	1454	7.2	20.7	1	12
GT/RW double	166	3.3	1.2	1	5
ECB/LL double	998	7.9	3.7	1	15
ECB/RW/GT triple	3215	2.6	1	1	6
ECB/GT/LL triple	631	3.9	1.1	1	5
ECB/RW/LL triple	206	4.2	2.2	1	8
ECB/GT/RW/LL quad.	797	3	1.3	1	5
Fertilizer (lbs/acre)	31799	130.2	47.3	0.5	236.3
Planting density (1000 seeds/acre)	31799	28.5	1.9	18.3	33.4
Four firms' annual market share ^a	14	0.55	0.23	0.18	0.86

^a Corn seed market share of these firms is zero prior to 1997. The summary statistics for this variable is for 1997 – 2010 only.

intended to capture possible interaction effects between traditional genes and biotech genes as they affect corn yield. Note that these interactions are subject to management. That is, they depend on where biotech genes are inserted in the germplasm. In this context, the “event lag” variables can affect yield. For example, they could have positive effects on yield if geneticists

and plant breeders manage to reduce over time negative interaction effects between particular events and traditional genes. In 2010, note that all single trait events except for the rootworm event, and two double stack events (ECB/LL, GT/LL) have been on the market for at least 10 years. Triple stack and quadruple stack hybrids entered the market only in recent years, and thus have less number of event years than most single and double stack events.

4. Econometric analysis

We use equations (9a) and (9b) to analyze and test hypotheses about genetic selectivity. Included among the explanatory variables for this analysis are the biotech traits \mathbf{G}_b reported in table 1. Also included are the following management variables, \mathbf{M} : crop rotation, planting density, fertilizer use, irrigation, insecticide use, and spring/fall tillage. Finally, the analysis controls for the environmental variables, \mathbf{V} , by including dummy variables for each location, and dummy variables for each year, as well as their interactions. These dummy variables capture agro-climatic conditions and the effects of weather and pest populations (which can vary across location as well as across years). Controlling for these effects is important in ensuring that the error term e_t in (4) or (9) does represent only the heterogeneity of the basic germplasm (and not weather effects or pest population effects).

A word of caution is needed about our model and its interpretation. We focus our attention on evaluating selectivity issues in the germplasm used in GM hybrids. In this context equations (9a)-(9b) provide a proper basis for our investigation. As just noted, a key argument is that the error term e_t in (9) represents the unobserved distribution of the germplasm. It means that weather and pest population effects are captured through the time dummies, the location dummies and their interactions. However, the effects of some GM traits (e.g., ECB or RW) are to control pest populations. In this case, our estimated model provides only partial measures of the

effects of GM traits on corn yield (as some of these effects being captured by the time/location dummies). It means that, while our econometric analysis gives a proper basis to evaluate selectivity effects related to GM hybrids, it should not be used to evaluate the productivity effects of GM traits. This should be kept in mind in the interpretation of our econometric results.

Each biotech hybrid includes at least one genetic trait, all being represented by corresponding dummy variable(s). In the presence of stacking, biotech hybrids would include more than one GM traits. To capture stack effects, we also include corresponding stack dummies. Therefore, biotech genes, \mathbf{G}_b , in equation (9a)-(9b) are captured by both trait dummies and stack dummies when applicable.⁷ Additionally, and as noted above, we also include “event lag” variables measuring the number of years since a particular event was first introduced in the market. Such variables were introduced in linear form, but also in quadratic form for those events with more than 10 years of market presence (to capture potential curvature effects).

We estimated equations (9a) and (9b) by ordinary least squares with robust errors, the term σ_t in (9b) being estimated as the standard deviation of the error term obtained from (9a). The analysis is conducted setting the threshold level m to zero ($m = 0$). This corresponds to a below-the-mean threshold for selectivity in germplasm. Given $S_t \equiv \mathbf{z}_t \boldsymbol{\gamma}$, it means that equation (9b) takes the form

$$Yield_t = g(\mathbf{G}_b^t, \mathbf{V}_t, \mathbf{M}_t, \boldsymbol{\beta}) + \mathbf{z}_t \boldsymbol{\gamma} \cdot \frac{4\sigma_t}{\sqrt{2\pi}} + u_t, \text{ when } \mathbf{G}_b^t \neq 0, \quad (9b')$$

where $[\mathbf{z}_t \boldsymbol{\gamma} \cdot \frac{4\sigma_t}{\sqrt{2\pi}}]$ is a Heckman-like factor correcting for selectivity bias. Thus, the null hypothesis of no selectivity bias corresponds to $\boldsymbol{\gamma} = 0$. Alternatively, finding evidence that $\boldsymbol{\gamma} \neq 0$ would document the presence of selectivity bias. In this latter case, (9b') provides measurement of such effects on productivity assessment.

We consider two specifications for (\mathbf{z}_t, γ) in (9b'). In the first specification, \mathbf{z}_t includes two variables: a constant and the market share of the four biotech firms. The constant coefficient allows the investigation of whether or not there exists selection bias associated with all GM hybrids compared to conventional hybrids, and the coefficient of the market share variable permits an analysis of whether selectivity bias may vary with changing market concentration. This first specification restricts the selectivity bias to be the same across all GM traits. This restriction is relaxed in the second specification, where we allow the selection bias to vary by 9 GM hybrid types: single trait ECB, single trait GT, and the stacked hybrids.⁸ As in the first specification, the second specification also includes the market share of the four biotech firms to capture possible effects of market concentration on selectivity.

We estimate equations (9a) and (9b') jointly, however, the selectivity term $\mathbf{z}_t \gamma \cdot \frac{4\sigma_t}{\sqrt{2\pi}}$ in (9b') is relevant only when $\mathbf{G}'_b \neq 0$. Practically, it means that the variables z_t in (9b') are specified as interaction variables with a dummy variable GM , where $GM = 0$ if $\mathbf{G}'_b = 0$ and $GM = 1$ if $\mathbf{G}'_b \neq 0$. Table 2 reports the estimation results from equations (9a) and (9b') under both specifications.

In both specifications, we found strong statistical evidence against the null hypothesis of no selectivity bias. In specification 1, the coefficient of the GM dummy is 0.32. It is statistically significant at the 1 percent level, showing the presence of selectivity. In the situation where the market share variable is zero, it would mean that only 68 percent of low-quality germplasm is being used in GM hybrids. In addition, the coefficient of the market share for the four biotech firms is positive and statistically significant at the 1 percent level. This indicates that increased market concentration contributes to increasing selectivity bias. It suggests that, as they get larger, biotech firms tend to insert GM traits into better quality germplasm.

Table 2. Estimation of yield effects and selectivity bias using robust errors.^a

Variables	Specification 1		Specification 2		
	Coefficient	t-Statistics	Coefficient	t-Statistics	
γ	$\frac{4\sigma_t}{\sqrt{2\pi}} \cdot GM$	0.32***	2.61	-	-
	$\frac{4\sigma_t}{\sqrt{2\pi}} \cdot ECB \text{ single}$	-	-	0.54***	3.93
	$\frac{4\sigma_t}{\sqrt{2\pi}} \cdot GT \text{ single}$	-	-	0.24	1.51
	$\frac{4\sigma_t}{\sqrt{2\pi}} \cdot ECB/GT \text{ double}$	-	-	0.32**	2.16
	$\frac{4\sigma_t}{\sqrt{2\pi}} \cdot GT/RW \text{ double}$	-	-	0.66***	2.98
	$\frac{4\sigma_t}{\sqrt{2\pi}} \cdot ECB/LL \text{ double}$	-	-	-0.04	-0.24
	$\frac{4\sigma_t}{\sqrt{2\pi}} \cdot ECB/RW/GT \text{ triple}$	-	-	0.35**	2.54
	$\frac{4\sigma_t}{\sqrt{2\pi}} \cdot ECB/GT/LL \text{ triple}$	-	-	0.44**	2.36
	$\frac{4\sigma_t}{\sqrt{2\pi}} \cdot ECB/LL/RW \text{ triple}$	-	-	-0.32	-1.64
	$\frac{4\sigma_t}{\sqrt{2\pi}} \cdot ECB/GT/LL/RW \text{ quad.}$	-	-	0.45**	2.41
	$\frac{4\sigma_t}{\sqrt{2\pi}} \cdot \text{market share} \cdot GM$	0.28***	2.81	0.30***	3.00
Previous crop	Alfalfa	23.3***	2.95	23.3***	2.94
	Wheat	15.7**	2.05	15.5**	2.03
	Soybean	-1.3	-0.15	-1.4	-0.17
	Cucumber	-31.5**	-2.52	-31.8**	-2.54
	Lupine	-10.6	-1.03	-47.1***	-6.33
	Pea	-27.6***	-4.23	-31.4**	-2.52
	Potato	-62.8***	-14.36	-63.1***	-14.42
	Triticale	16.7**	2.34	-19.8***	-3.53
GE traits	GT (glyphosate tolerant)	1.3	0.31	3.7	0.72
	LL (liberty link)	-0.9	-0.16	8.3*	1.89
	ECB (European corn borer)	-8.4**	-2.02	-14.7***	-3.27
	RW (Root worm)	-25.4***	-2.88	-17.2**	-2.10
Stacked GE dummies	ECB/RW	11.2	1.14	21.4**	2.05
	ECB/GT	7.7	1.33	11.4*	1.66
	GT/RW	7	0.69	-17.3	-1.38
	ECB/LL	9.5	1.39	17.9**	2.39
	ECB/RW/GT	25.6**	2.29	19.8*	1.93
	ECB/GT/LL	-0.8	-0.08	-10.4	-1.14
	ECB/LL/RW	34.6***	2.87	46.2***	3.97
	ECB/GT/LL/RW	20	1.39	2	0.17
EVTyrs of GT ^b	-5.4***	-6.17	-5.4***	-6.08	
EVTyrs of GT squared	0.4***	6.55	0.4***	6.48	
EVTyrs of LL	-3.8**	-2.15	-3.8**	-2.12	
EVTyrs of LL squared	0.2**	2.08	0.2**	2.13	

EVTyrs of ECB	1.8**	2.49	1.7**	2.23
EVTyrs of ECB squared	-0.2***	-3.86	-0.2***	-3.86
EVTyrs of RW	2.1	0.47	3	0.66
EVTyrs of ECB_RW	3.1*	1.91	2.4	1.54
EVTyrs of ECB_GT	-3.1***	-3.07	-3.2***	-3.02
EVTyrs of ECB_GT squared	0.2***	2.71	0.2***	2.63
EVTyrs of GT_RW	2.4**	2.19	2.6**	2.36
EVTyrs of ECB_LL	-1.8***	-2.73	-1.7***	-2.60
EVTyrs of ECB_LL squared	0.1	1.48	0.1	1.43
EVTyrs of ECB_RW_GT	-1.3***	-3.69	-1.3***	-3.54
EVTyrs of ECB_GT_LL	-0.2	-0.33	-0.3	-0.37
EVTyrs of ECB_LL_RW	-2.1***	-4.76	-3***	-6.36
EVTyrs of ECB_GT_LL_RW	0.5	1.23	0.5	1.10
Density	0.5*	1.95	0.5*	1.95
Fertilizer	0.1	0.88	0.1	0.88
Irrigation	136.5***	18.89	104.4***	13.99
Insecticide	7.2***	3.19	7.1***	3.15
Fall tillage	29.1***	7.44	29.1***	7.44
Spring tillage	-2.3	-0.58	-2.3	-0.58
Constant	165.6***	11.04	165.7***	11.04

^a To save space, the estimates of parameters associated with year dummies, location dummies and their interactions are omitted from the Table. They are available upon request.

^b “EVTyrs of XX” is the number of years since gene event XX has been first introduced and commercialized in corn.

Specification 2 allows for selectivity effects to vary across GM traits. As shown in Table 2, the effect of market share on selectivity remains similar: it is positive and statistically significant at the 1 percent level. However, specification 2 uncovers evidence of heterogeneous selectivity effects across GM traits. Using a Wald test, we tested the null hypothesis that selectivity bias is the same across GM traits, (i.e. testing specification 2 versus specification 1). The Wald test statistic had a p-value of 0.01. Thus we reject specification 1 and conclude that specification 2 is more appropriate as it captures heterogeneity in selectivity across GM traits.

Under specification 2, it can be seen from table 2 that there is a positive and statistically significant selectivity bias for single trait ECB and GT hybrids, ECB/GT and GT/RW double stack hybrids, ECB/RW/GT and ECB/GT/LL triple stack hybrids, and the quadruple stack

hybrids. The results show stronger selectivity effects for GT/RW double stack hybrids and ECB single trait hybrids (compared to other traits). Finally, from table 2, there is no statistical evidence of selectivity bias for ECB/LL double stack hybrids and ECB/LL/RW triple stack hybrids. The results document that selectivity biases are not uniform and vary across GM traits. We evaluate the extent of these selectivity biases in more detail below.

5. Implications

In this section, our empirical estimates are used to generate additional insights on sources of selection bias, and the changing patterns of selection bias over time. For illustrative purposes, our analysis focuses on GM hybrids in general (specification 1), and for two selected GM hybrids (specification 2) for which our results have shown strong selectivity bias: ECB and GT/RW. We focus our attention on the time period from 1997 to 2010. The estimated effects of selectivity bias on GM corn yield (in bushels per acre) are presented in table 3.

Table 3. Decomposition of selectivity bias (bushels/acre) over time.^a

Year	σ_t	Market share	Specification 1			Specification 2: Single trait ECB			Specification 2: GT/RW double stack		
			GM	Share	Total	GM Trait	Share	Total	GM Trait	Share	Total
1997	14.9	0.18	7.6	1.2	8.8	12.8	1.3	14.1	15.7	1.3	17
1998	16.9	0.18	8.6	1.4	10	14.6	1.5	16	17.8	1.5	19.3
1999	16.2	0.18	8.3	1.3	9.6	14	1.4	15.4	17.1	1.4	18.5
2000	15.2	0.52	7.8	3.5	11.3	13.1	3.8	16.9	16	3.8	19.8
2001	14.6	0.54	7.5	3.5	11	12.6	3.8	16.3	15.4	3.8	19.1
2002	19.5	0.53	10	4.6	14.5	16.8	4.9	21.7	20.5	4.9	25.5
2003	15.5	0.51	7.9	3.6	11.5	13.4	3.8	17.2	16.3	3.8	20.1
2004	17	0.50	8.7	3.8	12.5	14.6	4.1	18.7	17.9	4.1	22
2005	18.3	0.68	9.3	5.5	14.9	15.8	5.9	21.7	19.3	5.9	25.2
2006	19	0.69	9.7	5.9	15.6	16.4	6.3	22.7	20	6.3	26.3
2007	20.5	0.72	10.5	6.6	17	17.7	7	24.7	21.6	7	28.6
2008	11.1	0.79	5.7	3.9	9.6	9.6	4.2	13.8	11.7	4.2	15.9
2009	16	0.80	8.2	5.7	13.9	13.8	6.1	19.9	16.9	6.1	23
2010	18.5	0.86	9.4	7.1	16.6	15.9	7.6	23.6	19.5	7.6	27.1

For both specifications, we decompose the bias effects into two components, those associated with the GM trait(s), and those associated with the market dominance of biotech firms.. From table 3, the selectivity bias shows a rising trend over time, partly due to the expanding market share of the biotech firms. Under specification 1, the selectivity bias generates a sizable yield difference for GM hybrids compared with conventional hybrids: the yield differential ranges from 8.8 bushels per acre to 17 bushels per acre on average for GM hybrids. Recall that the average yield of conventional hybrids in our sample is 166.6 bushels per acre. This indicates that selectivity bias can raise GM corn yield by 5% to 10%. These estimates of the contribution of selectivity are relatively high compared to reported results for the total contribution of some GM traits (see, for example, Stanger and Lauer 2006). The expanding market share of the biotech firms contributes to a strengthening of the bias effects. From 1997 to 2010, the selection bias attributed to market concentration has increased from 13% of the total bias to around 43%.

Under specification 2, for single trait ECB hybrids and GT/RW double stack hybrids, the total bias effects range from 14.1 bushels per acre to 24.7 bushels per acre, and from 17 bushels per acre to 28.6 bushels per acre, respectively. On average, from 1997 to 2010, the total selection biases are 18.8 bushels per acre and 22 bushels per acre for single ECB and GT/RW double stack hybrids. For comparison, the sample mean yield during the same time period is 180 bushels per acre for conventional hybrids, and 18 and 25 bushels per acre higher for single ECB and GT/RW double stack hybrids, respectively. These strong bias effects reflect in part the fact that ECB and GT/RW exhibit the strongest GM selectivity bias effects among all GM hybrids (see table 2). From table 3, it is apparent that the GM trait selectivity effects for ECB and GT/RW are large and dominate the market share effects. However, selection bias effects associated with the

market concentration contributed to around 30% of the total bias for both hybrids in the most recent years, up from 8-9% prior to 2000,

6. Concluding remarks

This paper presents an analysis of potential selectivity issues in the productivity evaluation of GM technology. GM genes are at the heart of the current biotechnology revolution in agriculture. However, traditional genetic improvements have played, and continue to play, an important role in contributing to productivity increases in agriculture. Given the presence of many traditional genes in any germplasm, identifying the separate effects of GM genes versus conventional genes can be difficult. This is the essence of the selectivity problem addressed in this paper.

Our analysis provides a framework to investigate these issues, with an empirical application to GM corn yield in the US. It relies on a “Heckman-like” factor that accounts for selectivity. The selectivity can arise if biotech/seed firms insert GM traits only in high quality germplasm. In this case, observing high GM corn yield can be due in part to the high quality germplasm. Correcting for this selection effect is important to identify the true productivity effect of GM traits.

Using experimental data on Wisconsin corn yields covering the period 1990-2010, our analysis uncovers evidence of selection bias. It indicates that some of the observed yield gains associated with GM hybrids come from conventional genes. This stresses the important role that traditional breeding still plays in productivity improvements for corn, even after the widespread adoption of GM corn in the US. We also find that these selectivity effects are not uniform across GM traits. In some cases, the effects are small and not statistically significant (e.g., ECB/LL). In other cases, these effects can be large (e.g. ECB and GT/RW). We also investigate how market

concentration of biotech firms can affect selectivity bias in corn yield. We find that rising market concentration has contributed to significant increases in selectivity bias in corn yield over the last 15 years. This raises questions about whether current patenting laws appropriately differentiate between GM genes and conventional genes, and whether they provide appropriate incentives for investment in traditional breeding programs.

While our analysis was applied to corn yield data in Wisconsin, additional research is needed to explore selectivity issues in broader contexts. First, it would be useful to investigate such issues in other regions of the world. This is relevant to the extent that breeders often develop hybrids that are adapted to local agro-climatic conditions. This indicates that our findings may not hold in different regions. Second, more work is needed to explore whether selectivity bias may also arise in the productivity evaluation of other GM crops (e.g., cotton, soybean). Third, the implications of our findings raise questions about whether current patent laws provide appropriate incentives to support investments in both agro-biotechnology and traditional breeding programs. This appears to be a good topic for further research.

References

- Alston, J.M., and P.G. Pardey. *Making Science Pay: The Economics of Agricultural R&D Policy*. Washington, DC: AEI Press, 1996.
- Alston, J.M. and R.J. Venner Raymond J. “The effects of the US Plant Variety Protection Act on wheat genetic improvement.” *Research Policy* 31(2002): 527-542.
- Bouis, H. “The Potential of Genetically Modified Food Crops to Improve Human Nutrition in Developing Countries.” *Journal of Development Studies* 43(2007): 79-86.
- Duvick, D.N. “Plant Breeding: Past Achievements and Expectations for the Future.” *Economic Botany* 40(1992): 289-97.
- Duvick, D. “The Contribution of Breeding to Yield Advances in Maize.” *Advances in Agronomy* 86(2005): 84-145.
- Fernandez-Cornejo, J. *The Seed Industry in U.S. Agriculture: An Exploration of Data and Information on Crop Seed Markets, Regulation, Industry Structure, and Research and Development*, Resource Economics Division, Economic Research Service, U.S. Department of Agriculture, Agriculture Information Bulletin Number 786, 2004.
- Fulton M. and K. Giannakas. “Agricultural Biotechnology and Industry Structure.” *AgBioForum* 4(2001): 137-151.
- Goetz, S.J. “A Selectivity Model of Household Food Marketing Behavior in Sub-Saharan Africa.” *American Journal of Agricultural Economics* 74(1992): 444-452.
- Graff G.D., G.C. Rausser and A.A. Small. “Agricultural Biotechnology's Complementary Intellectual Assets.” *Review of Economics and Statistics* 85(2003): 349-363.
- Greene, W.E. *Econometric Analysis*. Prince Hall, Upper Saddle River, NJ, 2000.
- Griliches, Z. “Hybrid Corn: An Exploration in the Economics of Technological Change.” *Econometrica* 25(1957): 501-522.
- Griliches, Z. “Hybrid Corn and the Economics of Innovation.” *Science*, 132 (1960): 275-280.
- Heckman, J. “Sample Selection Bias as a Specification Error.” *Econometrica* 47(1979): 153-161.

- Herd, R.W. "Biotechnology in Agriculture." *Annual Review of Environment and Resource* 31(2006): 265-295.
- James, C. "Executive Summary of Global Status of Commercialized Biotech/GM Crops: 2010." *International Service for the Acquisition of Agri-Biotech Applications*, Briefs No. 42-2010, Ithaca, NY, 2010.
- Johnson, N.L. and S. Kotz. *Distributions in Statistics: Continuous Univariate Distributions* Houghton Mifflin, Boston, 1970.
- King, J.L., and D. Schimmelpfennig. "Mergers, Acquisitions, and Stocks of Agricultural Biotechnology Intellectual Property." *AgBioForum* 8(2005): 83-88.
- National Academies. *Impact of Genetically Engineered Crops on Farm Sustainability in the United States*. The National Academies Press, Washington, D.C., 2010.
- Qaim, M. "The Economics of Genetically Modified Crops." *Annual Review of Resource Economics* 1(2009): 665-693.
- Rausser, G.C., S. Scotchmer and L.K. Simon. "Intellectual Property and Market Structure in Agriculture." Working Paper Series 141901, Department of Agricultural & Resource Economics, UC Berkeley, (1999).
- Ronald, P. "Plant Genetics, Sustainability and Global Food Security." *Genetics* 1(2011): 11-20.
- Shi, G. "Bundling and Licensing of Genes in Agricultural Biotechnology." *American Journal of Agricultural Economics* 91(1) (2009): 264-274.
- Shi, G., J.P. Chavas and K.W. Stiegert. "An Analysis of the Pricing of Traits in the U.S. Corn Seed Market." *American Journal of Agricultural Economics* 92(5) (2010): 1324-1338.
- Springer, N.M. and R.M. Stupar. "Allelic Variation and Heterosis in Maize: How do Two Halves Make More than a Whole?" *Genome Research* 17(2007): 264-275.
- Stanger, T. F., and J. G. Lauer. "Optimum Plant Population of Bt and Non-Bt Corn in Wisconsin." *Agronomy Journal* 98(4) (2006): 914-921.
- USDA-NASS. *Crop Production Summary*. National Agricultural Statistics Service, Washington, DC (2011).

Vella, F. "Estimating Models with Sample Selection Bias: A Survey" *Journal of Human Resources* 33(1998): 127-169.

Wright, B.D. and P.G. Pardey. "The Evolving Rights to Intellectual Property Protection in the Agricultural Biosciences." *International Journal of Technology and Globalization* 2(2006): 12-29.

Footnotes

¹ The patenting of seeds followed the development of strong intellectual property rights applied to living organisms, as established by the 1980 US Supreme Court case *Diamond v. Charkrabarty*.

² Pioneer has been breeding corn seed since 1921, and owns a large bank of genetic material but was not strong in biotechnology traits. Its 1999 merger with DuPont was designed to address this problem (King and Schimmelpfennig 2005). Monsanto was a chemical company which acquired Asgrow, DeKalb and Holden's Foundation Seeds between 1997 and 2000 to obtain germplasm into which it could insert biotechnology traits (King and Schimmelpfennig 2005). Monsanto has a major brand in DeKalb, and has acquired more than 30 regional seed companies and brands through its Channel Biocorp and American Seeds businesses. Syngenta is European based, and in 2004 purchased J.C. Robinson, Golden Harvest, and the US operations of Advanta to add to its existing major brands of NK, Garst and AgriPro. Dow acquired a large number of biotechnology patents through its acquisition of Mycogen in 1998, and has continued to expand in the US, buying the US operations of Cargill in 2000, Triumph Seed, Dairyland Seed and Renze Seed in 2008, Pfister in July 2009, Grand Valley in 2010, and Prairie Brand Seed in June 2011. While DuPont has traditionally relied on germplasm acquired with Pioneer HiBred, and has not previously engaged in takeover activity, it purchased Pannar Seed, Curry Seeds, AgVenture, Hoegemeyer Hybrids, NuTech Seed, Seed Consultants and Terral Seeds in 2010, and Doeblers PA Hybrids in 2011. It has also entered into strategic alliances with Beck's Hybrids and Burrus Hybrids.

³ See Vella (1998) for a review of this literature and its extensions.

⁴ In this case, low quality germplasm would be less likely to be associated with GM seeds. It means that the lower tail of the observed quality distribution of conventional genes would change depending on how the germplasm gets combined with GM genes.

⁵ Note that, while the specification $S_t \equiv \mathbf{z}_t \boldsymbol{\gamma}$ does not restrict S_t to be between 0 and 1, this does not seem to be an issue in our empirical analysis. Indeed, the results reported in table 2 below corresponds to estimates of S_t that typically remain between 0 and 1.

⁶ Biotech traits are associated with “genetic events” corresponding to the chromosomal location of the transgene. An event refers to the unique DNA recombination event that took place in one plant cell, which was then used to generate entire transgenic plants.

⁷ Since “conventional seeds” are not included among these dummy variables, it means that coefficients of these dummies are interpreted as yield deviations from conventional seeds.

⁸ Single trait RW and LL hybrids, and the double stack ECB/RW hybrids are not included because they have a small number of observations.