AN ANALYSIS OF SELECTIVITY IN THE PRODUCTIVITY EVALUATION OF BIOTECHNOLOGY: AN APPLICATION TO CORN

GUANMING SHI, JEAN-PAUL CHAVAS, JOSEPH LAUER, AND ELIZABETH NOLAN

We investigate selectivity bias in the evaluation of biotech hybrid productivity. The analysis is applied to experimental data on Wisconsin corn yields from 1990 to 2010. Relying on a Heckman-like factor that accounts for selectivity, we find evidence of selection bias, indicating that some of the observed yield advantage associated with GM hybrids can be attributed to their conventional genes. We document how the rising market concentration of biotech firms has contributed to increasing selectivity bias in corn yield. The impact, however, is offset by the negative effect of the rising adoption rate of GM corn on selectivity bias.

Key words: biotechnology, corn, genetic modification, productivity, selectivity bias.

JEL codes: D02, L10, O10, Q01.

Genetic improvement has been a major source of agricultural productivity growth (Alston and Pardey 1996). For example, in the United States, 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-NASS 2011). Genetic improvements have been major contributing factors to this growth, and are considered to have contributed between 50-60% of corn yield gains during the 20th century (Duvick 1992, 2005).

Two innovations in corn breeding have played a key role in the crop's growth. 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 where private firms dominate the production and distribution of corn seed (Fernandez-Cornejo 2004). The second innovation was the introduction of genetically modified (GM) traits embedded in seeds, which took place in the 1990s following the development of gene transfer technology.

Historically, genetic selection has focused on traditional genes that have been selected mostly through "trial and error", in which case the genes contributing to higher yields were not explicitly identified. With the advent of biotechnology, GM genes (and their functions) are now identified, patented[1] and "transferred" (often across species) to a targeted organism. 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 adoption of GM corn hybrids has been rapid in the United States: in 2011, 88% of U.S. corn acreage was planted to hybrids with at least one GM trait, and 49% of the corn acreage was planted to hybrids with at least two GM traits (USDA-ERS 2011).

Guanming Shi is Associate Professor, Department of Agricultural and Applied Economics, University of Wisconsin, Madison. Email: gshi@wisc.edu. Jean-Paul Chavas is Professor, Department of Agricultural and Applied Economics, University of Wisconsin, Madison. Email: jchavas@wisc.edu. Joe Lauer is Professor, Department of Agronomy, University of Wisconsin, Madison. Email: jglauer@wisc.edu. Elizabeth Nolan is Lecturer, Department of Agricultural and Resource Economics, University of Sydney, 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 USDA/NIFA grant #2010-04833.

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

Amer. J. Agr. Econ. 95(3): 739–754; doi: 10.1093/ajae/aas169

Genetically modified genes do not work in isolation. Rather, they interact with and complement traditional genes, which helps the GM crops become more effective in their environment (e.g. by reducing pest damages). However, how much of the observed productivity improvement in GM crops can be attributed to GM genes, and how much to traditional genes? Answering this question involves complex issues because, while there are relatively few GM genes, many traditional genes in many different combinations contribute to corn productivity.[2]

Our paper compares the quality of traditional genes contained in GM crops to the quality of traditional genes contained in conventional crops. In general, higher quality germplasm leads to better performing seeds. Under competition, one expects firms to provide an efficient level of quality. However, a multiproduct monopolist has an incentive to offer sub-optimal levels of quality to low-valued consumers (e.g. Mussa and Rosen 1978; Tirole 1988, p. 150). Therefore, the provision of quality can be affected by market structure. In the case of GM seeds, biotech firms own patents giving them legal monopoly rights over the GM genes. Over the last 20 years, these biotech firms have been acquiring firms who previously owned basic hybrid germplasm and distribution networks (e.g. Alston and Venner 2002; Fernandez-Cornejo 2004; Wright and Pardey 2006; Howard 2009; National Academies 2010). The corn seed industry is now dominated by four large biotech firms, all of which own subsidiary seed companies (Fernandez-Cornejo 2004; Shi, Chavas and Stiegert 2010). These companies are: DuPont (Pioneer International), Monsanto, Syngenta, and Dow AgroSciences.[3] The market share of these 4 biotech firms in the U.S. corn seed market has increased over the last 15 years, reaching 86% in 2010.[4] While Pioneer and Monsanto each control about 35% of the corn seed market, Monsanto dominates the GM trait market; in 2009 their market share was over 80% (Moschini 2010). Such concentrated seed markets are associated with an increase in control over the quality of the germplasm. Chataway and Tait (2000) pointed out that part of Monsanto's incentives to acquire seed companies was to have access to elite germplasm because, "The Monsanto … gene was not originally inserted in the best variety." Unfortunately, we do not have direct observations on the quality of basic germplasm used in biotech and conventional seeds. However, it is possible to arrive at an indirect evaluation of the quality of the germplasm managed by biotech and seed companies.

In this paper, we rely on experimental corn yield data to provide indirect evidence on the management of germplasm quality in the U.S. corn seed industry. We develop a method to investigate whether the quality of the traditional germplasm is the same for GM seeds as it is for conventional seeds. We use arguments concerning "selectivity bias" presented in the econometric literature (as pioneered by Heckman, 1979), and adapt them to the evaluation of GM technology. Selection bias can arise when data are generated in a non-random fashion. Examples include self-selection, prescreening and the general evaluation of a sub-sample that excludes some members of the population of interest. Selectivity issues are relevant when analyzing labor, migration and marketing decisions (e.g. Heckman 1979; Goetz 1992).[5] In these cases, selectivity arises when individual performance is evaluated based on a sub-sample that is not representative of the population.

This paper investigates the possible presence of selectivity bias when evaluating GM seed productivity. The idea is simple: in the absence of selectivity bias, the quality of the germplasm does not differ between GM seeds and conventional seeds, implying that, for given agro-climatic conditions and management, any productivity difference between GM and conventional crops can be attributed to the GM genes. But "selectivity" would arise if GM genes were added more frequently to "high

---

[2] As pointed out by an anonymous referee, observing the effects of conventional germplasm can be difficult. Indeed, given the estimated 42,000 to 56,000 genes present in corn, evaluating the specific and joint effects of each gene is a complex task.

[3] 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). Since then, mergers and acquisitions have led to further structural changes and increased concentration in the corn seed industry (see Shi, Chavas and Stiegert 2010; Nolan and Santos 2012).

[4] This ratio is compiled using statistics reported by each firm's annual report; it only measures the total market share of the integrated firms and should not be interpreted as the commonly referred to four-firm concentration ratio (CR4). While some of the integrated firms have been among the top 4 firms in the industry for some years, not all of them were so highly-ranked during the study period.

[5] See Vella (1998) for a review of this literature and its extensions.

quality" germplasm. In that case, seeds with GM traits would appear to be more productive, but part of this increase would not be due to GM genes, but rather to the higher quality of the underlying germplasm. Therefore, observed yield differences between GM and conventional seeds are subject to "selectivity bias". The conceptual contribution of this paper is to show how Heckman's approach to selectivity bias can be adapted to evaluating the productivity of GM technology.

We also make an empirical contribution to the literature by applying the methodology to experimental data on GM corn yield. We evaluate the separate effect on yield of GM genes versus traditional germplasm using experimental data on Wisconsin corn yields from 1990 to 2010. Our study provides a refined empirical investigation of the relative effects that both GM and traditional genes have on agricultural productivity.

The empirical findings are timely and relevant in that we uncover evidence indicating that part of the observed yield gains from biotech seeds arise because patented GM traits are being inserted into superior germplasm. The analysis finds that such selectivity effects vary between different GM genes. We also demonstrate that selection bias in corn yields has been affected by the rising market concentration of biotech firms and the rate of GM adoption. Moreover, we show that selectivity effects become stronger as the dominance of biotech firms in the seed industry increases, but decline with increased GM adoption rates. The analysis indicates the importance of conventional genes in evaluating the productivity effects of biotech genes, and stresses the role played by traditional genetic selection in agricultural productivity improvements. All of this has important implications for research policy. It appears that, even given the widespread adoption of GM corn in the United States, traditional breeding continues to generate significant increases in agricultural productivity. Given the important part that traditional plant breeding continues to play in increasing agricultural productivity, it is possible that both the current emphasis on investment in biotech, and the patent protection extended to GM traits should be reconsidered.

Economics of Selectivity Bias in Biotech Seeds

Consolidation in the seed industry has been driven by a number of factors: exploiting asset complementarities; mitigating contractual hazards; seeking market power; and/or regulations (e.g. Kalaitzandonakes and Bjornsen 1997; Rausser, Scotchmer and Simon 1999; Fulton and Giannakas 2001; Graff, Rausser and Small 2003; Just, Alston and Zilberman 2006; Marco and Rausser 2008; Shi 2009). The presence of high fixed costs and low variable costs associated with biotechnology indicates a need to price above marginal costs to support high levels of research and development investment (Wright and Pardey 2006). Consolidation has also been driven by the need for access to elite germplasm (Wright and Pardey 2006; Chataway and Tait 2000). New and valuable traits cannot generate value unless they are incorporated in commercial hybrid lines sold to farmers. To sell their products and gain market share, biotech firms have the incentive to add their patented GM traits to very good basic germplasm. Since the marginal cost of incorporating a trait into a hybrid is small, the earnings from licensing a GM 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 helps explain how large biotech firms expand and integrate into retailing by purchasing regional seed companies.

Similar to the multi-product monopolist, integrated biotech/seed companies that can produce both conventional and biotech seeds are in a position to manage the quality of the germplasm used for GM seeds. As pointed out by Tirole (1988, p. 150), these companies have an incentive to provide suboptimal germplasm quality to low-valued consumers (in this context, farmers using conventional seeds) while providing high quality germplasm to high-valued consumers (farmers using GM seeds).[6]

To illustrate such a decision process, we follow Mussa and Rosen (1978) and consider the case where each farmer buys one unit of seed of quality q. There are two types of farmers ("type b" and "type c"), each type differing according to his/her valuation of quality. Assume that a farmer of "type b" receives gross benefit $B_b = \theta_b q_b$, while a farmer of "type c"

---

[6] Here, high-value and low-value farmers are defined according to each farmer's net willingness to pay for GM seeds. For example, when GM seeds help reduce pest damage, high-value farmers would be the ones facing more severe pest infestation problems on their farm. Alternatively, farmers facing low pest infestation may have little incentive to buy GM seeds: they would be low-value farmers likely to purchase conventional seeds.

receives gross benefit $B_c = \theta_c q_c$, with $\theta_b > \theta_c$. This assumption reflects heterogeneity in the marginal benefit of quality among farmers: "type b" farmers receive a higher incremental benefit of quality than do "type c" farmers (e.g. due to different agro-climatic conditions).[7] Let n_b and n_c be the number of farmers of "type b" and "type c", respectively. Let $C(q)$ denote the unit cost of producing quality q, where $C(q)$ is increasing and convex in q.

Three scenarios are of interest. First, consider the case of a social planner trying to maximize aggregate welfare: $n_c\ [\theta_c q_c - C(q_c)] + n_b[\theta_b q_b - C(q_b)]$. Assuming differentiability and interior solutions, the first-order conditions for an efficient choice of quality are:

$$(1a) \quad \theta_b = C'(q_b),$$

and

$$(1b) \quad \theta_c = \quad C'(q_c)$$

where $C'(q) = \partial C/\partial q$ denotes the marginal cost of quality. Equations (1*a*) and (1*b*) provide the standard efficiency results stating that marginal benefit equals marginal cost. Second, consider the case of a perfectly-discriminating monopolist who can charge different prices to different buyers. This monopolist would offer each farmer the socially-efficient quality, but would extract all economic surplus by charging a price higher than θ_c in (1*b*) for the lower quality level (Mussa and Rosen, 1978, p. 304).

Third, consider the case of a monopolist who offers all potential buyers the same price-quality combination. This is a situation of second-degree price discrimination, where buyers are heterogeneous but the monopolist can still price discriminate by inducing different buyers to self-select into buying price-quality combinations that increase monopoly profit (Tirole, 1988, p. p. 143). Denote by $P(q)$ the price charged by the monopolist for quality q. As shown by Mussa and Rosen (1978, p. 305), the monopolist's price-quality offers must satisfy $P(q_c) = \theta_c q_c$ and $P(q_b) = \theta_c q_c + \theta_b\ [q_b - q_c]$. In our example, such offers ensure that the two types of farmers are self-separated, leading only high-valued "type b" farmers to buy GM seeds. Then, total profit is $n_c\ [\theta_c q_c - C(q_c)] + n_b[\theta_c q_c + \theta_b[q_b - q_c] - C(q_b)]$, and the associated first-order conditions for interior solutions are:

$$(2a) \quad \theta_b = C'(q_b),$$

and

$$(2b) \quad \theta_c + (n_b/n_c)[\theta_c - \theta_b] = C'(q_c).$$

Equation (2a) is the same as (1*a*), and shows that the discriminating monopolist provides the efficient level of quality to "type b" farmers. But given $\theta_b > \theta_c$, comparing equation (2b) with (1*b*) implies that "type c" farmers face lower and suboptimal levels of quality. The extent of sub-optimality depends on the relative number of farmers of each type, n_b and n_c. Here, "type c" farmers could be those who buy conventional seeds; they are induced to buy lower quality seeds to improve the ability of the monopolist to charge higher prices to "type b" farmers who purchase GM seeds. In this case, the monopolist would choose suboptimal quality for seeds supplied to the low-valued conventional farmers.

While the analysis presented above is relatively simple,[8] it is relevant in the presence of heterogeneity of benefits generated by GM technology across farms. For example, the level of pest infestation often varies across farms due to agro-climatic conditions and site-specific pest population dynamics. In this context, when a GM trait is used to control a particular pest, the willingness-to-pay for this GM trait would vary across farms. Under second-degree price discrimination and the legal monopoly granted by a patent, biotech and seed firms would then have an incentive to create quality differences (besides GM genes) between traditional seeds and GM seeds. Such quality differences would generate selectivity issues in the productivity evaluation of GM hybrids.

Besides seeking evidence of selectivity bias, we are also interested in investigating factors that may contribute to such bias. On *a priori* grounds, there are two situations where selectivity issues would not be expected. The first situation is one of perfect competition in seed markets. Since intellectual property rights (such as patents, plant variety protection certificates and trade secrets) grant legal monopoly to these rights holders, the extreme case of perfect competition would require the

---

[7] For example, "type b" farmers (biotech seed farmers) may face higher pest pressure than "type c" farmers (conventional farmers), implying that biotech farmers would benefit more from a "high quality" seed that offers some protection against yield loss from pest damages.

[8] For a more general discussion, see Mussa and Rosen (1978).

absence of any such rights, implying open access to the germplasm. Such open access would prevent any firm from implementing a selectivity scheme when matching GM with conventional genes. In fact, the implementation of a selectivity scheme by any firm requires some level of control over the germplasm. In this case, selectivity issues would be more likely to arise in the presence of imperfect competition, which generates one of our testable hypotheses: selectivity bias is positively associated with increased market concentration. Indeed, while the ownership of technologies associated with GM seeds is now concentrated in four biotech firms, these firms have also acquired a significant amount of elite basic inbred germplasm (Wright and Pardey 2006).

Second, selectivity issues would not arise if all farmers possessed adopted GM seeds. In this case, all seeds would be GM seeds,[9] the market for conventional seeds would disappear, and there would be no prospect for any firm to implement a selectivity scheme between GM and conventional genes. Thus, the likelihood of finding selectivity issues would decrease with the adoption rate of GM technology, which generates another of our testable hypotheses: selectivity bias is negatively associated with the adoption of GM technology.

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:

$$(3) \qquad y = f(\mathbf{G}, \mathbf{V}, \mathbf{M})$$

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 (3) represents a generic agricultural production technology, allowing for possible interactions among genes, and between genotype, environmental conditions and management factors.

Equation (3) provides the information needed to evaluate yield. 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 yield can be measured by:

$$(4) \qquad \Delta y(\mathbf{G}^{t'}, \mathbf{G}^t; \mathbf{V}, \mathbf{M}) = f(\mathbf{G}^{t'}, \mathbf{V}, \mathbf{M}) - f(\mathbf{G}^t, \mathbf{V}, \mathbf{M}).$$

Note that $\Delta y(\cdot)$ in (4) 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 (4) 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 $\mathbf{G}_c$.

As discussed in the introduction, there are few biotech genes $\mathbf{G}_b$ 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 respective contribution to agricultural productivity. Nevertheless, conventional genes have historically been the subject of intense genetic selection by farmers, and more recently by plant breeders, and have contributed to major improvements in agricultural productivity (Alston and Pardey 1996; Duvick 1992). Such progress in conventional gene research should be recognized in any study on the effects of introducing GM traits.

At time t and in a given region, conditional on $\mathbf{G}_{bt}$, denote the population distribution of $\mathbf{G}_c$ by $\Gamma_t(\mathbf{G}_c|\mathbf{G}_{bt})$. The distribution function $\Gamma_t(\cdot|\mathbf{G}_{bt})$ has a time t subscript to account for the fact that genetic selection has been associated with significant changes in the quality

---

[9] This scenario is hypothetical. There are at least two reasons why the adoption of GM corn hybrids would not reach 100%: the presence of organic farming where planting GM seeds is not allowed; and regulatory-mandated refuge requirements where some percentage of acres must be planted in conventional hybrids (to slow the development of insect resistance to GM toxins). Still, this hypothetical scenario can be thought as approximating what would happen when the market share of GM seeds becomes "large".

of germplasm over time. Then, from (3) and conditional on $(\mathbf{G}_{bt}, \mathbf{V}_t, \mathbf{M}_t)$, average production at time t is given by:

$$(5) \quad g_t(\mathbf{G}_{bt}, \mathbf{V}_t, \mathbf{M}_t, \boldsymbol{\beta}) \equiv E_t[f(\mathbf{G}_{bt}, \mathbf{G}_c, \mathbf{V}_t, \mathbf{M}_t)] = \int_{\mathbf{G}_c} f(\mathbf{G}_{bt}, \mathbf{G}_c, \mathbf{V}_t, \mathbf{M}_t)\, d\Gamma(\mathbf{G}_c|\mathbf{G}_{bt})$$

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

Since the advent of biotechnology, we have faced a more complex situation where both biotech genes and conventional genes change over time. As noted earlier, the use of gene-transfer technology offers good prospects for additional productivity growth in agriculture. However, conventional genes are numerous and difficult to measure, and selectivity issues can arise when the biotech seed companies add biotech genes only to "high quality" conventional genes to generate superior productivity of the biotech seeds.

To analyze possible selectivity in the evaluation of biotech gene productivity, we start with the average production given in (5). At time t, it follows from (5) that equation (3) can be written as:

$$(6) \quad y_t = g_t(\mathbf{G}_{bt}, \mathbf{V}_t, \mathbf{M}_t, \boldsymbol{\beta}) + e_t$$

where $e_t \equiv y_t - g_t(\mathbf{G}_{bt}, \mathbf{V}_t, \mathbf{M}_t, \boldsymbol{\beta})$ is an error term satisfying $E(e_t) = 0$. Equation (6) is a regression model. At time t, the error term e_t in (6) measures the deviation from mean production reflecting heterogeneity in unobservable conventional genes. As such, the distribution of e_t provides an indirect measurement of conventional gene distribution.

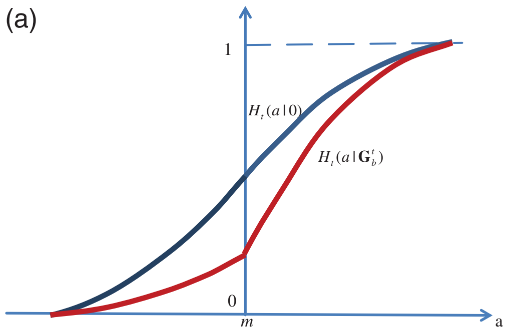
In principle, once specified equation (6) can be estimated using observations on production output y_t, and $(\mathbf{G}_{bt}, \mathbf{V}_t, \mathbf{M}_t)$. For given $(\mathbf{V}_t, \mathbf{M}_t)$, the estimation can provide useful information on how alternative biotech genes $\mathbf{G}_{bt}$ affect productivity. However, one needs to consider whether, in equation (6), the distribution of e_t is independent of $\mathbf{G}_{bt}$. If so, then a standard least-squares estimation of (6) would provide an unbiased estimate of the productivity effects of biotech genes, $\mathbf{G}_{bt}$.

However, this condition (e being independent of $\mathbf{G}_{bt}$) would fail to hold in the presence of genetic selectivity. Let $\mathbf{G}_{bt} = 0$ in the absence of biotech genes. Conditional on $(\mathbf{V}_t, \mathbf{M}_t)$, define $e_t(0) \equiv y_t - g_t(0, \mathbf{V}_t, \mathbf{M}_t, \boldsymbol{\beta})$, which captures the difference in quality of the basic germplasm in terms of yield. Denote the distribution of $e_t(0)$ by $H_t(a|0) = \text{Prob}[e_t(0) \le a]$. Next, consider situations where biotech genes are added to the basic germplasm, and, to address selectivity issues, we consider the case where biotech genes, $\mathbf{G}_{bt}$, may be bundled together with "high quality" conventional genes. Let $m \ge 0$ be a threshold of germplasm quality satisfying $H_t(m|0) \in (0,1)$, where $e_t \ge m$ 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 $\mathbf{G}_b^t \ne 0$, assume that the distribution of $e_t(\mathbf{G}_b^t)$ is given by:

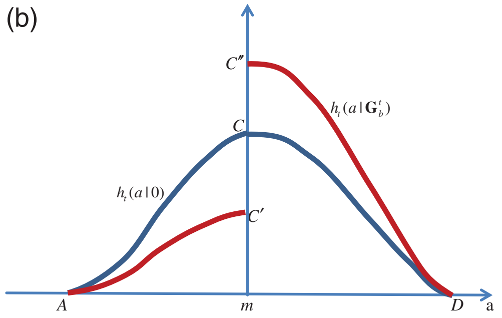
$$(7) \quad H_t(a|\mathbf{G}_b^t) = [1 - S_t]H_t(a|0) \text{ if } a < m, \quad = -\alpha_t + [1 + \alpha_t]H_t(a|0) \times \text{ if } a \ge m$$

where $\alpha_t \equiv S_t H_t(m|0)/[1 - H_t(m|0)]$. The parameter S_t in (7) acts as a selectivity index.

When $S_t = 0$, it follows from (7) that $H_t(\cdot|\mathbf{G}_b^t) = H_t(\cdot|0)$, and hence the distribution of basic germplasm quality does not differ between conventional and biotech seeds. In this case, $E[e_t(\mathbf{G}_b^t)] = 0$, and there is no selectivity bias in equation (6). However, when $S_t > 0$, selectivity bias exists since equation (7) implies that $H_t(a|0)$ tends to be larger than $H_t(a|\mathbf{G}_b^t)$. Figure 1 illustrates such a selectivity bias, and the distribution functions $H_t(a|0)$ and $H_t(a|\mathbf{G}_b^t)$, and the associated probability density functions $h_t(a|0)$ and $h_t(a|\mathbf{G}_b^t)$ are shown in (1*a*) and (1*b*), 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 downward as the probabilities of facing any event $a \in [-\infty, m]$ are rescaled by a factor $(1 - S_t) \in [0,1]$. When $S_t > 0$, there is a reduction in the probability that biotech seeds would have "low quality" germplasm. In a second step, all probabilities are shifted upward 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 $H_t(a|\mathbf{G}_b^t)$.



Distribution functions $H_t(a|0)$ and $H_t(a|\mathbf{G}_b^t)$



Probability density functions $h_t(a|0)$ **and** $h_t(a|\mathbf{G}_b^t)$

Figure 1. An illustration of selectivity effects

In the extreme case where $S_t = 1$, equation (7) would imply that $H_t(a|\mathbf{G}_b^t) = 0$ when $a < m$, that is, the distribution function $H_t(a|\mathbf{G}_b^t)$ becomes truncated below the threshold point m. Such extreme selection occurs where no low quality germplasm is used in biotech seeds. In intermediate situations where $S_t \in (0, 1)$, equation (7) allows for "partial selectivity", that is, the intensity of selection increasing with S_t.

Given $\mathbf{G}_b^t$, the expected value of output y_t at time t is:

$$(8)\quad \begin{aligned} &E_t(y_t|\mathbf{G}_b^t)\\ &= g(\mathbf{G}_b^t, \mathbf{V}_t, \mathbf{M}_t, \boldsymbol{\beta}) + E_t[e_t(\mathbf{G}_b^t)]\\ &= g(\mathbf{G}_b^t, \mathbf{V}_t, \mathbf{M}_t, \boldsymbol{\beta})\\ &\quad + \int_e e\, dH_t(e|\mathbf{G}_b^t), \text{ or using (7)},\\ &= g(\mathbf{G}_b^t, \mathbf{V}_t, \mathbf{M}_t, \boldsymbol{\beta})\\ &\quad + [1 - S_t]\int_{e<m} e\, dH_t(e|0)\\ &\quad + [1 - S_t + S_t + \alpha_t]\\ &\quad \times \int_{e \geq m} e\, dH_t(e|0),\\ &= g(\mathbf{G}_b^t, \mathbf{V}_t, \mathbf{M}_t, \boldsymbol{\beta})\\ &\quad + K_t(m, S_t)\int_{e \geq m} e\, dH_t(e|0),\\ &\quad \text{since } e_t(0) \text{ has mean zero} \end{aligned}$$

where

$$(9)\quad \begin{aligned} K_t(m, S_t) &\equiv S_t + \alpha_t\\ &= S_t/[1 - H_t(m|0)] \geq 0. \end{aligned}$$

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 (8) implies that $E(y_t) \geq g(\mathbf{G}_b^t, \mathbf{V}_t, \mathbf{M}_t, \boldsymbol{\beta})$ in general, with $E(y_t) = g(\mathbf{G}_b^t, \mathbf{V}_t, \mathbf{M}_t, \boldsymbol{\beta})$ when $S_t = 0$ and $E(y_t) > g(\mathbf{G}_b^t, \mathbf{V}_t, \mathbf{M}_t, \boldsymbol{\beta})$ when $S_t \in (0, 1]$. The selectivity bias is identified as:

$$(10)\quad D_t \equiv K_t(m, S_t)\int_{e \geq m} e\, dH_t(e|0) \geq 0.$$

Equation (10) provides 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 a mean of zero and a variance of 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 (10) becomes:

$$(10')\quad \begin{aligned} D_t \equiv K_t(m, S_t)&\{\phi(m/\sigma_t)/\\ &\times [1 - \Phi(m/\sigma_t)]\}\sigma_t. \end{aligned}$$

Note that the term $\{\phi(m/\sigma_t)/[1 - \Phi(m/\sigma_t)]\}$ in (10′) is the inverse Mills ratio, which is commonly used to analyze selectivity bias under normality (e.g. Heckman 1979). Equation (10′) 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 (8), (9) and (10′) provides:

$$(11a)\quad y_t = g(0, \mathbf{V}_t, \mathbf{M}_t, \boldsymbol{\beta}) + e_t, \text{ when } \mathbf{G}_b^t = 0,$$

and

$$(11b)\quad = g(\mathbf{G}_b^t, \mathbf{V}_t, \mathbf{M}_t, \boldsymbol{\beta}) + S_t[1 - H_t(m|0)]^{-1} \times \{\phi(m/\sigma_t)/[1 - \Phi(m/\sigma_t)]\}\sigma_t + u_t, \quad \text{when } \mathbf{G}_b^t \neq 0$$

where u_t is an error term with a mean of zero. Equations (11*a*)–(11*b*) provide a basis to empirically evaluate the presence of gene selectivity involving both traditional genes and biotech genes.

We must consider which factors would lead to genetic selectivity. As discussed above, selectivity would be absent if biotech genes were randomly inserted in the current germplasm. Alternatively, selectivity would arise if only the higher quality germplasm were chosen for GM seed production. If this were the case, selectivity would be associated with genetic selection that imposes some level of control on the distribution of germplasm and its quality. Thus, the selectivity parameter S_t in (11*b*) can vary depending on the situation.

To investigate the likelihood that selectivity may vary, 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.[10] Then, equations (11*a*)–(11*b*) 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 (11*b*) allows us to investigate and measure the effects of genetic selectivity on productivity assessment. As discussed earlier, the assertion of greater control over the quality of the germplasm may be associated with more concentrated seed markets. Similarly, selectivity may be less likely to arise when the adoption of GM hybrids is high. If so, genetic selectivity may vary with market conditions. Equation (11) (with $S_t \equiv \mathbf{z}_t\boldsymbol{\gamma}$) will allow us to investigate these issues empirically.

Data

To investigate the potential for genetic selectivity, we use data on corn yield obtained from field experiments conducted from 1990 to 2010 at the University of Wisconsin. The field experiments were undertaken at the university's Agricultural Research Stations and at long-term farmer cooperators located across the state of Wisconsin (see http://corn.agronomy.wisc.edu/HT/images/Map.jpg for a map of the research locations).[11] In these experiments, management practices were typical of those utilized on farms practicing mainly rainfed agriculture, including planting density, spring/fall tillage, fertilizer/herbicide/insecticide applications, and irrigation. A total of 4,748 hybrids have been tested in the past 21 years. Of these, 2,653 are conventional hybrids, and 2,095 are GM hybrids. All hybrids are tested in multiple sites and some for multiple years, yielding 31,799 usable observations for the analysis.

There are two major groups of GM traits in the corn hybrid market: those which provide insect resistance (IR) and those which provide herbicide tolerance (HT). The IR traits are designed to control specific pest populations, thus reducing corn yield losses due to pest damage. Two IR traits are examined here: those controlling the European Corn Borer (ECB), and those controlling corn rootworms (RW).[12] 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 (GF). Alt ogether these traits are embedded in a total of 12 different types of GM hybrids: four single-trait hybrids (ECB, GT, RW, and GF), four double-stack hybrids (ECB/GT, ECB/RW, ECB/GF, and GT/RW), three triple-stack hybrids (ECB/RW/GT, ECB/GT/GF, and ECB/RW/GF), and one quadruple stack hybrid (ECB/RW/GT/GF). 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.[13]

To investigate the potential selectivity bias associated with increasing market dominance

---

[10] While the specification $S_t \equiv \mathbf{z}_t\boldsymbol{\gamma}$ does not restrict S_t to be between 0 and 1, it does not seem to be an issue in our empirical analysis. Indeed, the results reported in table 2 below correspond to estimates of S_t that typically remain between 0 and 1.

[11] Our model applies to all biotech seeds in general. We focus on the corn hybrid case due to data availability.

[12] There are different GM events associated with GM seeds. 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. Our analysis focuses on broad GM traits (e.g. ECB, RW) and does not explore possible productivity differences across GM events across traits. Exploring such differences is a good topic for further research.

[13] As noted in footnote 12, biotech traits are associated with "genetic events" corresponding to the chromosomal location of the transgene.

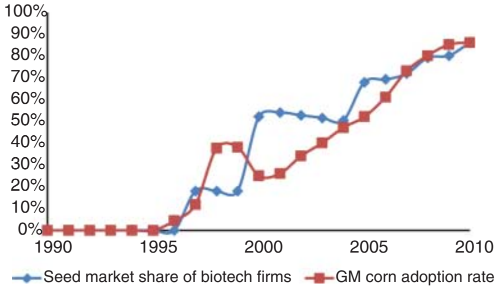


Figure 2. Seed market share of biotech firms and GM adoption rate in acreage, 1990–2010

of the biotech companies in the seed market, we use the aggregate seed market share of biotech companies involved in the corn seed market.[14] Following several waves of mergers and acquisitions since the mid-1990s, there are currently four such companies: Monsanto, DuPont (via the acquisition of Pioneer Hi-Bred International in 1999), Dow AgroSciences, and Syngenta. We constructed the aggregate market share of these four companies and their legacy companies (before mergers and acquisitions) when applicable from 2000 to 2007 using acreage data collected by **dmr**kynetec (DMR), of St. Louis, MO. Prior to the commercialization of biotech corn hybrids in 1996, no biotech firm was involved in corn seed production. Thus, we set the four biotech firms' corn seed market share at zero prior to 1997. From 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 from 2008 to 2010, we obtained information from the annual reports released by these firms.

Figure 2 shows the evolution of the hybrid corn seed market concentration by biotech/seed firms as measured by the seed market share of the vertically-integrated biotech firms, and also shows the GM corn adoption rate in terms of percentage of acreage over the last 21 years. DuPont's acquisition of Pioneer in 1999 increased biotech firms' aggregate seed market share from about 20% to over 50% in 2000. By 2010, these firms supplied about 86% of the U.S. corn seed market. Much of the expansion was through mergers and acquisitions of local seed companies by these vertically-integrated biotech firms. For GM adoption rate, the values from 1996 to 1999 are taken from Moss, Schmitz and Schmitz (2002), and those from 2000 onwards are from the USDA ERS.

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 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, the "event lag". These variables are intended to capture possible interaction effects between traditional genes and biotech genes, since they affect corn yield. 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 negative interaction effects over time 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/GF, GT/GF) had been on the market for at least 10 years. Triple-stack and quadruple-stack hybrids first entered the market in recent years, and thus have fewer event years than most single- and double-stack events.

Econometric Analysis

We use equations (11*a*) and (11*b*) 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, **M**: crop rotation, planting density, fertilizer use, irrigation, insecticide use, and spring/fall tillage. Finally, we control for the environmental variables, **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

---

[14] We define "biotech companies" as firms possessing patented GM trait technology. A company may have undertaken some biotech seed research and development, but will be classified in our definition as a "biotech" firm only if it has acquired patent(s) on its research output. Such a definition is empirically tractable and observable since patent information is publicly available while in-house innovation activities are often hidden from the public. Moreover, all the GM traits in the current commercial biotech seeds are patented.

Table 1. Summary Statistics

Variable	Number of observations	Mean	Standard Deviation	Min.	Max.
Yield (bushels/acre)					
Conventional	19,652	166.6	37.9	21	285.2
Glyphosate Tolerant (GT) single	972	182	37.9	56.2	276
Glufosinate Tolerant (GF) single	103	189.8	31.4	104.8	258
European Corn Borer (ECB) single	3,484	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	1,454	191.2	35.6	74.2	280.8
GT/RW double	166	204.6	35.6	109.9	268.7
ECB/GF double	998	197.6	39.9	65.7	285.5
ECB/RW/GT triple	3,215	202.8	31	95	288
ECB/GT/GF triple	631	201.1	36.4	98.3	283.9
ECB/RW/GF triple	206	209.1	31.7	104.8	285.1
ECB/GT/RW/GF quad.	797	208.5	33.2	78.5	289.8
Number of years since event introduction					
Glyphosate Tolerant (GT) single	972	7.1	2.8	1	13
Glufosinate Tolerant (GF) single	103	6.7	4.7	1	14
European Corn Borer (ECB) single	3,484	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	1,454	7.2	20.7	1	12
GT/RW double	166	3.3	1.2	1	5
ECB/GF double	998	7.9	3.7	1	15
ECB/RW/GT triple	3,215	2.6	1	1	6
ECB/GT/GF triple	631	3.9	1.1	1	5
ECB/RW/GF triple	206	4.2	2.2	1	8
ECB/GT/RW/GF quad.	797	3	1.3	1	5
Fertilizer (lbs/acre)	31,799	130.2	47.3	0.5	236.3
Planting density (1,000 seeds/acre)	31,799	28.5	1.9	18.3	33.4
Biotech firms' annual seed market share	14	0.55	0.23	0.18	0.86

Note: Corn seed market share of these firms is set to be zero prior to 1997 because there was no biotech seed in the market. The summary statistics for this variable is for 1997–2010 only.

location as well as across years). Controlling for these effects is important to ensure that the error term e_t in (6) or (11) represents 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 (11a)–(11b) provide a proper basis for our investigation. As noted, a key argument is that the error term e_t in (11) represents the unobserved distribution of the germplasm. The weather and pest population effects are captured through the time dummies, the location dummies, and their interactions. However, some GM traits (e.g. ECB or RW) contribute to yield by controlling 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 are being captured by the time/location dummies). While our econometric analysis provides a basis for evaluating selectivity effects related to GM hybrids, it should not be used to evaluate the productivity effects of GM traits.

Each biotech hybrid includes at least one genetic trait, each represented by a corresponding dummy variable(s). In the presence of stacking, biotech hybrids would include more than one GM trait. To capture stack effects, we also include corresponding stack dummies. Therefore, biotech genes, $\mathbf{G}_b$, in equation (11a)–(11b) are captured by both trait dummies and stack dummies when applicable.[15] Additionally, and as noted above, we also include "event lag" variables that measure the

---

[15] Since "conventional seeds" are not included among these dummy variables, coefficients of these dummies are interpreted as yield deviations from conventional seeds.

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 (11*a*) and (11*b*) by ordinary least squares with robust errors, the term σ_t in (11*b*) being estimated as the standard deviation of the error term obtained from (11*a*). We set the threshold level m to zero ($m = 0$), which corresponds to a below-the-mean threshold for selectivity in germplasm. Given $S_t \equiv \mathbf{z}_t \boldsymbol{\gamma}$, equation (11*b*) takes the form:

$$(11b') \quad Yield_t = g(\mathbf{G}_b^t, \mathbf{V}_t, \mathbf{M}_t, \boldsymbol{\beta}) + \mathbf{z}_t \boldsymbol{\gamma} \cdot \frac{4\sigma_t}{\sqrt{2\pi}} + u_t, \quad \text{when } \mathbf{G}_b^t \neq 0$$

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 case, (11*b*′) provides measurement of such effects on productivity assessment.

We consider three specifications for ($\mathbf{z}_t \boldsymbol{\gamma}$) in (11*b*′). In the first specification S1, $\mathbf{z}_t$, includes only a dummy variable for GM seeds, which allows us to investigate whether or not there is selection bias associated with all GM hybrids when compared to conventional hybrids. In the second specification, S2, $\mathbf{z}_t$ includes three variables: a constant, the 1-year lagged seed market share of the biotech firms, and the 1-year lagged GM adoption rate in percentage of acreage. The coefficients of the market share variable and adoption rate variable permit us to analyze whether selectivity bias may vary with changing market concentration and level of GM adoption. The first and second specifications restrict the selectivity bias to be the same across all GM traits. This restriction is relaxed in the third specification, S3, where we allow the selection bias to vary by 9 GM hybrid types: single-trait ECB; single-trait GT; and the stacked hybrids.[16] As in S2, the third specification also includes the 1-year lagged seed market share of the biotech firms, and the 1-year lagged GM adoption rate to capture possible effects of market concentration and level of adoption on selectivity.

We estimate equations (11*a*) and (11*b*′) jointly; however, the selectivity term $\mathbf{z}_t \boldsymbol{\gamma} \cdot \frac{4\sigma_t}{\sqrt{2\pi}}$ in (11*b*′) is relevant only when $\mathbf{G}_b^t \neq 0$. Effectively, the variables z_t in (11*b*′) are specified as interaction variables with a dummy variable GM, where $GM = 0$ if $\mathbf{G}_b^t = 0$ and $GM = 1$ if $\mathbf{G}_b^t \neq 0$. Table 2 shows the estimation results from equations (11*a*) and (11*b*′) under all three specifications. In all specifications, we found strong statistical evidence of selectivity bias. In S1, the coefficient of the *GM* dummy is 0.50; it is statistically significant at the 1% level, showing the presence of selectivity. In S2, the coefficient of the *GM* dummy is 0.36, while the coefficients of the seed market share for the integrated biotech firms and the GM adoption rate are positive and negative, respectively, and all are statistically significant at the 1% level. Increased market concentration appears to contribute to increasing selectivity bias, while a high rate of GM adoption decreases selectivity bias. As biotech firms grow, they tend to insert GM traits into better-quality germplasm. However, the incentive to be selective in their choice of germplasm is reduced when the conventional market is shrinking.

Specification 3 allows for selectivity effects to vary across types of GM trait. As shown in table 2, the effects of market share and GM adoption rate on selectivity remain similar. However, S3 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 S2 and conclude that S3 is more appropriate, as it captures heterogeneity in selectivity across different GM traits.

As can be seen from table 2, under S3 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/GF triple-stack hybrids, and the quadruple-stack hybrids. There is no statistical evidence of selectivity bias for ECB/GF double-stack hybrids and ECB/GF/RW triple-stack hybrids. The results demonstrate that selectivity bias is not uniform and varies across GM traits. We evaluate the extent of the different selectivity biases below.

---

[16] Single-trait RW and GF hybrids, and the double-stack ECB/RW hybrids are not included because they have a small number of observations.

Table 2. Estimation of Yield Effects and Selectivity Bias Using Robust Errors

	Specification 1		Specification 2		Specification 3	
Variables	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.
γ						
$\frac{4\sigma_t}{\sqrt{2\pi}} \cdot GM$	0.50***	4.58	0.36***	2.72	–	–
$\frac{4\sigma_t}{\sqrt{2\pi}} \cdot$ *ECB single*	–	–	–	–	0.55***	3.83
$\frac{4\sigma_t}{\sqrt{2\pi}} \cdot$ *GT single*	–	–	–	–	0.30*	1.78
$\frac{4\sigma_t}{\sqrt{2\pi}} \cdot$ *ECB/GT double*	–	–	–	–	0.40**	2.55
$\frac{4\sigma_t}{\sqrt{2\pi}} \cdot$ *GT/RW double*	–	–	–	–	0.75***	3.29
$\frac{4\sigma_t}{\sqrt{2\pi}} \cdot$ *ECB/GF double*	–	–	–	–	0.06	0.31
$\frac{4\sigma_t}{\sqrt{2\pi}} \cdot$ *ECB/RW/GT*	–	–	–	–	0.40***	2.81
$\frac{4\sigma_t}{\sqrt{2\pi}} \cdot$ *ECB/GT/GF*	–	–	–	–	0.51***	2.62
$\frac{4\sigma_t}{\sqrt{2\pi}} \cdot$ *ECB/GF/RW*	–	–	–	–	−0.25	−1.24
$\frac{4\sigma_t}{\sqrt{2\pi}} \cdot$ *ECB/GT/GF/RW*	–	–	–	–	0.53***	2.79
$\frac{4\sigma_t}{\sqrt{2\pi}} \cdot$ *lagged market share*	–	–	0.39***	3.93	0.39***	3.97
$\frac{4\sigma_t}{\sqrt{2\pi}} \cdot$ *lagged adoption rate*	–	–	−0.46***	−4.31	−0.44***	−4.07
Previous crop						
Alfalfa	23.3***	2.95	23.2***	2.94	23.2***	2.93
Wheat	15.6**	2.05	15.6**	2.05	15.5**	2.03
Soybean	−1.3	−0.15	−1.3	−0.15	−1.4	−0.17
Cucumber	−31.5**	−2.52	−41.9***	−4.93	−42.1***	−4.96
Lupine	−10.5	−1.03	−46.9***	−6.31	−47.0***	−6.31
Pea	−27.6***	−4.22	−27.9***	−4.26	−28.1***	−4.30
Potato	−62.8***	−14.36	−63.0***	−14.4	−63.2***	−14.45
Triticale	16.7**	2.35	−19.7***	−3.50	−19.7***	−3.51
GE traits						
GT (glysophate tolerant)	−2.2	−0.54	1.1	0.25	3.2	0.61
GF (glufosinate tolerant)	−5.6	−1.05	−0.1	−0.02	10.3**	2.35
ECB (European corn borer)	−13.5***	−3.55	−5.7	−1.22	−11.0**	−2.28
RW (Root worm)	−26.6***	−3.00	−2.37***	−2.68	−14.2*	−1.73
***Stacked GE* dummies**						
ECB/RW	17.3*	1.79	10.5	1.04	20.3*	1.89
ECB/GT	13.8**	2.51	6.0	0.94	8.1	1.12
GT/RW	11.1	1.11	8.5	0.83	−18.2	−1.45
ECB/GF	17.7***	2.92	9.1	1.19	13.6*	1.80
ECB/RW/GT	35.5***	3.29	25.8**	2.19	17.7*	1.68
ECB/GT/GF	12.7	1.48	0.1	0.01	−12.0	−1.24
ECB/GF/RW	45.6***	3.96	33.6***	2.63	40.8***	3.49
ECB/GT/GF/RW	34.8**	2.57	20.0	1.29	−2.5	−0.21
EVTyrs of GT	−4.6***	−5.49	−4.6***	−5.04	−4.7***	−4.99

(*Continued.*)

Table 2. Continued

	Specification 1		Specification 2		Specification 3	
Variables	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.
EVTyrs of GT squared	0.3***	6.01	0.4***	5.94	0.4***	5.89
EVTyrs of GF	−3.0*	−1.72	−3.3*	−1.86	−3.2*	−1.82
EVTyrs of GF squared	0.2*	1.80	0.2**	1.97	0.2**	2.02
EVTyrs of ECB	2.7***	4.13	1.5*	1.88	1.4*	1.72
EVTyrs of ECB squared	−0.2***	−4.85	−0.1***	−2.76	−0.1***	−2.81
EVTyrs of RW	2.4	0.53	2.8	0.61	3.7	0.82
EVTyrs of ECB_RW	3.2**	1.98	3.4**	2.09	2.6*	1.65
EVTyrs of ECB_GT	−2.6**	−2.56	−2.7***	−2.74	−2.8***	−2.73
EVTyrs of ECB_GT squared	0.2**	2.41	0.2***	2.85	0.2***	2.77
EVTyrs of GT_RW	2.5**	2.35	2.8**	2.56	3.0***	2.73
EVTyrs of ECB_GF	−1.6**	−2.42	−1.9***	−2.88	−1.8***	−2.75
EVTyrs of ECB_GF squared	0.1	1.38	0.1*	1.88	0.1*	1.82
EVTyrs of ECB_RW_GT	−1.2***	−3.57	−1.1***	−3.21	−1.1***	−3.05
EVTyrs of ECB_GT_GF	−0.2	−0.24	−0.002	−0.00	−0.03	−0.05
EVTyrs of ECB_GF_RW	−2.1***	−4.66	−2.0***	−4.47	−2.8***	−6.03
EVTyrs of ECB_GT_GF_RW	0.5	1.23	0.5	1.28	0.5	1.15
Density	0.5*	1.95	0.5*	1.95	0.5*	1.95
Fertilizer	0.1	0.88	0.1	0.88	0.1	0.88
Irrigation	136.5***	18.90	104.2***	13.87	104.5***	13.91
Insecticide	7.2***	3.20	7.2***	3.20	7.15***	3.16
Fall tillage	29.0***	7.44	29.0***	7.43	29.0***	7.43
Spring tillage	−2.3	−0.58	−2.3	−0.58	−2.3	−0.58
Constant	165.6***	11.04	165.6***	11.04	165.7***	11.04
R-squared	0.85		0.85		0.85	

Note: Statistical significance is noted by * at the 10% level, ** at the 5% level, and *** at the 1% level. To save space, the estimates of parameters associated with year dummies, location dummies and their interactions are omitted from the table but are available upon request. "*EVTyrs of XX*" is the number of years since gene event XX has been first introduced and commercialized in corn.

Implications

In this section, our empirical estimates are used to generate additional insights into sources of selection bias, and the changing patterns of selection bias over time. We focus on GM hybrids in general (S1 and S2), and for a selected GM hybrid with ECB single trait (S3) for which our data contain the most observations among all GM seeds. The time period under examination is from 1997 to 2010. The estimated effects of selectivity bias on corn yield (in bushels per acre) for GM seeds are presented in table 3.

For S2 and S3, we decompose the bias effects into three components, those associated with the GM trait(s), those associated with the market dominance of biotech firms, and those associated with the level of total GM adoption. From table 3, the selectivity bias shows a non-linear trend over time, first increasing, and then peaking in the early to mid-2000s, and then decreasing, especially since 2008. Such a pattern may be due in part to the offsetting effects of an expanding market share for biotech firms and an increasing GM adoption rate. Under S1, the selectivity bias generates a sizable yield difference for GM hybrids compared with conventional hybrids: the yield differential ranges from an average of 8.8 bushels per acre to 16.2 bushels per acre for GM hybrids. Recall that the average yield of conventional hybrids in our sample is 166.6 bushels per acre. The selectivity bias can raise GM corn yields by 5-10%. These estimates of the contribution of selectivity are relatively high compared to reported results for the total contribution of some GM traits (e.g. see Stanger and Lauer 2006).

When the total effects of bias are decomposed into separate components, results under S2 suggest that the expanding market share of biotech firms contributes to a strengthening of the bias effects. They also show that increased GM adoption has an offsetting effect on selectivity bias. From 1997 to 2010, the selection bias attributed to market concentration has increased, but the increase was balanced by the negative effect of the adoption rate on bias. Interestingly, the latter effect has dominated the former since 2005.

Under specification S3, for single-trait ECB hybrids, the total bias effects range from

Table 3. Decomposition of Selectivity Bias (bushels/acre) Over Time

				S1	S2				S3: Single trait ECB			
Year	σ_t	1-year lagged biotech-firm share	1-year lagged GM adoption rate	Total	GM trait	Biotech-firm share	GM adoption rate	Total	GM trait	Biotech-firm share	GM adoption rate	Total
1997	14.9	0	0.04	11.8	8.6	0	−0.5	8.2	13	0	−0.5	12.5
1998	16.9	0.18	0.12	13.4	9.8	1.9	−1.5	10.2	14.7	1.9	−1.4	15.2
1999	16.2	0.18	0.38	12.8	9.4	1.8	−4.5	6.7	14.1	1.8	−4.2	11.7
2000	15.2	0.18	0.38	12	8.8	1.7	−4.3	6.3	13.3	1.7	−4	11
2001	14.6	0.52	0.25	11.6	8.5	4.7	−2.7	10.5	12.7	4.7	−2.5	14.9
2002	19.5	0.54	0.26	15.4	11.3	6.5	−3.7	14.1	17	6.6	−3.5	20
2003	15.5	0.53	0.34	12.3	9	5.1	−3.9	10.2	13.5	5.1	−3.7	14.9
2004	17	0.51	0.4	13.5	9.9	5.5	−5	10.3	14.8	5.5	−4.7	15.6
2005	18.3	0.50	0.47	14.5	10.6	5.7	−6.3	10	16	5.7	−6	15.7
2006	19	0.68	0.52	15	11	8	−7.3	11.8	16.6	8	−6.9	17.7
2007	20.5	0.69	0.61	16.2	11.9	8.8	−9.2	11.5	17.9	8.8	−8.7	18
2008	11.1	0.72	0.73	8.8	6.4	5	−6	5.4	9.7	5	−5.7	9
2009	16	0.79	0.8	12.7	9.3	7.9	−9.4	7.7	14	7.9	−8.9	12.9
2010	18.5	0.80	0.85	14.6	10.7	9.2	−11.6	8.4	16.1	9.2	−11	14.4

9 bushels per acre to 20 bushels per acre. This result partially reflects the fact that ECB hybrids exhibit strong GM selectivity bias effects among all GM hybrids (see table 2). Table 3 shows that the GM trait selectivity effects for ECB are positive and large, with moderating effects generated by higher GM adoption rate effects in recent years.

Our analysis shows that because of selectivity bias, traditional genes' contribution to productivity may be greater than first anticipated. This result raises questions about the efficiency of current property rights supporting genetic improvements in agriculture. Over the last two decades, the shift toward patenting genes has stimulated private research on biotechnology and its applications to agriculture. Our findings that observed yield improvements often associated with GM technology are partially due to improvements in germplasm raise questions about whether current patent laws provide appropriate incentives to support investments in either agro-biotechnology or traditional breeding programs.

Concluding Remarks

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

Our analysis provides a framework to investigate selectivity issues related to how GM genes and conventional genes are combined in GM seeds. Building on the contributions made by Heckman, we present an empirical analysis of gene selectivity applied to GM corn yield in the United States. Selectivity arises when biotech/seed firms insert GM traits more frequently into high-quality germplasm than in lower-quality germplasm. We argue that such quality management schemes can increase profits for biotech/seed firms. However, because of the existence of such schemes, observed high GM corn yield should not be attributed only to GM genes, as there will be a selectivity bias in productivity evaluation. Identifying this selection effect is important when evaluating the true productivity of GM traits.

Using experimental data on Wisconsin corn yields from 1990 to 2010, our analysis uncovers evidence of selection bias. The results indicate that some of the observed yield gains associated with GM hybrids is attributable to conventional genes. This finding stresses the important role that traditional breeding still plays in productivity improvements for corn, even after the widespread adoption of GM corn hybrids. 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/GF). In other cases, these effects can be large (e.g. ECB and GT/RW). We also investigate how market concentration of biotech firms and the GM adoption rate 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. However, this positive effect is offset in part by the negative impact of the rising GM adoption rate in recent years. Our evidence of selectivity bias raises questions about whether current patent policy has shifted the incentives from investing in germplasm improvement to investing in GM trait development. To the extent that recent yield increases are due in part to improved germplasm, one must ask whether current patent policy provides appropriate incentives to invest in traditional breeding programs.

While our analysis was applied to experimental 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. Given that breeders often develop hybrids that are adapted to local agro-climatic conditions, our findings may not hold in different regions. Second, more work is needed to explore whether selectivity bias may also arise when evaluating the productivity of other GM crops (e.g. cotton, soybean). Third, our analysis focused on the broad effect of GM traits (e.g. ECB, RW) and did not explore possible differences related to the insertion of specific GM events. More research is needed to explore such differences. Fourth, we need to better understand the effects of GM technology on farmers' exposure to pest damages and weather risk. This seems particularly important given the prospects of future climate change. Finally, further research is needed to explore whether current property rights provide appropriate incentives to support investments in both agro-biotechnology and traditional breeding programs. These questions must be framed in a global context, where scientific progress can be uneven across industries and countries, with significant prospects for agricultural technology transfers both over space and across commodities.

References

Alston, J.M., and P.G. Pardey. 1996. *Making Science Pay: The Economics of Agricultural R&D Policy*. Washington, DC: AEI Press.

Alston, J.M. and R.J. Venner. 2002. The effects of the US Plant Variety Protection Act on wheat genetic improvement. *Research Policy* 31:527–542.

Bouis, H. 2007. The Potential of Genetically Modified Food Crops to Improve Human Nutrition in Developing Countries. *Journal of Development Studies* 43:79–86.

Chataway, J., and J. Tait. 2000. *PITA project: Policy influences on technology for agriculture: Chemicals, biotechnology, and seeds (Annex C11)*. Monsanto monograph. Targeted Social-economic Research (TSER) Programme, European Commission Project No. PL 97/1280, October.

Duvick, D.N. 1992. Plant Breeding: Past Achievements and Expectations for the Future. *Economic Botany* 40:289–97.

Duvick, D. 2005. The Contribution of Breeding to Yield Advances in Maize. *Advances in Agronomy* 86:84–145.

Fernandez-Cornejo, J. 2004. *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.

Fulton M. and K. Giannakas. 2001. Agricultural Biotechnology and Industry Structure. *AgBioForum* 4:137–151.

Goetz, S.J. 1992. A Selectivity Model of Household Food Marketing Behavior in Sub-Saharan Africa. *American Journal of Agricultural Economics* 74:444–452.

Graff, G.D., G.C. Rausser and A.A. Small. 2003. Agricultural Biotechnology's Complementary Intellectual Assets. *Review of Economics and Statistics* 85: 349–363.

Griliches, Z. 1957. Hybrid Corn: An Exploration in the Economics of Technological Change. *Econometrica* 25:501–522.

Griliches, Z. 1960. Hybrid Corn and the Economics of Innovation. *Science* 132: 275–280.

Heckman, J. 1979. Sample Selection Bias as a Specification Error. *Econometrica* 47: 153–161.

Herdt, R.W. 2006. Biotechnology in Agriculture. *Annual Review of Environment and Resource* 31:265–295.

Howard, P.H., 2009. Visualizing Consolidation in the Global Seed Industry: 1996–2008. *Sustainability* 1(4):1266–1287.

James, C. 2010. 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.

Johnson, N.L. and S. Kotz. 1970. *Distributions in Statistics: Continuous Univariate Distributions*. Boston: Houghton Mifflin.

Just, R., J. Alston and D. Zilberman, eds. 2006. *Regulating Agricultural Biotechnology: Economics and Policy*. Location: Springer Science+Business Media, LLC.

Kalaitzandonakes, N. and B. Bjornson. 1997. Vertical and Horizontal Coordination in the Agro-Biotechnology Industry: Evidence and Implications. *Journal of Agricultural and Applied Economics* 29(1): 129–139.

King, J.L., and D. Schimmelpfennig. 2005. Mergers, Acquisitions, and Stocks of Agricultural Biotechnology Intellectual Property. *AgBioForum* 8:83–88.

Marco, A.C. and G.C. Rausser. 2008. The Role of Patent Rights in Mergers: Consolidation in Plant Biotechnology. *American Journal of Agricultural Economics* 90(1): 133–151.

Moschini, G. 2010. Competition Issues in the Seed Industry and the Role of Intellectual Property. *Choices* 25(2). Available at http://www.choicesmagazine.org/magazine/print.php?article=120

Moss, C., T. Schmitz, and A. Schmitz. 2002. Differentiating GMOs and Non-GMOs in a Marketing Channel. *Florida Agricultural Experiment Station Journal*, Series No. XXXX, June 1.

Mussa, M., and S. Rosen. 1978. Monopoly and Product Quality. *Journal of Economic Theory* 18:301–317.

National Academies. 2010. *Impact of Genetically Engineered Crops on Farm Sustainability in the Unites States*. Washington, DC: The National Academies Press.

Nolan, E. and P. Santos. 2012. The Contribution of Genetic Modification to Changes in Corn Yield in the United States. *American Journal of Agricultural Economics* 94(5):1171–1188.

Qaim, M. 2009. The Economics of Genetically Modified Crops. *Annual Review of Resource Economics* 1:665–693.

Rausser, G.C., S. Scotchmer, and L.K. Simon. 1999. Intellectual Property and Market Structure in Agriculture. Working Paper Series 141901, Department of Agricultural & Resource Economics, University of California, Berkeley.

Ronald, P. 2011. Plant Genetics, Sustainability and Global Food Security. *Genetics* 1: 11–20.

Shi, G. 2009. Bundling and Licensing of Genes in Agricultural Biotechnology. *American Journal of Agricultural Economics* 91(1):264–274.

Shi, G., J.P. Chavas, and K.W. Stiegert. 2010. An Analysis of the Pricing of Traits in the U.S. Corn Seed Market. *American Journal of Agricultural Economics* 92(5):1324–1338.

Springer, N.M. and R.M. Stupar. 2007. Allelic Variation and Heterosis in Maize: How do Two Halves Make More than a Whole? *Genome Research* 17:264–275.

Stanger, T. F., and J. G. Lauer. 2006. Optimum Plant Population of Bt and Non-Bt Corn in Wisconsin. *Agronomy Journal* 98(4): 914–921.

Tirole, J. 1988. *The Theory of Industry Organization*. Cambridge, MA:The MIT Press.

U.S. Department of Agriculture, Economic Research Service. 2011. *Adoption of Genetically Engineered Crops in the U.S.: Extent of Adoption*. Washington, DC.

U.S. Department of Agriculture, National Agricultural Statistics Service . 2011. *Crop Production Summary*. Washington, DC.

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

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