Genetically Modified Crops
An Input Distance Function Approach[1]

Selected Paper
2008 Southern Agricultural Economics Association Annual Meeting
Dallas, TX
February 3-5

Keywords/Subject Areas: Production Economics, Genetically Modified Crops, Distance Function, Stochastic Frontier Analysis

By Justin G. Gardner[2] (Middle Tennessee State University), Richard F. Nehring (ERS) and Carl H. Nelson (University of Illinois at Urbana-Champaign)

Abstract:
GM crops have been widely adopted by U.S. Farmers. The welfare impact of this technology is not always clear cut, Fernandez-Cornejo *et al* (2005) for example found that the adoption of GM soybeans lead to increased off-farm income. In light of these findings it is necessary to estimate the impacts of GM crops using a whole-farm approach. Our initial findings indicate that GM crops do not contribute to the decline of traditional family farms. In addition we make a significant methodological impact by using the within transformation to remove unobserved individual effects (Mundlak, 1996) and estimate a distance function that is free of management bias. We show that the within transformation results in ML estimates that are identical to OLS estimates.

Introduction

Genetically modified organisms (GMOs) have been used in crop production for over a decade. According to Brookes and Barfoot (2005) the first genetically modified (GM) crop was the tomato in 1994, followed by GM soybeans. Currently the most common GMOs in agriculture crops can be classified into two groups: herbicide tolerant (HT) and insect resistant varieties of corn, cotton and soybeans. This technology was not developed via conventional crop breeding methods. Instead, a trait that is foreign to the organism was inserted into its genome. Roundup Ready soybeans are resistant to the herbicide glyphosate, which Monsanto markets under the brand name Roundup. Monsanto developed the Roundup Ready HT trait to serve as a complementary input to Roundup (Just and Hueth, 1993). Glyphosate resistance has also been inserted into corn and cotton. Insect resistance is achieved by inserting a gene from the bacteria *Bacillus thuringiensis* (*Bt*), which creates a toxin that affects Lepidoptera larvae. Currently the *Bt* trait has been inserted into corn and cotton to control the European Corn Borer, the Corn Rootworm, the Cotton Bollworm, the Pink Bollworm, the Asian Bollworm and the Tobacco Budworm.

[1] Working paper, please do not cite or circulate without permission form the authors.
[2] Contact author (jggardne@mtsu.edu)

GMOs are controversial. With the emergence of glyphosate resistant weeds (Gardner and Nelson, 2007) and recent accidental release of unapproved GM rice (Endres and Gardner, 2006) the debate over GM crops is not over. Therefore, it is prudent to continually re-assess the economic, environmental and regulatory issues regarding GM crops. There have been many studies on the welfare impacts of GM crops in US agriculture. In their review of the economics literature, Marra *et al*.(2002), broadly concluded that GM crops are profitable for U.S. farmers. However, some evidence suggests that GM crops may not be profitable (Bullock and Nitsi, 2001, Fernandez-Cornejo, et al., 2002). Our objective is to investigate the welfare impacts of GM crops with the goal of determining if GM crops are in fact profitable for US farmers and if not then why would farmers adopt a crop that is not profitable. Furthermore we seek to understand which farms have gained, or lost, the most from GM technology. *A priory* we suspect that farm-specific conditions such as location, farm size and pest pressure will have an effect on the profitability of adopting a GM crop. Specific farmer attributes such as education, off-farm employment and age may also have an effect. My idea is that these factors might help to identify which farmer types gain from adoption, and how economic gain varies by farmer type.

Using 2003 data, Hoppe and Banker (2006), classified 67.8% of US farms as limited resource, retirement, or lifestyle farms. These farms hold 44.4% of US farm assets and produce 8.1% of the US farm output. They tend to specialize in beef production, and the household of the principal operators do not rely on farm income. Small to medium size family farms (23.6% of all farms accounting for 18.9% of output) rely heavily on farm income, and large farms (8.8% of all farms accounting for 72.8% of output) rely almost exclusively on farm income. Understanding how emerging technology changes the distribution of farms is important, as farm distribution has important implications for farm policy. For example, Hoppe and Banker (2006) report that large farms tend to receive commodity-based support, while small farms tend to receive conservation-based support. If GM crops are only profitable for small and medium sized family farms, how will this change the cost of commodity programs? Thus, it is important to know how the technology impacts different types of farms.

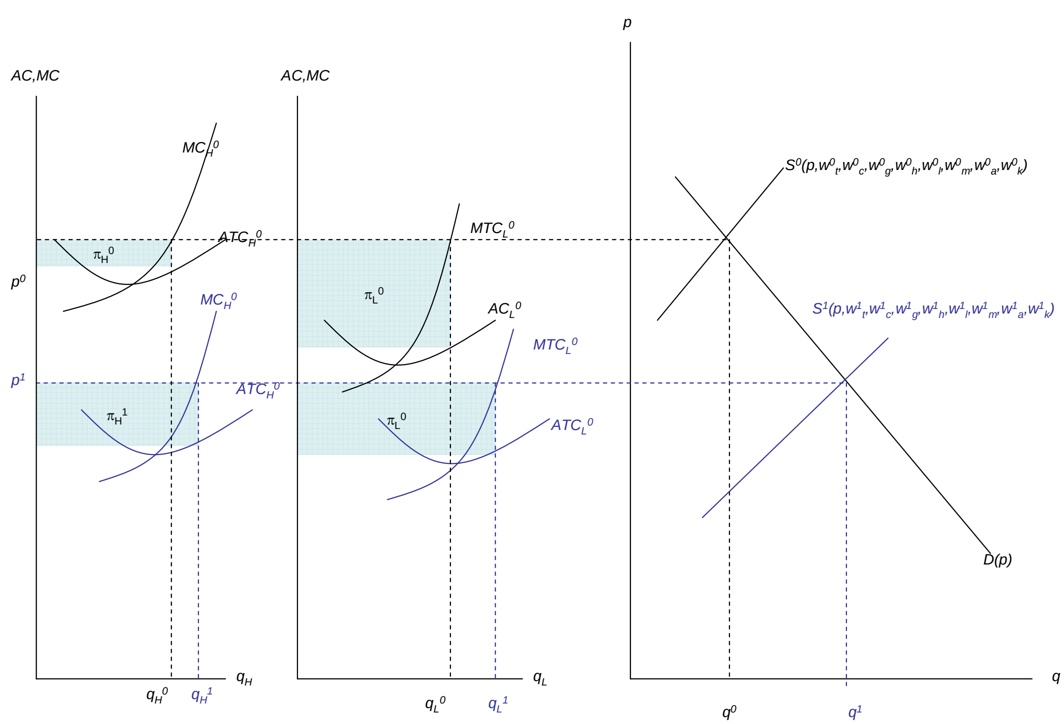
Theory

Consider two farms (see Figure 2.1.1), where one uses a high cost production method (denoted by *H*), while the other uses a low cost production method (denoted by *L*). In the initial equilibrium (Indexed by 0) GM inputs are not available. In the final equilibrium (Indexed by 1) both farms realize a decrease in costs relative to the initial conditions and both farms are assumed to adopt the technology. The market supply curve is the horizontal sum of the marginal costs curves of all the farms in the market. Thus, if farm costs decrease, the supply curve must shift out, lowering the market price (*p*). In Figure 1, after the introduction of the GM input, profits for both types of farms are about the same. As drawn the low-cost farmer has realized an increase in profit while the high-cost farmer's change in profit is small in absolute value. This is just one possibility; our objective is to devise a method to distinguish between low and high cost farmers, and then compare their welfare changes.

If production relationships, differentiated by farmer type, can be estimated, the research problem becomes a testable hypothesis. Consider a change in market price as illustrated in Figure 1, and compare the price change to the change in Average Total Cost (ATC). As shown by equation (1), the null hypothesis is that the changes to the market price and the average cost changes are equal. If the null is true then no farmer has gained; adoption does not guarantee a welfare increase. The alternative is that changes in average costs and price are not the same; only one type might gain.

(1)
$$\begin{aligned} H_0: &\quad \Delta ATC_H(q) = \Delta ATC_L(q) = \Delta p \\ H_a: &\quad \Delta ATC_H(q) \neq \Delta ATC_L(q) \neq \Delta p \end{aligned}.$$

Figure 1: High-Cost vs. Low-Cost Farmers



However, the problem is much more complex. Farms produce multiple outputs using multiple inputs. Consider an alternative way to express farm "type," the elasticity of scale for a firm with *K* outputs.

$$\varepsilon_s = \left(\sum_k \frac{\partial \ln C(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{g})}{\partial \ln y_k} \right)^{-1}, \quad (k = 1, 2, \cdots, K) \tag{2}$$

Where $\boldsymbol{y}$ is a vector of *K* outputs, $\boldsymbol{w}$ is a vector of input *M* prices and $\boldsymbol{g}$ is a vector of *I* GM crop variables. The elasticity of scale is the inverse of the elasticity of cost.

$$\varepsilon_c = \sum_k \frac{\partial \ln C(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{g})}{\partial \ln y_k}, \quad (k = 1, 2, \cdots, K) \tag{3}$$

The elasticity of cost is the percentage change in cost arising from a 1 percent increase in output. If ε_c is less than one (ε_s is greater than one) then the firm exhibits increasing returns to scale implying that average costs will decrease if the firm expands output. A farm that produces at minimum average cost is said to exhibit constant returns to scale and by differentiating ε_c with respect to the i^{th} GM crop variable,

$$\frac{\partial \varepsilon_c}{\partial g_i} = \sum_k \frac{\partial^2 \ln C(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{g})}{\partial \ln y_k \partial g_i} \tag{4}$$

one can investigate how ε_c changes when firms adopt GM crops[3]. If the derivative of ε_c with respect to the i^{th} GM crop variable in equation (4) is positive, then the percentage change in cost for a one percent increase in output increases as the GM adoption rate increases. Implying that the GM input makes output expansion more expensive thus favoring small farms; in other words, the optimal farm size would decrease. Alternatively, if the derivative were negative then the GM input would make output expansion less expensive. If farm size, measured by output, is used to differentiate farm type then the hypothesis that farm size has an influence on the returns to GM adoption can be phrased as:

$$H_0: \quad \frac{\partial \varepsilon_c}{\partial g_i} = 0, \quad H_a: \quad \frac{\partial \varepsilon_c}{\partial g_i} \neq 0 \tag{5}$$

If the alternative is true then GM crops will have an impact on the optimal farm size.

---

[3] In general, it is not possible to take this derivative, as GM adoption is not a continuous variable. This will be addressed below when we discuss pseudo-panel data.

Literature Review

To date the most comprehensive literature review of the welfare impacts of GM crops can be found in Marra (2001) and Marra *et al.* (2002) . The authors reached several broad conclusions regarding the literature on the current generation of GM field crops, *Bt* cotton is likely to be profitable in the cotton belt and reduces pesticide use, adopting *Bt* corn should provide a small yield increase, and in some cases adoption leads to significant increases in profit. For HT soybeans they conclude that cost savings should offset any revenue loss due to yield drag. These conclusions seem plausible as there are several effects that could induce a welfare gain. Bullock and Nitsi (2002) and Carpenter and Gianessi (2001) list four advantages of HT crops: 1) HT technology leads the farmer to substitute relatively less-expensive glyphosate for other herbicides;. 2) farmers realize a change in the shadow price of labor and management[4]; 3) due to glyphosate's effectiveness at killing larger weeds, weather induced spraying delays do not significantly affect weed control; 4) when farmers switch to HT technology substitution effects lead to a decrease in the price of alternative herbicides. The widespread adoption of GM crops may be evidence of a welfare gain. In 2005 Herbicide tolerant crops made up 87% and 60%, of U.S. soybean and cotton acreage respectively, while 35% of the corn acreage and 60% of cotton acres were insect resistant (Fernandez-Cornejo and Caswell, 2005). Bernard *et al.* (2004) found that farms in Delaware had yield increases and decreases in weed control costs when they adopted HT soybeans. So it would seem that adopting this technology results in a welfare gain for farmers. But, as noted above, some studies do not support this conclusion. The literature requires more examination before one can draw a conclusion from it.

Marra (2001) and Marra's *et al's.* (2002) evidence concerning the profitability of *Bt* cotton seems overwhelming. All of the 47 studies that were compiled indicated that *Bt* cotton is profitable. Only two HT cotton studies were compiled, both indicated that the technology was profitable, as did two studies where these two traits were "stacked". However, the GM corn and soybean evidence lacks the depth to be conclusive; the author(s) drew conclusions based on two studies for each GM crop. Studies using the United States Department of Agriculture's (USDA's) Agricultural Resource Management Survey (ARMS) were excluded from the analysis based on the argument that field level data, such as that collected in ARMS phase II, could not capture within-farm effects leading to an anti-GM bias.

[4] Farm labor and management has no market price, the price of a non-market good is its shadow price.

Fernandez-Cornejo *et al.* (2002) used ARMS data and concluded that HT soybean adoption did not have a statistically significant effect on farmer profit. The Fernandez-Cornejo *et al.* study made use of a flexible functional form to estimate a profit function and corrected for endogeneity using an instrumental variables method. Even though the study did not find a profit impact they did find a small positive yield impact. Bullock and Nitisi's (2001) study, which used a cost-minimizing simulation, found that GM soybean farmers are less profitable than their conventional counterparts. However, they did not take into account the labor and management savings that arise from convenience and timing factors, as these were not observable variables. This leads to a puzzling conclusion; it is uncertain whether or not HT soybeans are more profitable than conventional soybeans, but almost every farmer uses the technology. Perhaps the research community has been unable to measure an important component of farmers' welfare. Bernard *et al.* (2004), as mentioned before did find a positive impact in Delaware. However, they pointed out that Delaware soybean farms are larger then the national average, which may have biased their results. Fernandez-Cornejo *et al.* (2005) made an important advancement by explaining high adoption rates for roundup ready soybeans in spite of the technology's inability to increase farm profits. Using a household production framework they found a positive relationship between HT soybean adoption and off farm income. Their result suggests that adopting HT soybeans can free up resources for alternative uses. In addition the result highlights the importance of properly modeling farm production; farms are multi-output producers that generate off-farm income as well as commodity-derived on-farm income.

Data

We used data collected via the USDA's ARMS Cost and Returns Report (CRR) survey. CRR data are collected at the farm level and contain information on farm revenue, by crop and on input expenditures. This data also includes classification variables such as USDA farm typologies (Hoppe, *et al.*, 2000) which have been used to create cohort classifications, thus allowing the annual cross sections to be treated as if they were panel data (Nehring, et al., 2005, Paul and Nehring, 2005, Paul, et al., 2004). Starting in 1998 the CRR surveys began asking respondents about GMO usage. However, this information was collected in the 1996 and 1997 Production Practices Report (PPR) surveys, and therefore it is possible to use survey-based imputation to infer the specifics of GMO adoption (Nehring, et al., 2005, Paul, et al., 2004).

In addition to pooling ARMS cross sections from 1996 to 2005 we also create a pseudo-panel using the cohort definitions shown in table 1. These cohorts have been used extensively with ARMS data and the data has been shown to be robust to cohort specification (Paul, *et al.*, 2004). One of the objectives of this study is to compare all three of the primary GM crops, thus the geographic scope has been widened beyond the Corn Belt that is traditionally studied to include cotton-growing regions. This presents a problem with respect to cohort size. The number of individuals in each cohort must be large to prevent a measurement error bias (Verbeek and Nijman, 1992), and many states outside of the corn belt have low sampling rates. To overcome this problem, under-sampled states were combined with their geographic neighbors. The states that stand alone are Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Carolina, Ohio, South Dakota, Texas and Wisconsin. The remainder of the states were grouped into one of six regions, Southeast (Alabama, Georgia, South Carolina), Delta (Arkansas, Louisiana, Mississippi), Mid-South (Kentucky, Tennessee, Virginia), Northeast (Maine, New Hampshire, Vermont, New York, Massachusetts, Connecticut, Rhode Island, Delaware, Pennsylvania, West Virginia, New Jersey), Southwest (California, Nevada, Arizona, New Mexico, Colorado, Utah), and Northwest (Washington, Oregon, Idaho, Montana, Wyoming).

The pseudo panel covers 10 years, 1996-2005 and twenty geographic regions with 13 cohorts in each region. The pseudo sample size is 2600. The smallest annual average cohort size is 21.66, in 2001, while the largest is 57.57 in 2004.

Table 1: Cohort Definitions

Cohort	USDA Farm Typology	Gross Value of Sales
1	1-3	0-2499
2	1-3	2,500-29,999
3	1-3	>29,999
4	4	0-9.999
5	4	10,000-29,999
6	4	>29,000
7	5	<175,000
8	5	>174,999
9	6	250,000-329,999
10	6	330,000-409,999
11	6	>409,999
12	7-8	0-999,999
13	7-8	>999,999

Estimation

In order to estimate the impact of genetically modified crops in a multi-output multi-input setting, if one cannot observe the allocation of inputs among crops, it is necessary to use either an input or an output distance function; we choose the former. The derivation of the input distance function and its properties can be found in many sources (Coelli, et al., 1998, Fare and Primont, 1995, Irz and Thirtle, 2004, Kumbhakar and Lovell, 2000, Lovell, et al., 1994, Nehring, et al., 2005, Paul and Nehring, 2005, Paul, et al., 2004, Shephard, 1953, Shephard, 1970). Let $L(y')$ denote the input requirement set, which is the set of all inputs that can feasibly produce output y', or

(6) $$L(\boldsymbol{y}') = \{(\boldsymbol{x}, \boldsymbol{y}) : \boldsymbol{x} \text{ can be used to produce } \boldsymbol{y}'\}.$$

The input distance function is defined as[5]:

(7) $$D(y, x) = \max_{\lambda}\left(\lambda > 0 : st\left(\frac{x}{\lambda}\right) \in L(y)\right).$$

It is the maximum amount by which the input bundle can be scaled down and still produce y'.

The input distance function is nondecreasing and concave in *x*, and nonincreasing in y. It is homogenous of degree one, also known as linear homogenous, in inputs.

(8) $$D(\boldsymbol{y}, \rho\boldsymbol{x}) \equiv \rho D(\boldsymbol{y}, \boldsymbol{x}), \quad \rho > 0$$

We will exploit this fact when we parameterize an estimable functional form. If the underlying production technology exhibits constant returns to scale then the distance function is also homogenous of degree one in outputs.

(9) $$D(\rho\boldsymbol{y}, \boldsymbol{x}) = \rho D(\boldsymbol{y}, \boldsymbol{x}), \quad \rho > 0$$

The most significant problem associated with estimating a distance function is that distance is not an observable quantity. Therefore, one can not simply specify a functional form and estimate the model. Lovell *et al.* (1994) demonstrated how one can use the distance function's homogeneity property and the fact that distance functions are always greater than one to estimate the distance function using stochastic frontier analysis (SFA). In terms of a single-input single-output production model a SFA model is expressed as follows:

(10) $$y = \alpha + x\beta + v - u,$$

[5] Typically, the literature uses a subscript or superscript "i" to denote the input distance function, and an "o" to denote the output distance function. As we are only using an input distance function the notation is not necessary and has been dropped.

where the error term has two components, v is the “standard” identically and independently distributed regression error term with mean zero, and u is a “one-sided” error term. The purpose of the one-sided error term is to shift the regression estimate from the middle of the dataset to the edge of the dataset, allowing one to estimate a frontier. A SFA model allows for two effects that might cause an individual to deviate from the most efficient input-output combination. The stochastic effect, arising from random shocks due to things like weather, is captured by the standard regression error term. The inefficiency effect, resulting from non-random effects such as location or managerial ability, is captured by the one-sided error term.

Following Lovell *et al.* (1994) one can use linear homogeneity in inputs, $D(\rho \boldsymbol{x}, \boldsymbol{y}) = \rho D(\boldsymbol{x}, \boldsymbol{y})$, setting $\rho = 1/x_1$, where x_1 is any arbitrary input. So that

(11) $$D(\boldsymbol{x}^*, \boldsymbol{y}) = \frac{1}{x_1} D(\boldsymbol{x}, \boldsymbol{y})$$

where an asterisk (*) denotes a normalized input. Recalling that $D(\boldsymbol{x}, \boldsymbol{y}) \geq 1$, equation (11) can be partially parameterized by setting $D(\boldsymbol{x}, \boldsymbol{y}) = 1$ and including an error term that is always positive.

(12) $$D(\boldsymbol{x}^*, \boldsymbol{y}) = \frac{1}{x_1} e^{\varepsilon}, \quad \varepsilon \geq 0 .$$

Rearranging (12) yields:

(13) $$\frac{1}{x_1} = D(\boldsymbol{x}^*, \boldsymbol{y}) e^{-\varepsilon}, \quad \varepsilon \geq 0 .$$

Taking the natural logarithm of (13) the partially parameterized model can be expressed as a linear model.

(14) $$\ln \frac{1}{x_1} = \ln D(\boldsymbol{x}^*, \boldsymbol{y}) - \varepsilon, \quad \varepsilon \geq 0$$

The error term can then be decomposed into a standard term and a one-sided term,

(15) $$\ln \frac{1}{x_1} = \ln D\left(\boldsymbol{x}^*, \boldsymbol{y}\right) + v - u, \quad u \geq 0$$

After assuming a functional form, equation (15) takes the same form as equation (10) and a distance function can be estimated using SFA.

The literature (Coelli and Perelman, 2000, Irz and Thirtle, 2004, Nehring, et al., 2005, Paul and Nehring, 2005, Paul, et al., 2004) has adopted a translog functional form in favor of a simpler Cobb-Douglas functional form. Coelli and Perlman (2000), who estimate an output distance function, argue that the translog functional form is superior to the Cobb-Douglas because it is flexible and it is easy to impose homogeneity restrictions. The Cobb-Douglas does have drawbacks, unlike the translog it will not allow for the calculation of scope economies, which are based on second order interaction terms (Paul, *et al.*, 2004). The primary disadvantage of the translog model is the large number of estimated parameters, it is thus plagued by multicollinearity problems and a large number of insignificant parameters. Pervious studies (Irz and Thirtle 2004) have shown that the translog is preferable to the Cobb-Douglas although it may not make a significant difference when answering the research question.

Augmenting the normalized distance function with a vector ***g*** of GM crop variables equation (15) becomes

(16) $$\ln \frac{1}{x_1} = \ln D\left(\boldsymbol{x}^*, \boldsymbol{y}, \boldsymbol{g}\right) + v - u, \quad u \geq 0$$

The Cobb-Douglas functional form with GM interaction terms is:

(17) $$\begin{aligned} \ln D(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{g}) &= \alpha_0 + \sum_m \alpha_m \ln x_m^* + \sum_k \beta_k \ln y_k + \sum_r \gamma_r g_r \\ &+ .5 \sum_r \sum_m \gamma_{rm} g_r \ln x_m^* + .5 \sum_r \sum_k \gamma_{rk} g_r \ln y_k + v - u \end{aligned},$$

According to Irz and Thirtle (2004) the log derivative of the input distance function with respect to the m[th] input is that input's cost share,

(18) $$\varepsilon_m = \frac{\partial \ln D(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{g})}{\partial \ln x_m} = \frac{w_m x_m}{C(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{g})} = S_m ,$$

which can be interpreted as the m[th] input's relative importance in the production process. A cost share is by definition positive. Alternatively, the negative of the input cost share is the shadow value of the m[th] input relative to the normalizing input, x_1 (Paul *et al.* 2004). The log derivative of the input distance function with respect to the k[th] output,

(19) $$\varepsilon_k = \frac{\partial \ln D(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{g})}{\partial \ln y_k} = -\frac{\partial \ln C(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{g})}{\partial \ln y_k}$$

is the negative cost elasticity of that output. It is expected to be negative for all desirable outputs and, in absolute value, reflects the relative importance of each output. One can also compute the elasticity of costs (SEC) (Paul *et al.* 2004), which is the sum of the individual cost elasticities.

$$SEC = -\varepsilon = -\sum_k \varepsilon_k = -\sum_k \frac{\partial \ln D(\boldsymbol{x},\boldsymbol{y},\boldsymbol{g})}{\partial \ln y_k} = \sum_k \frac{\partial \ln C(\boldsymbol{y},\boldsymbol{w},\boldsymbol{g})}{\partial \ln y_k}. \tag{20}$$

The derivative of the distance function with respect to the GM variables,

$$\varepsilon_i = \frac{\partial \ln D(\boldsymbol{y},\boldsymbol{x},\boldsymbol{g})}{\partial \ln g_i}, \tag{21}$$

can be used to determine if GM crops increase efficiency, if (21) is negative then increasing the GM adoption rate decreases the distance to the efficiency frontier.

One drawback of using the distance function to estimate the elasticity of cost is that doing so requires a behavioral assumption, cost minimization, or the equivalent assumption that the firms are price efficient. Rather than using the elasticity of cost one can calculate the elasticity of scale. Following Fare and Primont (1995) one can specify a distance function where the inputs and outputs are scaled by the fixed factors ρ and θ.

$$D(\rho\boldsymbol{x},\theta\boldsymbol{y}) = 1 \tag{22}$$

The elasticity of scale is the percentage change in θ for a percentage change in ρ. In other words it answers the question, if producers increase all inputs by some fixed amount how much will outputs increase?

$$\varepsilon_s = \left.\frac{\partial \ln \theta}{\partial \ln \rho}\right|_{\rho=\theta=1} \tag{23}$$

By applying the Implicit Function Theorem, invoking Euler's Theorem and recalling that the distance function is equal to one when the firm is operating optimally Fare and Primont (1995) show that the elasticity of scale can be calculated as follows:

$$\varepsilon_s = -\frac{\sum_n \frac{\partial D(\boldsymbol{x},\boldsymbol{y})}{\partial x_n} x_n}{\sum_m \frac{\partial D(\boldsymbol{x},\boldsymbol{y})}{\partial y_m} y_m} = -\frac{D(\boldsymbol{x},\boldsymbol{y})}{\sum_m \frac{\partial D(\boldsymbol{x},\boldsymbol{y})}{\partial y_m} y_m} = -\left(\sum_m \frac{\partial D(\boldsymbol{x},\boldsymbol{y})}{\partial y_m} y_m\right)^{-1}. \tag{24}$$

Note that the elasticity of scale is the inverse of the elasticity of cost. The problem with using the elasticity of scale is that it assumes firms are able to scale all inputs. This is not realistic in agriculture as farms may not be able to increase items like land and capital equipment. Therefore, we propose to use a modified version of equation (24) where only the variable inputs, denoted by a *v* subscript, are scaled.

(25) $$\varepsilon_s' = -\frac{\sum_v \frac{\partial D(\boldsymbol{x},\boldsymbol{y})}{\partial x_v} x_v}{\sum_m \frac{\partial D(\boldsymbol{x},\boldsymbol{y})}{\partial y_m} y_m}.$$

Because the numerator in (25) will always be less than or equal to one, the modified scale measure will always be less than or equal to the elasticity of scale, $\varepsilon_s' \le \varepsilon_s$. We can take the derivative of equation (25) with respect to GM crops in order to determine if GM crops increase the elasticity of scale.

(26) $$\frac{\partial \varepsilon_s'}{\partial g_i} = \frac{\dfrac{\partial\left(\sum_m \frac{\partial D(\boldsymbol{x},\boldsymbol{y},\boldsymbol{g})}{\partial y_m} y_m\right)}{\partial g_i}}{\left(\sum_m \frac{\partial D(\boldsymbol{x},\boldsymbol{y},\boldsymbol{g})}{\partial y_m} y_m\right)^2}$$

If equation (26) is positive then we can conclude that the i^{th} GM technology favors large farms.

A major pitfall when estimating a production function is endogeneity. To illustrate this point Mundlak (2001, 1996) used a single input Cobb-Douglas model, and assumed that panel data are available, to show the source of endogeneity. Equation (30) illustrates Mundlak's (2001, 1996) model:

(27) $$Y_{it} = AX_{it}^{\beta} e^{m_{0i}+u_{0it}}$$

This model has an individual fixed effect, m_0, that does not vary over time, it is unobservable and correlated with the input *X*. As the individual fixed effect is not observed it manifests itself in the error term. The errors are thus correlated with the independent variable, giving rise to endogeneity. When panel data are available endogeneity can be eliminated by using the within estimator. This is done by averaging the variables for the i^{th} individual over time and mean differencing the data. Thus any endogeneity arising from unobserved management ability disappears from the model. The primary motivation for building the pseudo panel is to correct for endogeneity. Consider a distance function, which could take on any arbitrary functional form, expressed as an estimable model:

(28) $$-x_{ct,1} = D\left(\boldsymbol{x}_{ct}^{*}, \boldsymbol{y}_{ct}, \boldsymbol{g}_{ct}\right) + m_{ct} + v_{ct} - u_{ct}.$$

Unlike the distance function in equation (16), this model includes an additional error term, m_{ct}. This error term, described by Mundlak (1996; 2001), is an individual fixed effect that is assumed to be constant over time,

(29) $$m_{ct} = m_c$$

it is unobservable by the researcher and correlated with the inputs ***x***.

(30) $$Cov(x_{ct}, m_{ct}) \neq 0, .$$

As the individual fixed effect is not observed, it manifests itself in the error term. The errors are thus correlated with the independent variables, giving rise to endogeneity.

The model also includes the "traditional" error term, v_{ct}, which is not correlated with the inputs and is assumed to be identically and independently distributed with mean zero.

(31) $$cov(x_{ct}, v_{ct}) = 0, \quad v_{ct} \sim N(0, iid)$$

Further complicating matters, Lovell *et al.'s* (1994) exposition demonstrated that one could use SFA to estimate a distance function. Lovell *et al.'s* (1994) method leads to the third error term in equation (28). Conceptually, this error term will "shift" a line fitted by OLS from the center of the data so that the fitted line sits on the edge, or frontier of the data. The only assumption about u_{ct} is that it is always positive.

(32) $$u_{ct} > 0$$

As shown by Mundlak (1996; 2001) we can correct for endogeneity by using the within transformation. In addition Verbeek and Nijman (1992) demonstrated that, when n_c is large, it is safe to assume that the cohort fixed effect, m_{ct}, would remain constant through time and hence the within estimator will remove m_{ct} from the error term because

(33) $$m_{ct} - \frac{\sum_t m_{ct}}{T} = 0 .$$

In this case the within transformation also has an impact on the one-sided error term.

(34) $$v_{ct} - \frac{\sum_t v_{ct}}{T} = \phi_{ct} .$$

The one-sided error term, which was added to the model in order to shift the fitted model to the frontier of the data, is absorbed into the traditional error term. In effect, the fitted model is shifted back to the middle of the data and frontier estimates converge to OLS estimates. Therefore, a distance function can be estimated via OLS when panel data is available. By analogy, if a researcher is using panel data to estimate a frontier production function, correcting for endogeneity using the within transformation will result in the same parameter estimates as a traditional production function that has been corrected for endogeneity using the within transformation.

Results

A pseudo-panel may be rendered invalid if cohort members are able to switch cohorts. In a footnote Paul *et al*. (2004) suggest that a pseudo-panel can be validated by performing a regression using the pooled observations. Hence we will proceed by estimating a pooled distance function model using SFA methods, and then compare the results to a pseudo-panel that has been mean-differenced so as to estimate a within model. In addition we will compare the results of OLS and SFA versions of a within model. Outputs and inputs are expressed in natural logarithms, inputs have been normalized by the natural logarithm of the dependant variable, land. Variable definitions can be found in table 2. The estimated models can be found in table 3. Inputs are the total expenditure on the input, and outputs are total revenues from the output. All inputs are normalized using the dependant variable, land. Land and capital were converted from stocks to annualized flows. In a pseudo-panel, dummy variables are interpreted as the percentage of individuals in the cohort for which the variable is equal to one. i.e an education dummy variable in the pooled data become the percentage of farmers within a cohort that has at least a high school education.

The pooled model was estimated using the survey weights supplied with the ARMS data, and variances were estimated using Huber-White Standard errors. We compare OLS and SFA versions of the within-transformed input distance function to demonstrate that, given the within transformation, stochastic frontier and OLS estimates are identical. Stochastic frontier and OLS models were estimated using Stata's frontier and regress commands respectively.

When using robust or cluster standard errors Stata's frontier command will not provide a χ^2 statistic in order to test the null hypothesis that $u = 0$, therefore, to illustrate this test, we also include a frontier within estimate with non-robust standard errors. If the null is rejected then the one-sided error term is statistically significant and the model cannot be estimated via OLS. For the SFA model using the within transformation we fail to reject the null, implying that we can use OLS. When a robust variance estimator is used the estimated value of u is zero. The reader will notice that the parameter estimates for the OLS and SFA within models are identical. This is because the likelihood function for SFA within is just the normal likelihood function where β is obtained by minimizing the sum of squared errors.

For the within model with the exception of off-farm income (E), all of the output coefficients are statistically significant. All of the output coefficients are negative and all of the input coefficients are positive, which is a marked improvement over the pooled and the pseudo-panel estimates. When a non-robust variance estimator is used fuel, (f) is the only output that is not statistically significant at the five percent level. For the robust variance models fertilizer (n) is statistically significant at the ten percent level and off farm employment (j) is statistically significant in the OLS cluster-variance model. These results highlight the true value of creating a pseudo-panel. The input and output cost shares, computed using Stata's lincom command, are reported in table 5.

Table 2: Variable Definitions, Input Distance Function

Outputs		Inputs	
M	Corn	s	Seed
C	Cotton	n	Fertilizer
L	Livestock	l	Labor
O	Other Crops	c	Other Crop Expenses
E	Off-Farm Income	o	Other Miscellaneous Expenses
		a	Feed and Animal Expenses
		f	Fuel and Electricity
		e	Capital Equipment
		j	Off-farm Employment
		Other Variables	
s1-s19	State or Region Dummy Variables	Mbtm	Corn**Bt* Corn interaction term
y1-y9	Year Dummy Variables	Mhtm	Corn*HT Corn interaction term
abtcorn	Acres of *Bt* Corn	Cbtc	Cotton**Bt* Cotton interaction term
abtcotton	Acres of *Bt* Cotton	Chtc	Cotton*HT Cotton interaction term
ahtcorn	Acres of HT Corn	Shts	Soybean*HT Soybean interaction term
ahtcotton	Acres of HT Cotton	OP_AGE	Age of the Primary Opertor
ahtsoy	Acres of HT Cotton	vsmall	Very Small Farm Dummy Variable
hs	Completed High School Dummy Variable	small	Small Farm Dummy Variable
crop	Crop Farm Dummy Variable	large	Large Farm Dummy Variable

All of the input shares have the hypothesized sign; fertilizer is the only input share that is not statistically significant. Following Irz and Thirtle, it is possible to calculate an input share for the normalizing input, labor. Unlike Irz and Thirtle we are able to report a significance level for the normalizing input. The results show that the most important inputs, the ones with the largest cost shares, are labor and land, which together make up 50 percent of all input costs, this is a predictable result. However, the next largest cost share is "other miscellaneous expenses"

Table 3: Input Distance Function Results[3]

Dependant Variable: land	Pooled SFA[1] Huber-White Robust Standard Errors	OLS[2] Within	SFA[1] Within	OLS[2] Cluster Standard Errors	SFA[1] Cluster Standard Errors	SFA[1] Robust Standard Errors
M	-0.037	-0.023	-0.023	-0.023	-0.023	-0.023
	(14.12)***	(6.89)***	(6.94)***	(6.85)***	(6.90)***	(4.88)***
C	-0.061	-0.009	-0.009	-0.009	-0.009	-0.009
	(15.06)***	(3.14)***	(3.16)***	(2.31)**	(2.33)**	(2.55)**
L	-0.040	-0.046	-0.046	-0.046	-0.046	-0.046
	(18.70)***	(8.59)***	(8.65)***	(2.42)**	(2.43)**	(3.50)***
O	-0.043	-0.038	-0.038	-0.038	-0.038	-0.038
	(22.62)***	(9.65)***	(9.72)***	(4.03)***	(4.06)***	(5.52)***
S	-0.049	-0.017	-0.017	-0.017	-0.017	-0.017
	(15.61)***	(4.99)***	(5.02)***	(6.31)***	(6.36)***	(3.10)***
E	-0.080	-0.004	-0.004	-0.004	-0.004	-0.004
	(32.14)***	(0.94)	(0.94)	(0.59)	(0.60)	(0.67)
s	0.041	0.031	0.031	0.031	0.031	0.031
	(13.93)***	(4.25)***	(4.28)***	(4.18)***	(4.21)***	(2.98)***
n	-0.001	0.028	0.028	0.028	0.028	0.028
	(0.28)	(3.05)***	(3.08)***	(1.84)*	(1.85)*	(1.76)*
p	0.027	0.056	0.056	0.056	0.056	0.056
	(10.13)***	(6.44)***	(6.49)***	(2.82)**	(2.84)***	(3.29)***
l	0.242	0.310	0.310	0.310	0.310	0.310
	(17.93)***	(17.91)***	(18.04)***	(3.40)***	(3.43)***	(8.16)***
c	0.028	0.014	0.014	0.014	0.014	0.014
	(13.67)***	(3.31)***	(3.34)***	(2.46)**	(2.48)**	(2.41)**
o	0.252	0.154	0.154	0.154	0.154	0.154
	(16.52)***	(8.02)***	(8.08)***	(6.20)***	(6.24)***	(5.37)***
a	0.033	0.047	0.047	0.047	0.047	0.047
	(10.17)***	(8.29)***	(8.35)***	(2.47)**	(2.49)**	(3.94)***
f	-0.037	0.012	0.012	0.012	0.012	0.012
	(6.84)***	(1.08)	(1.08)	(0.63)	(0.63)	(0.74)
e	0.014	0.127	0.127	0.127	0.127	0.127
	(2.60)***	(8.50)***	(8.57)***	(2.71)**	(2.72)***	(3.55)***
j	0.345	0.013	0.013	0.013	0.013	0.013
	(53.26)***	(2.99)***	(3.01)***	(2.06)*	(2.07)**	(3.14)***
abtcorn	-0.000	-0.001	-0.001	-0.001	-0.001	-0.001
	(3.87)***	(5.70)***	(5.75)***	(6.39)***	(6.43)***	(3.25)***
abtcotton	0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.77)	(0.82)	(0.83)	(0.60)	(0.61)	(0.74)
ahtcorn	0.000	0.001	0.001	0.001	0.001	0.001
	(0.15)	(1.39)	(1.40)	(1.37)	(1.38)	(1.48)

* significant at 10%; ** significant at 5%; *** significant at 1%
1. z-statistics in parenthesis
2. t-statistics in parenthesis
3. Year and state dummy variables have been omitted from the pooled model due to space considerations.

Table 3, Continued[3]

Dependant Variable: land	Pooled SFA[1] Huber-White Robust Standard Errors	OLS[2] Within	SFA[1] Within	OLS[2] Cluster Standard Errors	SFA[1] Cluster Standard Errors	SFA[1] Robust Standard Errors
ahtcotton	-0.000	-0.001	-0.001	-0.001	-0.001	-0.001
	(1.34)	(2.12)**	(2.13)**	(1.22)	(1.23)	(1.52)
ahtsoy	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(1.56)	(0.67)	(0.67)	(0.60)	(0.61)	(0.61)
hs	-0.158	0.016	0.016	0.016	0.016	0.016
	(7.11)***	(0.59)	(0.60)	(0.44)	(0.44)	(0.50)
crop	0.350	0.003	0.003	0.003	0.003	0.003
	(14.91)***	(0.19)	(0.19)	(0.14)	(0.14)	(0.19)
OP_AGE	0.008	-0.002	-0.002	-0.002	-0.002	-0.002
	(10.77)***	(3.51)***	(3.54)***	(2.92)**	(2.94)***	(3.24)***
Mbtm	0.011	0.009	0.009	0.009	0.009	0.009
	(3.89)***	(1.95)*	(1.96)**	(3.03)**	(3.05)***	(1.85)*
Mhtm	-0.001	0.000	0.000	0.000	0.000	0.000
	(0.24)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)
Cbtc	-0.001	-0.006	-0.006	-0.006	-0.006	-0.006
	(0.07)	(0.27)	(0.27)	(0.25)	(0.25)	(0.22)
Chtc	0.020	0.009	0.009	0.009	0.009	0.009
	(2.19)**	(0.52)	(0.52)	(0.47)	(0.47)	(0.45)
Shts	0.002	-0.011	-0.011	-0.011	-0.011	-0.011
	(0.59)	(3.15)***	(3.18)***	(1.89)*	(1.90)*	(2.57)**
vsmall	0.278					
	(13.08)***					
small	-0.022					
	(1.14)					
large	-0.136					
	(5.40)***					
Constant	-5.372	N/A	N/A	N/A	N/A	N/A
	(44.48)***	N/A	N/A	N/A	N/A	N/A
Observations	105883.83	2574	2574	2574	2574	2574
R-squared	N/A	0.74	N/A	0.74	N/A	N/A
Overall F	N/A	N/A	7241.21	N/A	N/A	N/A
P>F	N/A	N/A	0.000	N/A	N/A	N/A
Overall Chi2	105883.83	187.75	N/A	N/A	N/A	N/A
P>Chi2	0.000	0.000	N/A	N/A	N/A	N/A
sigmav	N/A	N/A	0.23	N/A	0.23	0.23
sigmau	N/A	N/A	0.00	N/A	0.00	0.00
$\chi 2$	N/A	N/A	0.00	N/A	N/A	N/A
P>Chi2	N/A	N/A	1.00	N/A	N/A	N/A

* significant at 10%; ** significant at 5%; *** significant at 1%

1. z-statistics in parenthesis
2. t-statistics in parenthesis
3. Year and state dummy variables have been omitted from the pooled model due to space considerations.

Table 4: Output and Input Shares, Within Model

Outputs		Inputs			
M	-0.023	s	0.031	a	0.047
	(4.07)***		(4.21)***		(2.49)**
C	-0.009	n	0.028	f	0.012
	(2.57)***		(1.85)*		(0.63)
L	-0.046	p	0.056	e	0.127
	(2.42)**		(2.84)***		(2.72)***
O	-0.038	l	0.310	j	0.013
	(4.03)***		(3.43)***		(2.07)**
S	-0.019	c	0.014	land	.207
	(3.52)***		(2.48)**		(9.84)***
E	-0.004	o	0.154		
	(0.59)		(6.24)***		

followed by capital equipment. Compared to those four inputs seed, fertilizer, pesticides fuel and animal specific inputs make up a very small percentage of a farm's total cost.

The adjusted elasticity of scale is surprisingly small at 0.75. Paul *et al*. (2004) calculated cost elasticities using a translog input distance function, using ARMS data from 1996-2001, which were between 0.554 for lifestyle farms and 0.839 for very large farms. When inverted this would result in an elasticity of scale of 1.818 and 1.19 respectively. This difference could be due to the functional form, the Paul *et al*. (2004) paper used a translog model which could have lead to differences due to economies of scope. The Paul *et al.* (2004) paper focused specifically on farms in the Corn Belt, our data include farms outside of the Corn Belt that are believed to be less efficient than Corn Belt farms. These factors could explain the differences between my model and Paul *et al.'s* model.

The output*GM interaction coefficients in table 2 demonstrate how the elasticity of scale changes as the GM adoption increases. Recalling that the coefficients must be inverted and multiplied by -1 to determine the impact on the elasticity of scale (see equation **Error! Reference source not found.**), we know that if an interaction coefficient is positive then the GM technology in question is biased in favor of large farms. For HT soybeans the interaction coefficient (Shtm) is negative and significant for the within model. In the pooled model the coefficient is not statistically significant. The within estimate implies that the output level that minimizes average costs has increased, thus larger farms will realize an efficiency gain. One might also conclude that the within estimate implies that HT soybeans will put upward pressure on farm size. The *Bt* corn interaction term is statistically significant and positive in both the within model and the pooled model. The implication is that adopting *Bt* corn decreases the elasticity of scale. The GM*Cotton interaction coefficients are not statistically significant

The estimated models can also be used to determine if GM crops increase efficiency. According to the pooled model results, increasing the number of acres planted with *Bt* corn will increase efficiency, as evidenced by a statistically significant negative coefficient for the pooled model. The within model reports that *Bt* corn increases efficiency and the result is statistically significant regardless of the method used to compute variances. For the within model, when robust variance estimates are not used the HT cotton variable is negative and statistically significant, which would indicate that HT cotton increases efficiency. The remainders of the GM variables are not significant. These results strongly suggest that planting *Bt* corn increases efficiency, while the remainder of the GM crops do not.

Conclusion

In this paper we present two versions of an input distance function; a pooled model using ARMS data from 1996 through 2005, and a pseudo-panel model that has been first-differenced to remove unobserved cohort-level endogeneity. Both models have similar results. In the context of ARMS, where the data are collected by a stratified random sample, it is difficult to argue that a pseudo-panel model is superior to cross-section models or a model where repeated cross sections are pooled. However, the advantage to using a pseudo-panel lies in the fact that it allows researchers to use panel data methods that are not available in a pooled cross-section model. One can mean-difference the data and estimate a within model that is free of unobserved management bias. Furthermore, when using SFA to estimate a distance function using panel data that has been mean-differenced we can see that the OLS and the SFA estimation methods provide the exact same results. In this example, the within model results are similar to the pooled model, but any time-invariant unobserved cohort-level heterogeneity has been removed from the data. All of the input coefficients are of the correct sign, and only off-farm income is not statistically significant. Likewise all of the output coefficients have the correct sign and only fuel is insignificant. The within model shows that *Bt* corn increases efficiency, while decreasing the elasticity of scale and HT soybeans increase the elasticity of scale.

References

Bernard, J. C., J. D. Pesek, Jr., and C. Fan. "Performance Results and Characteristics of Adopters of genetically Engineered Soybeans in Delaware." *Agricultural and Resource Economics Review* 33, no. 2(2004): 282-292.

Brookes, G., and P. Barfoot. "GM Crops: The Global Economic and Environmental Impact â€“ The first Nine Years 1996-2004." *AgBioForum* 8, no. 2(2005): 187.

Bullock, D. S., and E. I. Nitsi. "Roundup Ready Soybean Technology and Farm Production Costs, Measuring the Incentive to Adopt Genetically Modified Seeds." *American Behavioral Scientist* 44, no. 8(2001): 1283.
Coelli, T., and S. Perelman. "Technical efficiency of European railways: a distance function approach." *Applied Economics* 32, no. 15(2000): 1967.
Coelli, T., P. D. S. Rao, and G. E. Battese. *An Introduction to Efficiency and Productivity Analysis*. Boston, Massachusetts: Kluwer Academic Publishers, 1998.
Endres, B. A., and J. G. Gardner. "Coexistence Failures and Damage Control: An Initial Look at Genetically Engineered Rice." *Agricultural Law* 16(2006).
Fare, R., and D. Primont. *Multi-Output Production and Duality: theory and Applications*. Boston Massachusettts: Kluwer Academic Publishers, 1995.
Fernandez-Cornejo, J., and M. Caswell. "The First Decade of Geneticlly Engeneered Crops in the United States." USDA.
Fernandez-Cornejo, J., C. Hendricks, and A. K. Mishra. "Technology Adoption and Off-Farm Household Income." *Journal of Agricultural and Applied Economics* 37, no. 3(2005).
Fernandez-Cornejo, J., C. Klotz-Ingram, and S. Jans. "Farm-Level Effects of Adopting Herbicide-Tolerant Soybeans in the U.S.Aa." *Journal of Agricultural and Applied Economics* 34(2002): 149.
Gardner, J. G., and G. C. Nelson. "Assessing the Environmental Consequences of Glyphosate Resistant Weeds in the United States." *Pest Management Science* (2007).
Hoppe, R. A., and D. E. Banker. "Structure and Finances of US Farms: 2005 Family Farm Report." United States Department of Agriculture Economic Research Service.
Hoppe, R. A., J. E. Perry, and D. Banker. "ERS Farm Typology For a Diverse Farm Sector." USDA, ERS.
Irz, X., and C. Thirtle. "Dual Technological Development in Botswana Agriculture: A Stochastic Input Distance Function Approach." *Journal of Agricultural Economics* 55, no. 3(2004): 455.
Just, R. E., and D. L. Hueth. "Multimarket exploitation: The case of biotechnology and chemicals." *American Journal of Agricultural Economics* 75, no. 4(1993): 936.
Kumbhakar, S. C., and C. A. K. Lovell. *Stochastic Frontier Analysis*. Cambridge, United Kingdom: Cambridge University Press, 2000.
Lovell, C. A. K., et al. (1994) Resources and Functionings: A New View of Inequality in Australia, ed. W. Eichorn. Berlin, Springer-Verlag Press.
Marra, M. C. (2001) The Farm Level Impacts of Transgenic Crops: A Critical Review of the Evidence, ed. P. G. Pardy. Baltimore, MD, International Food Policy Research Institute.
Marra, M. C., P. G. Pardy, and J. M. Alston. "The Payoffs to Transgenic Field Crops: An Assessment of the Evidence." *Agbioforum* 5(2002): 1-8.
Mundlak, Y. (2001) Production and Supply, ed. B. Gardner, and G. Rausser. New York, NY, Elsevier Publishing.
Mundlak, Y. "Production Function Estimation: Reviving the Primal." *Econometrica* 64, no. 2(1996): 431.
Nehring, R., J. Fernandez-Cornejo, and D. Banker. "Off-farm labour and the structure of US agriculture: the case of corn/soybean farms." *Applied Economics* 37, no. 6(2005): 633.
Paul, C. M., and R. Nehring. "Product diversification, production systems, and economic performance in U.S. agricultural production." *Journal of Econometrics* 126, no. 2(2005): 525-548.
Paul, C. M., et al. "Scale Economies and Efficiency in U.S. Agriculture: Are Traditional Farms History?" *Journal of Productivity Analysis* 22, no. 3(2004): 185.
Shephard, R. W. *Cost and Production Functions*. Princeton: Princeton University Press, 1953.
Shephard, R. W. *Theory of Cost and Production Functions*. Princeton: Princeton University Press, 1970.
Verbeek, M., and T. Nijman. "Can Cohort Data Be Treated as Genuine Panel Data?" *Empirical Economics* 17, no. 1(1992): 9.