NBER WORKING PAPER SERIES

ECONOMIES OF DENSITY VERSUS NATURAL ADVANTAGE: CROP CHOICE ON THE BACK FORTY

Thomas J. Holmes
Sanghoon Lee

Working Paper 14704
http://www.nber.org/papers/w14704

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2009

Holmes is grateful to NSF Grant 0551062 for support of this research. We have benefited from the comments of seminar participants at the NBER agglomeration conference, the University of British Columbia, the University of Minnesota, and the RSAI North American meetings. We are grateful to Julia Thornton for research assistance. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Economies of Density versus Natural Advantage: Crop Choice on the Back Forty
Thomas J. Holmes and Sanghoon Lee
NBER Working Paper No. 14704
February 2009
JEL No. Q10,R12,R14

ABSTRACT

We estimate the factors determining specialization of crop choice at the level of individual fields, distinguishing between the role of natural advantage (soil characteristics) and economies of density (scale economies achieved when farmers plant neighboring fields with the same crop). Using rich geographic data from North Dakota, including new data on crop choice collected by satellite, we estimate the analog of a social interactions econometric model for the planting decisions on neighboring fields. We find that planting decisions on a field are heavily dependent on the soil characteristics of the neighboring fields. Through this relationship, we back out the structural parameters of economies of density. Setting an Ellison-Glaeser dartboard level of specialization as a benchmark, we find that of the actual level of specialization achieved beyond this benchmark, approximately two-thirds can be attributed to natural advantage and one-third to density economies.

Thomas J. Holmes
Department of Economics
University of Minnesota
4-101 Hanson Hall
1925 Fourth Street South
Minneapolis, MN 55455
and Federal Reserve Bank of Minneapolis
and also NBER
holmes@umn.edu

Sanghoon Lee
Sauder School of Business
University of British Columbia
2053 Main Mall, Vancouver, BC V6T1Z2, Canada
Sanghoon.Lee@sauder.ubc.ca

1 Introduction

A basic principle in economics is that a location may specialize in a particular activity for two main reasons. First, the location might have some underlying characteristics that give it a comparative advantage in the activity. David Ricardo provides the example of wine production in Portugal. Second, some type of scale economies or agglomeration benefits may be associated with the activity. Adam Smith provides the example of the pin factory. When we observe location specialization, we can ask about the role that these two factors play. For example, Chicago became a major city because of its specialization as a transportation hub. How much credit is due to its comparative advantage (through its access to Lake Michigan and the Chicago River), and how much is due to agglomeration benefits? (See Cronon (1991) on this issue.) Los Angeles specializes in making movies. How much of this is due to locational advantages (the weather, quick access to mountains and beaches, the large number of beautiful people who live in the area), and how much of this is due to agglomeration benefits?

We tackle the question of why a location specializes in a setting where the geographic scale is extremely narrow and the issues are illustrated in stark terms. We look at crop choice on 160-acre square parcels of farmland called *quarter sections*. Agriculture is the textbook case for the important role that natural factors can play in the location of economic activity. Even within narrow geographic areas, soil conditions can vary, and these conditions determine a location's comparative advantage for crop choice. But scale economies also matter in this setting. When a farmer plants a particular crop on a particular field, there can be cost savings—*density economies*—from planting the same crop on nearby fields. Doing so enables the farmer to plant and harvest the crop as part of a single operation across the fields. By planting nearby fields the same, the farmer can spread out the overhead for procuring specialized equipment and specialized knowledge.

We are drawn to this agricultural setting because its features permit a particularly clean analysis of the question. Agriculture is a unique industry in terms of the extent to which it allows us to get a handle on the natural land characteristics that determine comparative advantage, such as the soil type, the slope of the land, and moisture. Detailed maps of land characteristics make it possible to determine how characteristics vary throughout a quarter section (again, a 160-acre square parcel). We combine this data with newly available maps of crop choice. Analogous to the data in Burchfield et al. (2006), our data are based on pictures from the sky (satellite imagery), and no confidentiality restrictions impede us from determining how a farmer is planting individual quarter sections. In short, with the choice of this setting, we cleanly measure both the crucial location characteristics and the activity choices at high geographic resolution.

Moreover, agriculture is a unique industry in terms of the extent to which the crucial location characteristics can be taken as exogenous, since it is mainly dependent on natural factors. The movie industry in Los Angeles benefits from its large supply of beautiful people, but this characteristic depends upon the decisions of people to move there. The high-tech industry in Silicon Valley benefits from the eminence of nearby Stanford University, but Stanford's eminence has benefited, in part, from the surrounding high-tech industry. Our analysis will rely heavily on comparing the characteristics of neighboring fields. In most related contexts, we would need to worry about a selection process for neighbors, with the underlying units of analysis choosing who their neighbors are. But a field cannot pick itself up and move around to select its neighbors.[1] Glacial activity determined the characteristics of a field's neighbors long ago.

Our econometric technique follows the literature on social interactions (see Manski (1993); Glaeser and Scheinkman (2001) and Brock and Durlauf (2001b) provide surveys). Papers in this literature study the connection in the behavior of neighboring decision units. For example, is a person more likely to commit a crime if his neighbor commits a crime (an "endogenous effect," in Manski's terminology)? An analogous question arises here: Is it more likely that soybeans will be planted in a field if soybeans are planted in an adjacent field? One issue in this literature is how to model the game theoretic interactions of the decision makers and what to do in the case of multiple equilibria (see Brock and Durlauf (2001a)). More generally, when modeling the role of increasing returns and location specialization, studies have to confront the issue that some economies will be internal and others external. None of these complications arise in our setting because quarter sections are generally operated as part of a single farm operation. (We will see evidence on this point in Section 6.) The farmer's choice of how to plant a quarter section is appropriately modeled as a social planner's problem in which all economies are internalized and in which the planting decisions across all the fields are determined jointly.

Before discussing results, we say a little more about our data. We focus on the planting decisions in the Red River Valley region of North Dakota. We picked a narrow geographic area because of computational considerations. The fertile Red River Valley is ideal for

[1] See Evans, Oates, and Schwab (1992) for an example of a paper in the social interactions literature that has to confront a situation in which neighbors are endogenous.

our purposes because many years of crop data are available for this area and because a wide variety of crops are planted in the area, making the analysis of which crop to plant interesting.

We now provide some background about quarter sections. A quarter section is the land unit that was distributed for free through the 1862 Homestead Act to individuals who promised to settle and farm the land. It is one-half mile on each side, so the area is a quarter square mile. A virtually perfect grid of squares over North Dakota (and many other states) was laid out in the early 1800s. A quarter section can be subdivided into *quarter quarters* of forty acres each, which we call *fields.* The reader may be familiar with the terms "back forty" and "front forty," which refer to these units. We aggregate our data to the level of these forty-acre fields and study the joint planting decisions of the four fields of a quarter section.

We turn now to our results. In the reduced form of the structural model, if density economies matter, the planting decision on a field depends not only on the soil characteristics of the given field, but also on the soil characteristics of neighboring fields. We find strong evidence of this link to neighbors. We estimate that for most crops, the weight placed on a field's neighbors is on the order of one-third, compared to two-thirds on the field's own characteristics. With the structural parameter estimates in hand, we can determine what would happen to plantings if we were to shut down density economies across fields for a particular crop. We estimate that planting levels of the particular crop would typically fall on the order of 40 percent.

We address the issue that our results may be driven by correlated effects (in Manski's terminology) lurking in the background. That is, there may not be any connection in decision making; the adjacent fields may simply have similar unobserved characteristics that are not being adequately controlled for. We show that our findings cannot all be attributed to correlated effects through a boundary analysis. We find that the link between a field's planting decision and its neighbor's characteristics is attenuated when the neighboring field is on the other side of one of several kinds of boundaries, including a quarter section boundary and an ownership or administration boundary. We show that in terms of *observed* soil characteristics, neighboring fields across such boundaries are no more different than neighboring fields within such boundaries. Since the pattern of *observed* soil characteristics does not change at a border, there is no reason to believe the pattern of *unobserved* soil characteristics would change either. We conclude that the attenuation of the link between neighbors is due to a reduction in the magnitude of density economies enjoyed across such boundaries.

We use our estimates to quantify the factors leading to crop specialization by quarter sections within counties. That is, why do we note that the four fields of one quarter section all tend to be planted with wheat, and in another quarter section within the same county, all four fields tend to be planted with corn? Ellison and Glaeser (1997) have shown that any analysis of geographic concentration with "small numbers" needs to take into account that some concentration can emerge from "dartboard reasons." In our analysis of specialization of quarter sections, we have a small numbers issue because there are only four fields. Consider the following extreme model of crop planting within a county. Suppose there are no density economies and that the crop suitability of particular fields is independently and identically distributed (i.i.d.) across the county, analogous to randomly throwing darts labeled "corn" and darts labeled "wheat," and so on, at a county map. By chance, this process will result in some quarter sections with all four fields that are wheat and other quarter sections with all four fields that are corn. We are interested in concentration that emerges beyond dartboard reasons. We expect that land is not i.i.d. across a county; rather, there is likely to be geographic autocorrelation, since a natural event such as a glacial river or lake extends over a wider area than a single field. Because of such a process, fields that are near each other—in particular, those in the same quarter sections—will tend to specialize in the same crops because they will have similar soils.

Further, there is specialization in the same crop by the four fields in a quarter section because of density economies. In Ellison and Glaeser (1997), concentration beyond the dartboard level through natural advantage and increasing returns is formally equivalent. But here—with our structural estimates of the density economy technology parameters and our estimates of the soil quality of each field—natural advantage and increasing returns can be distinguished. We take the dartboard level of concentration as a benchmark and decompose the contribution of natural advantage and density economies in determining the actual level of crop specialization by quarter sections within counties beyond the dartboard level. We estimate that natural advantage goes about two-thirds of the way. Given our priors of a high degree of geographic autocorrelation in soils, it is not surprising that the natural advantage contribution is big. We find it interesting that the share accounted for by density economies, about one-third, is as big as it is.

Our specialization measure is defined by long-run land use averaged over a ten-year period. The long-run average differs from planting in any one particular year because of crop rotation. Farmers often rotate their crops (including leaving a field fallow) primarily because of soil issues (nutrients, pestilence) and also to potentially spread work over the year (when different crops have different harvest and planting times). These are reasons to diversify rather than specialize. In certain prime agricultural regions in the United States, almost all of the farmers in a county are on a two-year corn-soybean crop cycle. If we look at a crop map in such an area in a single year, we can see extreme specialization at the quarter section level, on the order of half of them completely specializing in corn and the other half specializing in soybeans, with quarter section boundaries clearly traced out by the yellow and green colors on the map, indicating corn and soybeans. But when we look at the next year's map, the yellow and green colors flip, so in terms of a long-run average, there are no differences across quarter sections in land use, even though there are sharp differences at the annual frequency. In the Red River Valley region that we look at, there is substantial crop choice diversity within a county, even over the long run. So we have something interesting to go after, namely: Why are the fields of some quarter sections planted one way on average, while the fields in other quarter sections within the same county are planted a different way?

There is a long-standing interest in measuring economies of scale in farming and estimating farm production functions more generally (see, for example, the survey by Battese (1992)). For many studies, the primary interest is how average cost changes as farm operations incorporate more land. Our analysis holds fixed the land margin at the four fields of a quarter section, and examines how costs vary when those four fields are planted more intensively for a particular crop, i.e., at higher density. This is analogous to the way, with respect to the airline industry, that Caves, Christensen, and Tretheway (1984) distinguish between an airline increasing the number of routes it serves and increasing the frequency of flights. They call cost savings from the latter economies of density, and we follow their terminology.

Holmes (2008) provides a recent analysis of economies of density in Wal-Mart's store location problem. The cost saving that Wal-Mart can achieve by locating its stores close together is conceptually similar to what a farmer can achieve by planting neighboring fields the same way. One important way in which our study differs from typical farm productivity analyses such as those cited in Battese (1992) is that we do not directly observe measures directly related to productivity, such as bushels of output or labor or capital inputs. Rather, we see soil conditions and crop choices. It is from the revealed preferences underlying these choices that we infer density economy parameters. We finally cite the early study of Johnston (1972) that discussed cost savings achieved when farmers operate land parcels that are close together rather than dispersed.

This paper is most closely related to the spatial literature on the economics of industry location. The focus of much of this literature is determining the relative agglomerating force of various types of scale economies (e.g., knowledge spillovers), leaving natural advantage in the background (see Rosenthal and Strange (2004) for a survey). Ellison and Glaeser (1999) and Ellison, Glaeser, and Kerr (2007) are exceptions in that they jointly consider the forces of natural advantage and scale economies as we do. One obvious way their work differs from ours is that they look at manufacturing in all of the United States, whereas we look at crops in the Red River Valley. Our work also differs substantively in approach. We take a within-industry approach and estimate a structural economic model. Their paper takes a cross-industry nonstructural approach. By being very narrow in our application, we are able to precisely measure natural advantage in a way that would be impossible in an aggregate analysis of all manufacturing industries in the United States.

2 Theory

We model the planting decisions on the four quadrants of a square piece of farmland. We refer to the quadrants as fields and index them by $j \in \{1, 2, 3, 4\}$. The fields are arranged as illustrated in the following diagram:

1	2
3	4

Fields 2 and 3 are directly adjacent to field 1. We call directly adjacent pairs like these *A neighbors*. Field 4 is diagonal to field 1 and 3 is diagonal to 2; we call such diagonal pairs *B neighbors.* In our empirical work, a field corresponds to a 40-acre quarter quarter. The four fields together make up a 160-acre quarter section.

We develop a variant of the linear-in-means social interactions model exposited in the survey paper of Brock and Durlauf (2001b). For each field j, the farmer chooses an action y_j. We will interpret this as the level of planting of a particular crop, e.g., the quantity of soybeans planted on field j. For each field j there is a variable x_j that determines the suitability of growing the particular crop. This reflects the underlying soil characteristics of field j. We will call this the *soil quality measure.* The farmer faces the static problem of picking a vector of plantings (y_1, y_2, y_3, y_4), given a vector of soil quality (x_1, x_2, x_3, x_4), to maximize total profit over the four fields,

$$\max_{(y_1,y_2,y_3,y_4)} \pi_1 + \pi_2 + \pi_3 + \pi_4. \tag{1}$$

Assume that the profit on field 1 can be written as

$$\pi_1 = \lambda x_1 y_1 - \frac{1}{2} y_1^2 + \frac{1}{2}\theta_A y_1 y_2 + \frac{1}{2}\theta_A y_1 y_3 + \frac{1}{2}\theta_B y_1 y_4. \tag{2}$$

Observe that the profit on field 1 depends upon its own quality x_1 and own planting y_1. But it also depends upon the interactions of its own planting y_1 with the plantings on the other fields. Fields 2 and 3 are the A neighbors (i.e., directly adjacent) to field 1, and the coefficient on these interactions is θ_A. Field 4 is the B neighbor (i.e., diagonal), and the coefficient on this interaction is θ_B. The profit on the other fields is the symmetric equivalent to (2). Note the coefficients of 1/2 on the quadratic terms are a normalization on the units of profit that we impose without loss of generality.

Profit on field j does not *directly* depend upon the soil quality characteristic x_k of a neighboring field k. So this specification zeros out what are variously called exogenous or contextual effects.[2] The profit on field j *indirectly* depends upon the qualities of neighboring fields because these qualities influence the planting levels of the neighboring fields, which in turn affect profitability on field j. This is called an *endogenous* effect in the literature. If $\theta_A > 0$, then plantings on adjacent fields are complements; raising the planting level one field makes it more profitable to raise planting on the adjacent fields. The parameter θ_B has the analogous role for diagonal neighbors. If θ_A and θ_B are big, there are economies of density enjoyed from setting a high level of plantings throughout the four fields. We will refer to θ_A and θ_B as the *density economy parameters*. We define the *composite density parameter* Θ as

$$\Theta \equiv 2\theta_A + \theta_B. \tag{3}$$

The first-order necessary condition of problem (1) for field 1 (the other fields are sym-

[2]It is possible to think of stories in which exogenous effects might be present as well. For example, perhaps the characteristics of neighboring fields affect the likelihood that there will be pests, and the pests from the neighboring fields might spill over. As shown by Manski (1993), in linear structures such as this, endogenous and exogenous effects cannot be separately identified. Since the latter strike us as second order, we zero them out a priori.

metric) yields

$$
\begin{aligned}
0 &= \lambda x_1 - y_1 + \frac{1}{2}\theta_A y_2 + \frac{1}{2}\theta_A y_3 + \frac{1}{2}\theta_B y_4 + \frac{1}{2}\theta_A y_2 + \frac{1}{2}\theta_A y_3 + \frac{1}{2}\theta_B y_4 \\
&= \lambda x_1 - y_1 + \theta_A(y_2 + y_3) + \theta_B y_4.
\end{aligned} \tag{4}
$$

Note that in the choice of y_1, the farmer takes into account not only how y_1 impacts π_1, but also how y_1 impacts π_2, π_3, and π_4 (the last three terms in the first line). This differs from the standard social interaction model where each unit is a separate maximizer, playing a game with the other units (Brock and Durlauf (2001a), Bajari et al. (2006)). The game theoretic aspect of these analyses yields many complications, including the possibility of multiple equilibria. Typically, in these analyses, the decision maker for unit j has to take expectations of the decisions made by other units, whereas here the choices y_k are known when y_j is selected. Here, thinking of the back forty as playing a game with the front forty is not sensible.

Solving the first-order conditions, we obtain the policy function that specifies the planting on each field j as a function of the vector of field qualities (x_1, x_2, x_3, x_4). The policy function for field 1 (the other fields are symmetric) is

$$
y_1 = \gamma_O x_1 + \gamma_A(x_2 + x_3) + \gamma_B x_4, \tag{5}
$$

where

$$
\begin{aligned}
\gamma_O &= \frac{\lambda\left(1 - 2\theta_A^2 - \theta_B\right)}{(1 - 2\theta_A - \theta_B)(1 + 2\theta_A - \theta_B)\left(1 + \theta_B\right)} \\
\gamma_A &= \frac{\lambda\theta_A}{(1 - 2\theta_A - \theta_B)(1 + 2\theta_A - \theta_B)} \\
\gamma_B &= \frac{\lambda\left(2\theta_A^2 + \theta_B - \theta_B^2\right)}{(1 - 2\theta_A - \theta_B)(1 + 2\theta_A - \theta_B)\left(1 + \theta_B\right)}.
\end{aligned}
$$

We will refer to γ_O, γ_A, and γ_B as the *policy function parameters.* The parameter γ_O is the coefficient on the field's own quality, and γ_A and γ_B are the coefficients on neighboring qualities. If the policy function parameters are known, we can work backward and solve for the three structural parameters λ, θ_A, and θ_B as follows:

$$\begin{aligned}\lambda &= \frac{(\gamma_O - \gamma_B)\left(\gamma_O^2 - 4\gamma_A^2 + 2\gamma_O\gamma_B + \gamma_B^2\right)}{\gamma_O^2 - 2\gamma_A^2 + \gamma_O\gamma_B} \\ \theta_A &= \frac{\gamma_A(\gamma_O - \gamma_B)}{\gamma_O^2 - 2\gamma_A^2 + \gamma_O\gamma_B} \\ \theta_B &= \frac{-2\gamma_A^2 + \gamma_B(\gamma_O + \gamma_B)}{\gamma_O^2 - 2\gamma_A^2 + \gamma_O\gamma_B}.\end{aligned}$$

Average plantings across the four fields equals

$$\begin{aligned}\bar{y} &\equiv \frac{y_1 + y_2 + y_3 + y_4}{4} = \frac{1}{4}(\gamma_0 + 2\gamma_A + \gamma_B)(x_1 + x_2 + x_3 + x_4) \\ &= \left(\frac{\lambda}{1 - \Theta}\right)\frac{x_1 + x_2 + x_3 + x_4}{4} \\ &= \frac{\lambda}{1 - \Theta}\bar{x},\end{aligned} \tag{6}$$

where $\bar{y}$ is average plantings, $\bar{x}$ is average quality, and again Θ is the composite density parameter in (3). We assume that

$$1 - \Theta > 0, \tag{7}$$

as otherwise the density economies are so big that there is no solution. We normalize λ and the scaling of field quality so that

$$\frac{\lambda}{1 - \Theta} = 1. \tag{8}$$

With this normalization, average plantings across the four fields equals average field quality and the policy function coefficients sum to 1 ($\gamma_0 + 2\gamma_A + \gamma_B = 1$). If there are no density economies, $\theta_A = \theta_B = 0$, then the optimal planting y_j at each field j simply equals the quality x_j of the field.

From (6) we can see that the composite density parameter Θ has a simple interpretation. Suppose that for a particular quarter section, the density benefits are shut down. For example, let us say that walls are set up between the fields preventing any sharing of fixed cost for this particular crop.[3] This sets $\theta_A = \theta_B = \Theta = 0$ holding the other parameters fixed.[4]

[3] In our model, there is only one crop with production level y_j. But implicitly when y is low, the land is being used for something else, an outside alternative. In our experiment as we shut down density economies for the crop in question, we are leaving matters alone for the outside good.

[4] For this discussion, we set $\Theta = 0$ for one particular quarter section, holding it fixed in other quarter sections. Otherwise, this aggregate change in technology might impact prices and ultimately the λ parameter

We can see from (6) that the average planting $\bar{y}$ of the crop on the quarter section would decrease by Θ.

2.1 Specialization, Density Economies, and Natural Advantage

As we will see below, farmers tend to plant the same crops within the four fields of a quarter section. The main task of this paper is to quantify the roles that density economies and natural advantage play in this specialization.

The variance of plantings over the four fields equals

$$var_y^{within} = \frac{\sum_{j=1}^{4} (y_j - \bar{y})^2}{4}.$$

We use this variable to measure specialization (or, more precisely, the inverse of specialization). The superscript *within* is used to indicate the standard deviation is taken over the four fields within a quarter section. If var_y^{within} equals zero, the plantings in each field are exactly the same, $y_j = \bar{y}$, and we say that the four-field quarter section is *completely specialized*.

The degree of specialization depends upon both the extent of density economies and the degree to which soil qualities are similar across the four fields. If there were no density economies, the variance in plantings would equal the variance in soil conditions,

$$var_y^{within} = var_x^{within}, \text{ if } \theta_A = \theta_B = 0,$$

since $y_j = x_j$ in this case, as noted above. If density economies exist, the standard deviation of plantings is lower than for soil quality, as the farmer harmonizes plantings across the fields to realize gains from complementarities. For the next result, we simplify calculations by assuming $\theta_B = 0$. We find that specialization strictly increases in θ_A and goes to complete specialization when θ_A is near its upper limit.

Proposition 1. Set $\theta_B = 0$ and vary Θ over its range $[0, 1]$ by varying θ_A. As Θ varies, rescale λ according to (8) to leave average plantings fixed. The variance measure $var_y^{within}(\Theta)$ strictly declines in Θ and goes to zero as Θ approaches its theoretical upper bound of 1.

Proof. See the Appendix.■

(which can be interpreted as output price).

So far we have considered the problem of a single quarter section with four fields. Now suppose there are a collection of N quarter sections and index each by i. So $x_{i,j}$ is the quality of field j on quarter section i and $y_{i,j}$ is optimal plantings. Let $\overline{var}_y^{within}$ and $\overline{var}_x^{within}$ be the mean of the within quarter section variances of plantings and soil qualities over the N quarter sections. Let var_x^{across} be the overall variance of soil qualities across all the fields and all N quarter sections.

The comparative advantage contribution to specialization depends upon how soil quality is distributed across fields. To the extent that fields in the same quarter section tend to be similar in quality, this is a force that will drive plantings to be similar across the fields within the quarter section. To establish a benchmark for comparison, suppose the soil quality for each field is drawn randomly, and let $F_{i,j}(x_{i,j})$ be the cumulative distribution function for the quality draw on field j of quarter section i. Consider an extreme case where soil quality at the field level is i.i.d. across all fields and quarters, $F_{i,j}(x_{i,j}) = F(x_{i,j})$. Let N be large so that the distribution in the sample is essentially the same as the underlying distribution. We then have the following proposition.

Proposition 2. In the case where field quality is distributed i.i.d., $F_{i,j}(x_{i,j}) = F(x_{i,j})$, then

$$\overline{var}_x^{within} = \frac{3}{4} var_x^{across}. \tag{9}$$

Proof. See the straightforward calculations in the Appendix.■

The case where field qualities are i.i.d. is analogous to the Ellison and Glaeser (1997) pure dartboard special case. With this random process, by chance some quarter sections will get lucky and draw four high-quality soils for its four fields. And by chance, others will be unlucky and get four low-quality draws. The former will specialize in planting the crop, and the latter will specialize in not planting the crop. This kind of realized specialization is the dartboard-driven approach highlighted by Ellison and Glaeser (1997) and Duranton and Overman (2005). The factor shows up in the multiplicative $\frac{3}{4}$ term in (9) and is the contribution of the dartboard process in reducing the average variation within a quarter section compared to the aggregate variation.

For the next step, introducing the following simplified notation is helpful:

$$\begin{aligned} v_0 &\equiv \frac{3}{4} var_x^{across} \\ v_1 &\equiv \overline{var}_x^{within} \\ v_2 &\equiv \overline{var}_y^{within}. \end{aligned}$$

We take v_0 as the benchmark measure of specialization, the amount achieved if fields are distributed i.i.d. in a dartboard fashion and if density economies are zero. The variable v_1 is the specialization measure achieved at the actual distribution of soil quality but still with zero density economies. Finally, the variable v_2 is the specialization measure achieved at the actual distribution soil qualities and the actual density economies. (Note again that the smaller the measure, the greater the specialization.)

We use the following ratio to define the contribution of comparative advantage to specialization:

$$\textit{Natural Advantage Specialization Ratio} = \frac{v_0 - v_1}{v_0}. \tag{10}$$

In the extreme case where quality draws are i.i.d. across fields and quarters, Proposition 2 implies that the natural advantage specialization ratio equals zero. In the opposite case where the draws within a quarter section are perfectly correlated, $v_1 = 0$, so the ratio equals one. Our measure (10) is in the spirit of the Ellison and Glaeser (1997) index in that it is positive only for specialization that occurs beyond dartboard reasons. It increases the more that fields in the same quarter section tend to be similar in soils.

We construct an analogous ratio to define the contribution of density economies to specialization:

$$\textit{Density Economies Specialization Ratio} = \frac{v_1 - v_2}{v_1}. \tag{11}$$

When density economies are zero, $\theta_A = \theta_B = 0$, this ratio is zero. From Proposition 1, at the opposite extreme where density economies approach the theoretical upper bound, the ratio is one.

Finally, we can look at the spread $v_0 - v_2$ between the actual measure of specialization v_2 and the measure v_0 in the extreme dartboard case with no economies of density and decompose the spread into shares:

$$\textit{Natural Advantage Share} = \frac{v_0 - v_1}{v_0 - v_2} \tag{12}$$

$$\textit{Density Economy Share} = \frac{v_1 - v_2}{v_0 - v_2}. \tag{13}$$

These shares are, of course, closely related to the ratios we just defined. The shares are useful to report because they are a summary answer to the main question of the paper. The ratios provide information about the relative amount of specialization that is occurring beyond the dartboard case of v_0 (and not just how the change is split between v_1 and v_2).

3 Econometric Issues

Rather than observe the scalar quality index $x_{i,j}$, we observe a vector of characteristics $z_{i,j}$ with K elements for each field j on quarter section i. As we will explain below, this vector consists of variables such as dummy variables for soil type and local ground characteristics such as slope. We assume the following:

$$x_{i,j} = z'_{i,j}\beta + \varepsilon_{ij}, \tag{14}$$

where the weight vector β on characteristics is unknown. Let $\varepsilon_i \equiv (\varepsilon_{i,1}, \varepsilon_{i,2}, \varepsilon_{i,3}, \varepsilon_{i,4})$ be the vector of unobserved quality components for quarter section i and analogously $z_i \equiv (z_{i,1}, z_{i,2}, z_{i,3}, z_{i,4})$. For simplicity, assume ε_i are drawn i.i.d. across quarter sections. Make the orthogonality restriction,

$$E[\varepsilon_i z'_i] = 0. \tag{15}$$

To summarize, our assumption here is that there is a vector of soil characteristics z_i that we do observe (such as soil and weather variables) and there are other things ε_i that we miss. And this measurement error ε_i is unrelated to the soil characteristics that we do observe.

Inserting (14) into (5) yields

$$y_1 = \gamma_O z'_1\beta + \gamma_A(z'_2 + z'_3)\beta + \gamma_B z'_4\beta + (\gamma_O\varepsilon_1 + \gamma_A(\varepsilon_2 + \varepsilon_3) + \gamma_B\varepsilon_4). \tag{16}$$

Conceptually, there is no difficulty here because we can consistently estimate β and γ jointly with nonlinear least squares. In practice, this estimation approach is difficult when the dimension of β is large, thereby complicating nonlinear optimization. A convenient feature here is that once we fix $\gamma = (\gamma_O, \gamma_A, \gamma_B)$, the equation is linear in β and we can use the ordinary least squares method to calculate the minimum sum of squared errors conditional on γ. It is then easy to find the γ giving the minimum sum of squared errors. (See Appendix A.3 for how we calculate the standard errors of the estimates obtained with this method.)

We emphasize that no restriction is imposed on the correlation of the error term across fields *within* a quarter section. In particular, we allow

$$E[\varepsilon_{i,j}\varepsilon_{i,j'}] \geq 0$$

for $j \neq j'$. Thus, we allow for correlated effects. It is likely that fields in the same quarter section have an unobservable component of quality that will be correlated across the fields. Because of this, even if $\theta_A = \theta_B = 0$, if y_1 were regressed on its own measured soil quality $z_1'\beta$ and the other planting levels y_2, y_3, and $y_4 ... y_J$, we would expect to see positive coefficients on the neighboring planting levels. But when we regress y_1 on its own measured soil quality $z_1'\beta$ and the neighboring measured soil qualities $z_2'\beta, z_3'\beta, z_3'\beta$, the coefficients on the latter will be zero if density economies are zero.

We note that we also get consistent estimates if we allow for measurement error in the planting variables $y_{i,j}$ that is correlated across neighboring fields. Correlated measurement error like this might show up due to clouds blocking the satellite view of neighboring fields.

4 Data

Three main data elements are used in our analysis. The first element is the boundary information we use to define fields. The second element is data on crop choice. The third is data on soil and other land characteristics. The analysis in Section 6 uses data on land ownership and administration, but we defer description of this until later. Our data are available online.[5]

4.1 Fields

The Public Land Survey System imposed a grid of squares upon the new lands of the young United States. The Fifth Principal Meridian governing the origin of the grid for North Dakota and nearby states was established in 1815 (Committee on Integrated Land Data Mapping, 1982, p. 14). The grid consists of a hierarchy of different size squares. There are four *quarter sections* (half mile a side) in a *section* (one mile a side). There are thirty-six sections in a *township* (six miles a side). Figure 1 illustrates the section grid for Pembina County, one of the counties in North Dakota included in our study.[6] The eastern boundary of the county is irregular, following the Red River. The northern boundary meets Canada, so the top row is not full height. Otherwise, the section grid is a virtually perfect system of one-by-one-mile squares.

[5] The link to the data can be found at http://strategy.sauder.ubc.ca/lee/research.html.

[6] The boundary files for sections are posted by the North Dakota State University Extension Geospatial Education Project at http://134.129.78.3/geospatial/default.htm. We constructed the boundaries for quarters and quarter quarters ourselves by subdividing the section boundaries.

We will analyze the farmer's problem at the quarter section level, dividing it up into the four quarter quarters that we call *fields.* A field is 40 acres or 1/16 of a square mile. Our crop and soil data are at a higher resolution than the field level, but we aggregate up to the field to make the analysis more tractable and interpretable.

We study crop choice in the North Dakota Red River region. We define this region to include all counties along the Red River on the eastern border of North Dakota as well as the next layer of counties in. The twelve counties in this region are illustrated in Figure 2 and listed in Table 2.[7] There are a total of 231,000 fields in the region.[8] Equivalently, there are $14,400 \approx 231,000/16$ square miles, since there are 16 fields in each square mile section.

4.2 Crops

Our crop data are from the Crop Data Layer (CDL) program of the National Agricultural Statistics Service (NASS). The data are based on satellite images combined with survey information on the ground. Using the survey data, the NASS estimates a model of how satellite images correspond to crops. The map product contains the fitted values.

Compared to other states, North Dakota is special in that there exists a relatively long panel of annual CDL data that begins in 1997. There is significant variation across states in the availability of CDL data, since its collection depends upon the cooperation of state agencies. For example, there currently is no analogous CDL data available for Minnesota, which lies on the eastern side of the Red River Valley. The availability of many years of data with which we can determine long-run average land use is an important reason why we picked North Dakota over other states.

The resolution of the crop data is at the level of 30 meter by 30 meter squares (approximately four points per acre). We use the program ArcGIS to strip the crop information from the map product offered by NASS. We take a fixed grid of points 30 meters apart and for each year locate the point in the CDL map to determine the crop associated with this point in each year. There are 40.8 million points in the grid for our twelve-county region.

Table 1 lists the land use classifications and the fraction of grid points in each category averaged over our 1997–2006 sample period.[9] The most common category is "Spring wheat,"

[7] We excluded Barnes County, which is the second layer in from Cass County, because of concerns we had about the data quality of our soil information for this county.

[8] We discard quarter quarters (QQs) that do not consist of a regular quarter mile by quarter mile square. These nonregular QQs are negligible in land area. In footnote 10 we explain our precise criteria.

[9] The table uses the category definitions from the 2005 CDL. In 2006, the Conservation Reserve Program was shifted from the "Pasture" category to the "Fallow/idle cropland" category, resulting in a large shift

which has a .223 share. “Soybeans” is next, and then “Pasture” and “Fallow/idle cropland.” Urban activity is negligible in this area, as can be seen from the .017 share for “Urban.” The category “Clouds,” with a .028 percent share, is for observations where the satellite view of the point in a given year is blocked by clouds.

Next we map each grid point from the crop data into the field that contains it. Figure 3 illustrates some fields in Pembina County and the grid points they contain. The dark lines are the quarter section boundaries. The lighter lines are field boundaries within a quarter section. As can be seen in Figure 3, we trim off the points near the border of each field and use only interior points. We want to be careful not to misclassify a point near a border as being in an adjacent field. Each field side is .25 miles (400 meters). We trim the points that are .03 miles (48 meters) on each side of the border. Each field has approximately 100 points in the interior.[10]

Suppose crops are indexed by c and grid points indexed by g. Let $y^c_{g,t} = 1$ if crop c is planted at grid point g in year t and set it to zero otherwise. Let y^c_g be the mean value of $y^c_{g,t}$ over the years in the sample. For example, if wheat is planted at grid point g in five of the ten years, then $y^{wheat}_g = \frac{1}{2}$. Farmers in the region commonly practice crop rotation, and one such rotation is to alternate between wheat and soybeans. If this rotation is practiced at point g, then $y^{wheat}_g = \frac{1}{2}$.

To aggregate the grid point crop information to the level of a field, we define $y^c_{i,j}$ to be the mean of y^c_g over all the grid points g in the interior of field j on quarter section i. This long-run average for each field corresponds to variable $y_{i,j}$ in our econometric model.[11] Table 2 presents the summary statistics for $y^c_{i,j}$ by county and overall for the two major crops, spring wheat and soybeans. We can see in this table that there is variation in crop choice across counties. For example, the soybean share is relatively low in the northern counties (.04 in Cavalier, .09 in Pembina) and high in the southern counties (.23 in Sargent, .27 in Richland). There is also substantial variation across fields *within* each county.

between these categories. Over the years, there have also been a few reclassifications for some small crops like canola. These reclassifications do not matter for any of the major crops.

[10] We drop fields that have more than 130 points or less than 66 points. The soil information is missing for some points, and we also drop fields if more than 10 percent of their points have missing information. These cases represent a negligible land area.

[11] The selected data and programs used in the paper are posted at http://strategy.sauder.ubc.ca/lee/research.html.

4.3 Soil

Our soil data are taken from the Soil Survey Geographic (SSURGO) database maintained by the U.S. Department of Agriculture. Any given country can have hundreds of different soil types. The SSURGO data map the location of the various soils at a high level of resolution and provide underlying soil and ground characteristics for each soil type. The soil taxonomy is a standard soil classification system based on soil-forming processes, wetness, climatic environment, major parent material, soil temperature, soil moisture regimes, and so on. It has different classification levels: order–suborder–great group–subgroup–family–series, and the SSURGO data set has the first four. With the four levels, there are 1,200 types of soils in the system, and the areas we consider have 67 of them. The data come in the form of a map boundary file. We take each point in our above-mentioned 30 meter by 30 meter grid and use the SSURGO data to determine the soil and ground characteristics of that point.

As with the crop variables, we average the soil variables across points in a field to obtain $z_{i,j,k}$, the value of soil characteristic k on field j of quarter section i. Let $z_{i,j}$ be the vector of the K characteristics. In our analysis, this vector will include 67 dummy variables for different soil taxonomy codes, slope, aspect, air temperature, annual precipitation, elevation, annual unfreezing days, soil loss tolerance factor (t factor), wind erosion index (wei), latitude, longitude and their quadratic terms, and 12 dummy variables for each county. Altogether, there are $K = 110$ characteristics.

The SSURGO soil map data are considered reliable enough to have widespread use in practical applications. The maps can be found in real estate listings for farm property analogous to the way in which listings of houses for sale include pictures of each room. The detailed soil information is used in North Dakota to determine land value assessments for property taxes.

We demonstrate the utility of the soil data for our purposes by running some preliminary regressions. We regress the planting choice $y_{i,j}$ at the field level on the field's own soil characteristics, ignoring the characteristics of the field's neighbors. We run the regression for each of the twelve counties for each crop separately, so all of the variation in field characteristics is coming from within-county variation in soil variables. To interpret this regression in terms of our model, note that if there are no density economies, $\theta_A = \theta_B = 0$, then the policy function coefficients γ_A and γ_B equal zero and it is possible to consistently identify (up to a multiplicative scalar) the field attribute coefficients β in (16) through an ordinary least squares (OLS) regression. Table 3 reports, for each crop, the mean value of the R^2

of this regression averaged over the 12 counties. It also reports the minimum, median and maximum. Recall that the crops are sorted so that the most important crops come first. The R^2 tends to be fairly high for the more important crops. The mean R^2 across the twelve counties is .34 for spring wheat and .24 for soybeans. The R^2 is less for the smaller crops, but it is still non-negligible. Some of these smaller crops, such as sugar beets or potatoes, tend to be concentrated in particular counties, so the R^2 is naturally higher in the places where the crops are grown.

One final point about soil is that human behavior can impact soil properties—what soil scientists call the anthropogenic impact. In the Red River Valley, the largest anthropogenic impact was due to cooperative efforts to drain most of the land beginning in the early 1900s. This effort resulted in a legal drain system regulated by county governments. Since these efforts were regional in nature, they rarely led to differences in soil conditions at property line boundaries. A producer's management decisions can impact nutrients and stored soil moisture conditions for next year's crop, but these seasonal use-dependent variables are not measured as part of routine soil surveys. Farming practices employed over a long period of time can impact erosion and result in changes in near-surface soil properties. However, in the Red River Valley region, farmers have tended to use similar, proven practices on this valuable land. Most of the soils in the Red River Valley did not suffer the Dust Bowl erosion problems in the 1930s, as areas farther west and south did. According to Mike Ulmer, the USDA-NRCS senior regional soil scientist for the Northern Great Plains, most significant variations in soil variables for the Red River Valley region are due to natural soil genesis rather than human behavior.[12]

5 Basic Analysis

This section conducts the basic empirical analysis. The first part estimates the structural model parameters. The second part uses the model estimates to examine the contributions of density economies and comparative advantage to specialization.

[12] We are grateful to Mike Ulmer for his help with this paragraph. The USDA-NRCS is the U.S. Department of Agriculture, Natural Resources Conservation Service.

5.1 Parameter Estimates

The structure parameters of our model consist of density economy parameters θ_A and θ_B and the field quality coefficients β. (The field quality coefficients β and λ are scaled so that (8) holds, and λ drops out.) We use the nonlinear least squares procedure discussed in Section 3 to estimate the structural parameters of equation (16) for each crop.

Our baseline estimates are obtained by estimating the model on a crop-by-crop basis jointly for all twelve counties together.[13] The density parameters θ_A, θ_B are assumed to be the same in each county and the soil characteristics vector β is the same, except that we allow for county dummies in the soil vector. These estimates are reported in Table 4. The policy function estimates as well as the structural parameter estimates are reported. (The coefficients on soil quality are too numerous to report here but are available upon request.)

We begin our discussion with the policy parameter estimates. The robust pattern across all the crops is that the planting rule for a given field depends heavily on the neighboring fields. Furthermore, as one would expect, the effect is stronger with the type A neighbors that are immediately adjacent as compared to the type B diagonal neighbor. Given the scaling (8), the policy parameters sum to one ($\gamma_0 + 2\gamma_A + \gamma_B = 1$). If there were no density economies, then the own quality coefficient γ_O would equal one. It is apparent in the table that γ_O is substantially less than one for all of the crops. Consider spring wheat, for example. The weight on own field quality is .66. The weight γ_A on the two adjacent fields is .13 each, and the weight γ_B on the diagonal is .09. Altogether, fully one-third of the weight in the planting decision of spring wheat for a particular field depends upon the qualities of the other fields in the quarter section.

We turn now to the structural parameters. The robust pattern across all the crops is that the density economy parameter θ_A for adjacent fields is significantly positive. The parameter θ_B is substantially smaller in each case. The last column contains Θ, the composite density parameter (equal to $2\theta_A + \theta_B$). Recall the interpretation for Θ discussed above. If density economies are shut down for a particular crop at a particular quarter section, this is the decline in planting of the crop, expressed as a share of the initial planting. For wheat, the decline share would be .389. For most of the other crops, the decline is even bigger. The estimated density economies are sufficiently big that if the farmer were precluded from enjoying them for a particular crop, we predict there would be a substantial reduction in the planting of the crop.

[13] We use only those quarter sections with four complete fields.

Next we discuss the robustness of our estimates under alternative specifications and data restrictions. We focus on the robustness of our estimate of the composite Θ because this is a useful summary statistic. The baseline specification imposes that the structural parameters are constant across all counties including the soil coefficients. Our first robustness check is to estimate the model separately for each of the twelve counties. In Table 5 we report our average estimate of Θ. There is little difference in the result. The average Θs are higher than the baseline model estimates for seven crops and lower for the other five crops.

Next we consider what happens when we throw out quarter sections that contains land that is other than prime farmland. Our soil data contain a ranking of soil quality with categories like "prime farmland," "farmland of local importance," and "not prime farmland." For our baseline, we leave in all of the categories because the issues we are interested in very much apply here. A farmer might be more willing to plant wheat on a field that is not prime farmland if it is next to a field that is. Still, it is interesting to see what happens when we condition on all of the land being prime farmland so that all variation in soils is then *within* the prime farmland category. When we restrict attention to quarter sections where all four fields are 100 percent prime farmland, we eliminate 46 percent of the observations. Table 5 shows the results. The composite density parameter Θ actually tends to increase when the model is estimated on the prime farmland subsample, going from .389 to .578 for wheat, .422 to .598 for soybeans, and .672 to .825 for corn.

5.2 What Determines Specialization?

In the theory section, we consider three cases. Case 0 is where measured specialization occurs because of dartboard reasons but nothing else. Case 1 is where we use the actual soil distribution but do not take density economies into account. Case 2 is where we use the actual soil and actual density economies.

We do not have the actual soil qualities and the actual density economies. But with our parameter estimates, we can compute fitted values. We take the fitted values of soil quality and the fitted values of crop choice and plug these into our formulas for the specialization ratios and shares that we defined in Section 2. We evaluate means at the level of a county. For example, we calculate var_x^{across} for a particular crop by differencing out the mean in each county. Then we average over the county-level variances. So the specialization we examine compares quarter sections within the same county. Table 6 reports the results.

Recall that the Natural Advantage Specialization Ratio (NASR) would be zero if soil quality were distributed i.i.d. within each county. The ratio would equal one if instead the four fields of each quarter section were perfectly correlated in soil quality. Table 6 shows this ratio is roughly .6 throughout the various crops, which is in between i.i.d. and perfect correlation but closer to perfect correlation. We conclude that there is a significant amount of geographic autocorrelation in soil qualities leading the natural advantage factor to play a big role in specialization.

Next observe that the Density Economies Specialization Ratio (DESR) is quite high for most crops. If DESR were at its upper limit of one, the NASR (column one) would equal the Natural Advantage Share (column three). If DESR is high but not quite equal to one, then column one is bigger than column three, but not much bigger. This is the basic pattern in Table 6. The bottom line is that natural advantage creates significant specialization beyond what would happen with a dartboard, and then density economies contributes even more specialization beyond that, with a roughly two-thirds/one-third contribution split, as seen in the last two columns.

6 Further Analysis

The key empirical finding of the previous section is that planting in a land parcel depends on neighboring soil characteristics in addition to those of the parcel itself. From this exhibited behavior, we recover structural parameters in which density economies are significant. A natural concern in interpreting any result like this is that there are some unobservable characteristics of the given land parcel that are somehow being captured by the measured soil characteristics of the neighbors. The most plausible candidate here is some kind of measurement error in soil classification. Perhaps a soil scientist made a mistake in classifying one field but got things correct on an adjacent field. For example, suppose type C soil is good for corn and type W is good for wheat. Actual soil types of nearby fields tend to be correlated, and it may be that all of the fields of a quarter section are type W. If one field is mistakenly classified as C—and we see wheat planted on this field—we might mistakenly attribute this to density economies flowing from the adjacent wheat fields.

Here we explore the issue by examining what happens across the borders of quarter sections, taking an approach in the spirit of Holmes (1998). The idea is that if all of our results are entirely due to soil measurement error issues and the like, then we should get similar results when we look at fields that cross quarter section boundaries and ownership boundaries. But if our results are arising from density economies, then we would expect the results to be attenuated at such boundaries, because such boundaries are relatively more likely to be boundaries between farm operations. We expect the potential for density economies to be less when adjacent fields are managed by different operations.

The first part of this section introduces additional data on farm operations. The second part presents simple descriptive evidence to make our point. In the last part, we reestimate our model with some of the additional data brought in.

6.1 Ownership Data

Here we introduce additional data related to ownership. The data make the point that quarter section boundaries are closely connected to ownership and administration boundaries.

Recall we earlier defined the A neighbors of a field to be the two directly adjacent neighboring fields in the same quarter section. As illustrated in Figure 4a, a field has two additional adjacent neighbors in different quarter sections. Call these the C neighbors of a field.

For one of the counties of our study, Cass County, we have obtained a file containing all the land parcels in the county and the name of the owner of each parcel.[14] We take each point in our above-mentioned 30 meter by 30 meter grid and map it to our parcel information. We then aggregate up to the field level.[15]

In Table 7, we examine differences in ownership at field boundaries. For this analysis, we exclude fields categorized as urban from the soil file information (approximately 2 percent of the observations) and fields where one of the owner names is blank at the field boundaries (slightly more than 1 percent of observations). We classify boundaries of adjacent fields as to whether the fields are A neighbors or C neighbors. We see from Table 7 that for A neighbors, in a fraction .87 of the time, the two fields are part of the same legal land parcel. In contrast, if the fields are C neighbors—again, meaning that they are separated by a quarter section boundary—in only .01 of the time are the fields part of the same legal parcel. Table 7 makes clear that in this county, legally defined land ownership parcels are essentially quarter sections.

Even when adjacent fields are contained in different legal parcels, it still may be the case

[14] This GIS shapefile is posted by the Cass County government at http://www.casscountynd.gov/departments/gis/Download.htm/. We used the file that was current as of 2007/07/30.

[15] In the rare cases where there are different owner names within the same field, we assign ownership to the modal name.

that the two parcels are held by the same owner. The table also shows the fraction of cases where owner name is identical for the adjacent fields.[16] At A borders, the match rate goes from .87 for parcels to .92 for owner name. At C borders, the match rate goes from .01 to .29. So we see that ownership commonly crosses quarter section boundaries. Nevertheless, in well more than half of the cases, adjacent fields that cross quarter section boundaries are held by owners with different names.

Even when ownership names differ across quarter section boundaries, the fields may be operated as part of the same operation. Fields held within the same family can be listed under different names (e.g., a wife's or grandmother's maiden name). And farmers can operate land that they lease. The Department of Agriculture maintains a database of farm operation boundaries but does not publicly release this information. However, before 2008, it released a data product that we can use to draw inferences about farm operation boundaries. (The 2008 Farm Act bans release of the data from this point forward, so we are lucky to get the data when we did.) The data are geospatial information on "Common Land Units" (CLUs). These are the reporting units for government subsidy programs. CLUs are typically quarter sections, through there is much variation. The public release of the CLU data did not disclose the individual operators. Nevertheless, it was published in such a way that we were able to manipulate it to determine which county office administers the federal farm programs for each field.[17] Typically, a field is administered by the office in the same county as the field. But there are cases where farm operations cross county boundaries, and in such cases it is typically convenient for the farmer to work with a single administrative office. In such a case, a field can be administered by an office in a different county from where the field is located.

The bottom part of Table 7 shows our results for the county administrator variable in Cass County. Adjacent fields being administered by different counties is relatively rare. Out of about 54,000 adjacent field pairs, this happens only 1,002 times. When this does happen, it is 10 times as likely to occur when the pair crosses quarter section boundaries (type C) than not (type A)—1,002 instances versus 110. For these type C borders, the fraction of cases with the same ownership name falls from .30 if the administrator is the same to only .06 if the administrator is different. We take this as solid evidence that a difference in county

[16] Identical owner name is defined as a match on the first five characters. Last name is listed first, so this permits matches on different first names. It does not make much difference if we require a match on all the characters.

[17] The county-level CLU data happen to contain (1) all the CLUs in the county plus (2) CLUs outside the county that are administered by the county.

administrator across field boundaries is a good signal of a difference in farm operations across field boundaries. This is useful because we have the county administrator variable for all our counties but have the legal parcel information just for Cass County. Below, we use both variables.

6.2 Evidence of Planting Discontinuities at Quarter Section Boundaries

This part makes the point that soil quality does not change discontinuously at quarter section boundaries, but planting does. We begin with a graphical illustration. Figure 5a provides a map of soil boundaries for a particular area in our sample and an overlay of the quarter section boundaries. It illustrates there is heterogeneity in soils within a field. Given the arbitrary way in which quarter section boundaries were drawn back in the early 1800s, we expect to see no connection with soil boundaries and no connection is evident here.

Figure 5b provides a crop map over the same area. The connection between crop borders and quarter section borders is readily evident. So we see that crops change at quarter section boundaries but soil does not.

We now make the same point in a table. In column 1 of Table 8, we report $\overline{std}_x^{within}$, the mean within quarter section deviation of soil quality for each crop, normalized by the mean plantings for the crop.[18] The statistic reported is like a coefficient of variation. Note that the variation of soil quality within quarter sections is significant, the statistic ranging from about .12 to .24 throughout the various crops. The existence of this within quarter section soil heterogeneity is a key part of our identification strategy.

To explain the second column, we introduce the concept of a *fake quarter section.* As illustrated in Figure 4b, we imagine the quarter section boundaries were drawn one-quarter mile to the west and one-quarter mile to the north, compared to the way they were actually drawn. As before, there are four fields in a fake quarter section. But now we see that each field is actually in a different true quarter section. Now field boundaries are actually quarter section boundaries.

In column 2, we do the same calculation as for column 1, except we calculate the standard

[18] For each quarter section i we compute the standard deviation across the four fields and then take the mean over all quarter sections over all 12 counties. We divide by mean plantings for each crop, which approximately equals the means in Table 1. (The slight difference arises because a few incomplete quarter sections are thrown out here.)

deviation of soils within each fake quarter section. The two columns are virtually the same. Just as in Figure 5a, soil changes are unrelated to quarter section boundaries.

The last two columns report the standard deviation within each quarter section in actual average plantings, again normalized by the mean levels. The variation is much greater across boundaries in the fake quarters than within boundaries for the actual quarters. This is consistent with the sharp delineation in crop boundaries illustrated in Figure 5b.

6.3 Extended Model Estimates

We reestimate our earlier model in three different ways and show how taking into account boundary considerations impacts the results. The results are in Table 9. For the sake of comparison with our earlier work, we repeat in the first column of Table 9 our baseline estimate of the Θ from Table 4.[19]

The first exercise reestimates the model exactly as we did in Table 4, except we use the fake quarters rather than the actual quarters. Recall that the distribution of soils for the fake quarters is the same as for the actual quarters. We get very different results with the fake quarters. The estimate of Θ is attenuated for all of the crops. For example, for wheat, the coefficient falls from .39 to .32, soybeans .42 to .37, corn .67 to .56. Now we are not surprised that we are still getting estimates of significant density economies even in the fake quarters, because we expect density economies are larger and extend beyond quarter sections, an issue we raise in the conclusion. Our main point is that a measurement error story for why we are getting positive estimates for Θ cannot account for why these estimates would be attenuated at quarter section boundaries.

The second exercise estimates the model with actual quarters, as in our original approach. But now we use information about county administration. We estimate a specification of the reduced form policy function so that plantings in neighboring given fields are weighted by γ_A and γ_B as before if they are administered by the same county. But if a different county (and then likely a different operation), we assume the weights are $\delta\gamma_A$ and $\delta\gamma_B$. So the parameter δ is like a discount factor. In the estimates, there is clear pattern of substantial discounting. We focus our discussion on the major crops. For spring wheat, soybeans, and corn, the discount factors are .59, .70, .51. These are substantially below one.

The third exercise is analogous to the second exercise. But now we discount when the

[19] We get a slightly different number of observations because we use only the fields that have all three neighbors in the same *fake* squares.

name is different rather than when the county administrator is different. We use data from Cass County because that is all that is available. For all but the negligible crops at the bottom of the list where there is little data, there is a clear pattern of discounting. For example, for spring wheat, soybeans, and corn the estimated discount factors are .78, .67, and .81. Again, these are well less than one, but not zero. We do not expect these to come out to zero, because farm operation boundaries clearly can cross ownership name boundaries through rental markets and through different names in the same family. Again, the point here is that alternative explanations for our positive estimates of γ_A and γ_B based on some kind of measurement error that is averaged out across field boundaries cannot account for why the estimates are significantly attenuated at name change boundaries. Our density economy explanation can account for this pattern.

7 Conclusion

For the quarter sections in North Dakota's Red River Valley, we estimate the determinants of crop specialization. We quantify the relative contributions of Ellison-Glaeser dartboard effects, natural advantage (land characteristics), and scale economies (density economies here). These kinds of decompositions are difficult to provide in most settings. We are able to get somewhere in this setting because (1) the natural advantage factor in agriculture is overwhelming, (2) we are able to get extremely detailed data at a narrow geographic level on land characteristics and choice, and (3) in the early 1800s, the United States government drew an arbitrary square grid of quarter sections in the landscape, and we make heavy use of this grid.

We believe the major limitation of this paper is that it does not take into account density economies that extend beyond the quarter section level. The average farm size in North Dakota from recent Census figures is eight quarter sections.[20] Of course, planting decisions at the individual farm operation level will extend more broadly over a farm's operations and not just a single quarter section. Moreover, we expect that scale economies extend beyond individual farm operations because the fixed costs of infrastructure such as grain elevators, sugar beet processing plants, and research in location-specific seeds are spread over many farms, and as neighboring farmers share knowledge. Once we start expanding the geographic scope to be big enough to cross individual farm operations, we need to bring

[20] In the 2002 Census, the average farm size in North Dakota is reported to be 1,238 acres or 8.01=1,238/160 quarter sections. If we were to weight farms by acreage, the mean would be substantially higher.

in various game-theoretic coordination issues and externality issues. As a first step, we picked a land unit—a quarter section—small enough that we could be confident it was all under the same management, but large enough so that it is possible to conduct an interesting geographic analysis. The next step in this research line is to broaden the geographic reach and confront externality issues.

We expect that our approach of combining micro soil data with the satellite crop data will have other applications. Indeed, having such detailed information about what is happening at such a narrow geographic level is rare in any industry. In particular, these data could be used to look at the impacts of government policies, perhaps including policies related to ethanol. The corn-soybean rotation mentioned in the introduction has actually been discontinued recently by some farmers wishing to take advantage of ethanol-induced high prices for corn and planting corn every year instead. The detailed data permit us to determine exactly where the switches are taking place.

A Appendix

A.1 Proof of Proposition 1

Setting $\theta_A = \theta$ and $\theta_B = 0$ in the equations for γ_O, γ_A, and γ_B in the text and imposing the normalization $\lambda/(1-2\theta) = 1$ yields

$$
\begin{aligned}
\gamma_O &= \frac{\left(1-2\theta^2\right)}{(1+2\theta)} \\
\gamma_A &= \frac{\theta}{(1+2\theta)} \\
\gamma_B &= \frac{2\theta^2}{(1+2\theta)}.
\end{aligned}
$$

So we can write $y_1 - \bar{y}$ as (noting $\bar{y} = \bar{x}$)

$$
\begin{aligned}
y_1 - \bar{y} &= \frac{1-2\theta^2}{1+2\theta}x_1 + \frac{\theta}{1+2\theta}x_2 + \frac{\theta}{1+2\theta}x_3 + \frac{2\theta^2}{1+2\theta}x_4 - \frac{1+2\theta}{1+2\theta}\bar{x} \\
&= \frac{1}{1+2\theta}\left(\begin{array}{c} x_1 - 2\theta^2 x_1 + \theta x_2 + \theta x_3 + 2\theta^2 x_4 - \frac{x_1}{4} - \frac{x_2}{4} - \frac{x_3}{4} - \frac{x_4}{4} \\ -\frac{\theta}{2}x_1 - \frac{\theta}{2}x_2 - \frac{\theta}{2}x_3 - \frac{\theta}{2}x_4 \end{array}\right) \\
&= \frac{1}{1+2\theta}\left(\frac{3}{4}x_1 - 2\theta^2 x_1 + \frac{\theta}{2}x_2 + \frac{\theta}{2}x_3 + 2\theta^2 x_4 - \frac{x_2}{4} - \frac{x_3}{4} - \frac{x_4}{4} - \frac{\theta}{2}x_1 - \frac{\theta}{2}x_4\right) \\
&= \frac{1}{1+2\theta}\left(x_1 - \bar{x} - 2\theta^2\left(x_1 - x_4\right) + \frac{\theta}{2}\Delta\right) \\
&= \frac{1}{1+2\theta}B_1
\end{aligned}
$$

for B_1 and Δ defined by

$$
\begin{aligned}
B_1 &\equiv x_1 - \bar{x} - 2\theta^2\left(x_1 - x_4\right) + \frac{\theta}{2}\Delta \\
\Delta &\equiv x_2 + x_3 - x_1 - x_4.
\end{aligned}
$$

From symmetry we have

$$\begin{aligned} B_2 &\equiv x_2 - \bar{x} - 2\theta^2 (x_2 - x_3) - \frac{\theta}{2}\Delta, \\ B_3 &\equiv x_3 - \bar{x} - 2\theta^2 (x_3 - x_2) - \frac{\theta}{2}\Delta, \\ B_4 &\equiv x_4 - \bar{x} - 2\theta^2 (x_4 - x_1) + \frac{\theta}{2}\Delta. \end{aligned}$$

We need to show that

$$var_y^{within} = \sum_{j=1}^{4} \frac{1}{4} \left[\frac{1}{1+2\theta}\right]^2 B_j^2$$

decreases in θ. Clearly, the effect of θ on the middle term is decreasing. So it is sufficient to show that

$$\sum_{j=1}^{4} 2B_j \frac{dB_j}{d\theta} < 0.$$

Now

$$\begin{aligned} B_1 \frac{dB_1}{d\theta} &= \left[x_1 - \bar{x} - 2\theta^2 (x_1 - x_4) + \frac{\theta}{2}\Delta\right] \left[-4\theta (x_1 - x_4) + \frac{1}{2}\Delta\right] \\ B_2 \frac{dB_2}{d\theta} &= \left[x_2 - \bar{x} - 2\theta^2 (x_2 - x_3) - \frac{\theta}{2}\Delta\right] \left[-4\theta (x_2 - x_3) - \frac{1}{2}\Delta\right] \\ B_3 \frac{dB_3}{d\theta} &= \left[x_3 - \bar{x} - 2\theta^2 (x_3 - x_2) - \frac{\theta}{2}\Delta\right] \left[-4\theta (x_3 - x_2) - \frac{1}{2}\Delta\right] \\ B_4 \frac{dB_4}{d\theta} &= \left[x_4 - \bar{x} - 2\theta^2 (x_4 - x_1) + \frac{\theta}{2}\Delta\right] \left[-4\theta (x_4 - x_1) + \frac{1}{2}\Delta\right]. \end{aligned}$$

Observe that the $\bar{x}$ term will cancel out when we add these up, so set it to zero. If we add the first and fourth line, we get

$$\begin{aligned} & [x_1 + x_4 + \theta\Delta] \frac{1}{2}\Delta + \left[x_1 - x_4 - 4\theta^2 (x_1 - x_4)\right] [-4\theta (x_1 - x_4)] \\ = \; & [x_1 + x_4 + \theta\Delta] \frac{1}{2}\Delta - 4\theta \left[1 - 4\theta^2\right] (x_1 - x_4)^2. \end{aligned}$$

By symmetry, when we add the second and third lines, we get

$$-[x_2 + x_3 - \theta\Delta] \frac{1}{2}\Delta - 4\theta \left[1 - 4\theta^2\right] (x_2 - x_3)^2.$$

Adding these together yields

$$\begin{aligned}
& [x_1 + x_4 - (x_2 + x_3) + 2\theta\Delta] \frac{1}{2}\Delta - 4\theta \left[1 - 4\theta^2\right] (x_1 - x_4)^2 \\
= & -[1 - 2\theta] \frac{1}{2}\Delta^2 - 4\theta \left[1 - 4\theta^2\right] (x_1 - x_4)^2 - 4\theta \left[1 - 4\theta^2\right] (x_2 - x_3)^2 .
\end{aligned}$$

This is negative as long as $1 - 2\theta > 0$ and $1 - 4\theta^2 > 0$, which is true for $\theta < \frac{1}{2}$, as claimed. *Q.E.D.*

A.2 Proof of Proposition 2

In the case where field quality is i.i.d., the population mean for large N of the within-quarter section variance is

$$\begin{aligned}
\overline{var}_x^{within} &= E[x_1 - \bar{x}]^2 \\
&= E\left[\left(\frac{3}{4}x_1 - \frac{x_2}{4} - \frac{x_3}{4} - \frac{x_4}{4}\right)^2\right] \\
&= \frac{12}{16}E[x_1^2] - \frac{12}{16}E[x_1]^2 \\
&= \frac{3}{4}var_x,
\end{aligned}$$

where we use independence of the draws of x_1, x_2, x_3, x_4, for the equalities in lines 3 and 4.

A.3 How to Calculate Standard Errors of the Nonlinear Least Square Estimates Obtained with the Two-Stage Method

This section shows how to calculate the standard errors (covariance matrix) of the estimates $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ obtained with the two stage method. (See Davidson and MacKinnon (2004) for a more detailed explanation.) We express the estimation equation (16) as the following:

$$y_1 = \mathbf{x}(\boldsymbol{\gamma}, \boldsymbol{\beta}) + u,$$

where $\mathbf{x}(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \gamma_O z_1'\beta + \gamma_A(z_2' + z_3')\beta + \gamma_B z_4'\beta$ and $u = \gamma_O\varepsilon_1 + \gamma_A(\varepsilon_2 + \varepsilon_3) + \gamma_B\varepsilon_4$.

Let $\left(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\beta}}\right)$ be the nonlinear least square estimates and let $\mathbf{X}(\boldsymbol{\gamma}, \boldsymbol{\beta}) = [\partial\mathbf{x}(\boldsymbol{\gamma}, \boldsymbol{\beta}) / \partial\boldsymbol{\gamma}$

$\partial \mathbf{x}(\boldsymbol{\gamma}, \boldsymbol{\beta}) / \partial \boldsymbol{\beta}]$. The consistent estimator of the covariance matrix of the nonlinear estimator $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ is

$$\widehat{Var}\left(\hat{\gamma}, \hat{\beta}\right) = s^2 \left(\hat{\mathbf{X}}'\hat{\mathbf{X}}\right)^{-1}, \tag{17}$$

where $\hat{\mathbf{X}} = \mathbf{X}\left(\hat{\gamma}, \hat{\beta}\right)$ and

$$s^2 \equiv \frac{1}{n-k} \sum \hat{u}^2 = \frac{1}{n-k} \sum \left(y_1 - \mathbf{x}\left(\hat{\gamma}, \hat{\beta}\right)\right)^2. \tag{18}$$

One easy way to calculate the covariance matrix (17) is to regress $\mathbf{y}_1 - \mathbf{x}\left(\hat{\gamma}, \hat{\beta}\right)$ on $\mathbf{X}\left(\hat{\gamma}, \hat{\beta}\right)$ and use the resulting covariance matrix.

$$\mathbf{y} - \mathbf{x}\left(\hat{\gamma}, \hat{\beta}\right) = \hat{\mathbf{X}}\mathbf{b} + residuals$$

This is called the Gauss-Newton regression (GNR) and the covariance matrix for the OLS parameter estimates $\hat{\mathbf{b}}$ is

$$Var\left(\hat{\mathbf{b}}\right) = (s')^2 \left(\hat{\mathbf{X}}'\hat{\mathbf{X}}\right)^{-1}, \tag{19}$$

where

$$(s')^2 = \frac{1}{n-k} \sum residuals^2 = \frac{1}{n-k} \sum^2 \left(\mathbf{y} - \mathbf{x}\left(\hat{\gamma}, \hat{\beta}\right) - \hat{\mathbf{X}}\mathbf{b}\right)^2.$$

Since the regressor does not have any explanatory power ($\mathbf{b} = \mathbf{0}$), $(s')^2$ is equal to s^2 in equation (18). Therefore, it holds

$$\widehat{Var}\left(\hat{\gamma}, \hat{\beta}\right) = Var\left(\hat{\mathbf{b}}\right).$$

We can easily obtain $Var\left(\hat{\mathbf{b}}\right)$ by regressing $\mathbf{y}_1 - \mathbf{x}\left(\hat{\gamma}, \hat{\beta}\right)$ on $\mathbf{X}\left(\hat{\gamma}, \hat{\beta}\right)$ in any statistical package.

References

Bajari, Patrick, Han Hong, John Krainer, and Denis Nekipelov (2006), "Estimating Static Models of Strategic Interactions," Working paper, University of Michigan.

Battese, George E. (1992),"Frontier Production Functions and Technical Efficiency: A Survey of Empirical Applications in Agricultural Economics," *Agricultural Economics*, 7(3–4): 185–208.

Brock, William A., and Steven N. Durlauf (2001a), "Discrete Choice with Social Interactions," *Review of Economic Studies*, 68(2): 235–260.

Brock, William A., and Steven N. Durlauf (2001b), "Interactions-Based Models," in *Handbook of Econometrics*, Vol 5, ed. J. J. Heckman and E. E. Leamer, Amsterdam: North Holland.

Burchfield, Marcy, Henry G. Overman, Diego Puga, and Matthew A. Turner (2006), "Causes of Sprawl: A Portrait from Space," *Quarterly Journal of Economics*, 121(2): 587–633.

Cass County GIS.ed, Cass county property parcels [computer file]. Available online at http://www.casscountynd.gov/departments/gis/Download.htm, accessed August 2008.

Caves, Douglas W., Laurits R. Christensen, and Michael W. Tretheway (1984), "Economies of Density versus Economies of Scale: Why Trunk and Local Service Airline Costs Differ," *RAND Journal of Economics*, 15(4): 471–489.

Cronon, W. (1991), *Nature's Metropolis: Chicago and the GreatWest,* New York: Norton.

Davidson, Russell, and James G. MacKinnon (2004), *Econometric Theory and Methods*, New York: Oxford University Press.

Duranton, Gilles, and Henry Overman (2005), "Testing for Localisation Using Micro-Geographic Data," *Review of Economic Studies*, 72(4): 1077–1106.

Committee on Integrated Land Data Mapping (1982), *Modernization of the Public Land Survey System*, Washington, DC: National Academy Press.

Ellison, Glenn, and Edward L. Glaeser (1997), "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy*, 105(5): 889–927.

Ellison, Glenn, and Edward L. Glaeser (1999), "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?" Papers and Proceedings, *American Economic Review*, 89(2): 311–316.

Ellison, Glenn, Edward L. Glaeser, and William Kerr (2007), "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," Working Paper 13068, National Bureau of Economic Research, April.

Evans, William N., Wallace E. Oates, and Robert M. Schwab (1992), "Measuring Peer Group Effects: A Study of Teenage Behavior," *Journal of Political Economy*, 100(5): 966–991.

Glaeser, Edward L., and Jose A. Scheinkman (2001), "Measuring Social Interactions," in *Social Dynamics*, ed. S. N. Durlauf and H. P. Young, pp. 83–102, Cambridge: MIT Press.

Holmes, Thomas J. (1998), "The Effect of State Policies on the Location of Manufacturing: Evidence from State Borders," *Journal of Political Economy*, 106(4): 667–705.

Holmes, Thomas J. (2008), "The Diffusion of Wal-Mart and Economies of Density," Working Paper 13783, National Bureau of Economic Research.

Johnston, Warren E. (1972), "Economies of Size and the Spatial Distribution of Land in Farming Units, *American Journal of Agricultural Economics*, 54(4): 654–656.

Manski, Charles F. (1993), "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60(3): 531–542.

Rosenthal, Stuart S., and William C. Strange (2004), "Evidence on the Nature and Sources of Agglomeration Economies," in *Handbook of Urban and Regional Economics*, Vol. 4, *Cities and Geography,* ed. J. V. Henderson and J.-F. Thisse, pp. 2119–2172, Amsterdam: Elsevier.

Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Soil Survey Geographic (SSURGO) Database for North Dakota [computer file]. Available online at http://soildatamart.nrcs.usda.gov, accessed 2007.

USDA, National Agricultural Statistics Service, North Dakota Cropland Data Layer 1996–2007 [computer file]. Available online at http://datagateway.nrcs.usda.gov/.

Figure 1. Section Boundaries, Pembina County, ND

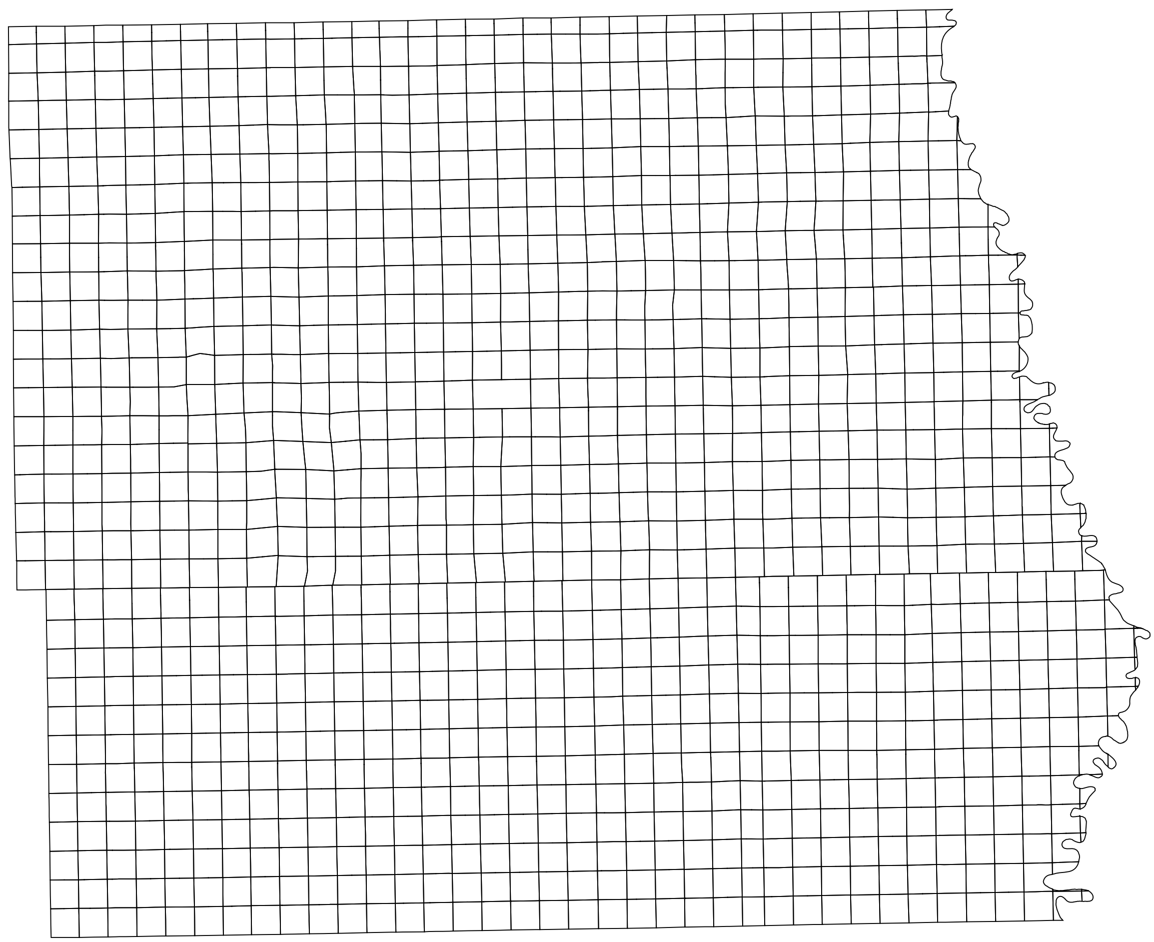


Figure 2. Counties Used in Our Analysis
(North Dakota Red River Region)

Figure 3. Example Fields and Grid Points
from Pembina County

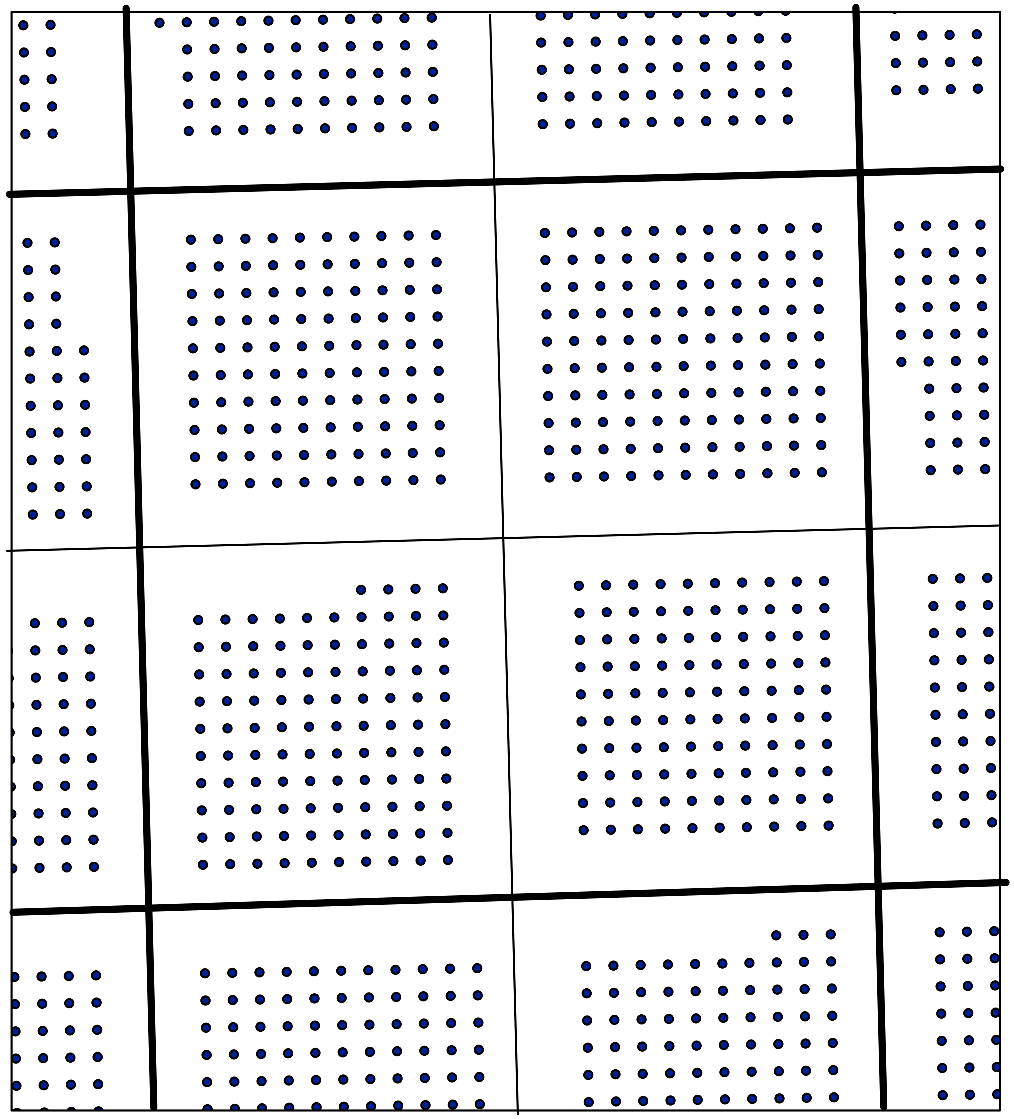
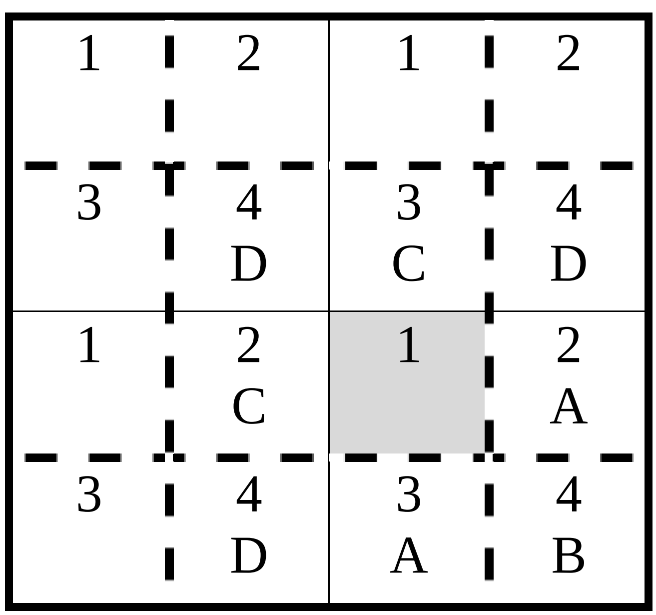


Figure 4a. Extended Neighborhoods

1	2	1	2
3	4 D	3 C	4 D
1	2 C	1	2 A
3	4 D	3 A	4 B

Figure 4b. Fake Quarter Neighborhoods



Note: Dotted lines indicate fake quarter section boundaries. Note that in a fake quarter section, each field belongs to a different true quarter section.

Figure 5a. Example Soil Map in a Quarter Section

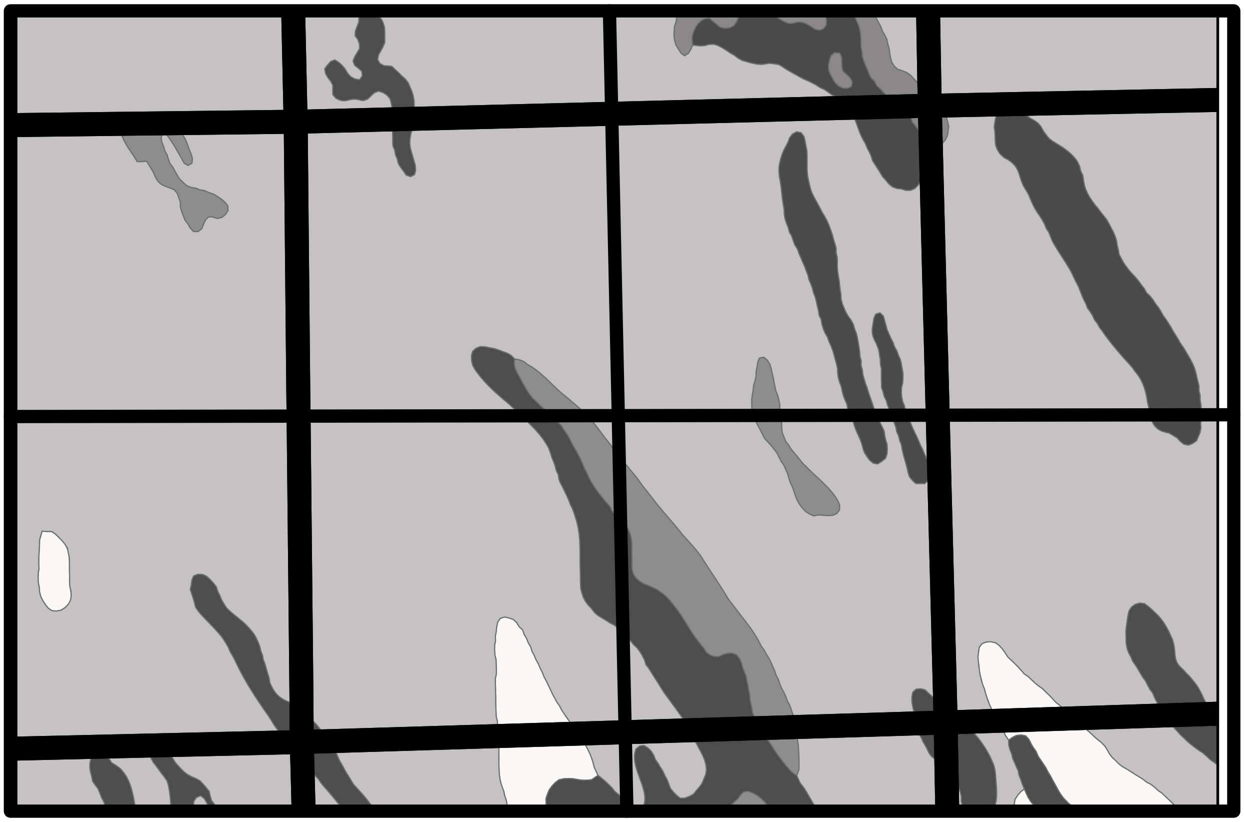


Figure 5b. Example Crop Map in a Quarter Section

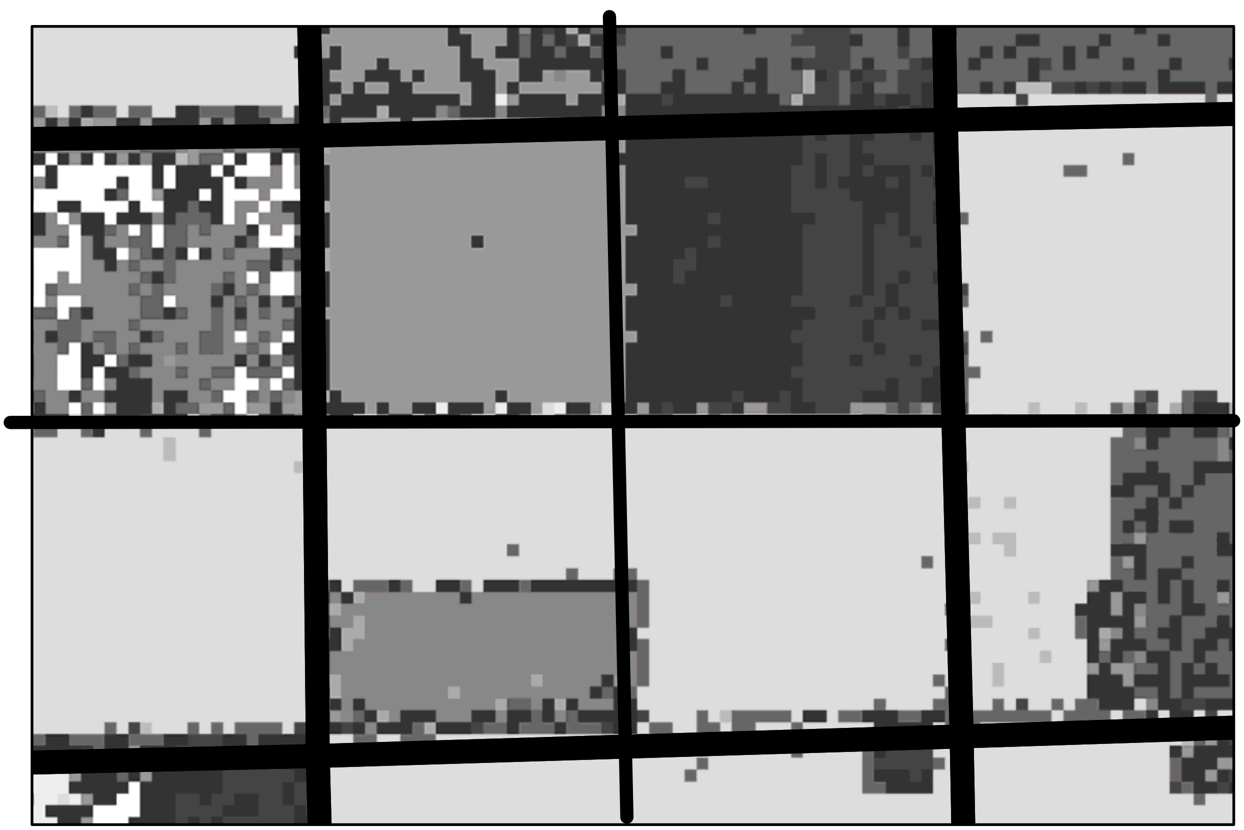


Table 1
Land Use in North Dakota Red River Region
Averages over 1997–2006 Period

Crop	Share
Spring wheat	.223
Soybeans	.158
Pasture/range/Conservation Reserve Program/farmstead	.154
Fallow/idle cropland	.098
Corn	.049
Dry edible beans	.037
Water	.037
Sunflowers	.036
Barley	.034
Clouds	.028
Durum wheat	.026
Other small grains & hay (oats, millet, rye & winter wheat, alfalfa & other hay)	.020
Canola	.019
Urban	.017
Beets	.015
Other crops (canola, flaxseed, safflower & very small acreage crops)	.012
Woods, woodland pasture	.010
Potatoes	.009
Miscellaneous (15 residual categories)	.018

Source: Authors' calculations with Cropland Data Layer North Dakota, 1997–2006

Table 2
Summary Statistics of Crop Planting Share at the Field Level
Two Major Crops in North Dakota Red River Region

		Spring Wheat				Soybeans			
County	Number of Fields	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max
All Counties	231,595	.23	.16	.00	.86	.17	.15	.00	.98
By County									
Cass	28,750	.25	.14	.00	.70	.34	.15	.00	.98
Cavalier	24,576	.30	.15	.00	.79	.04	.04	.00	.33
Grand Forks	23,426	.25	.15	.00	.72	.16	.11	.00	.65
Nelson	16,128	.20	.16	.00	.70	.08	.09	.00	.57
Pembina	18,599	.31	.16	.00	.86	.09	.08	.00	.54
Ramsey	21,102	.18	.13	.00	.65	.06	.07	.00	.41
Ransom	13,824	.14	.15	.00	.75	.17	.15	.00	.68
Richland	24,013	.15	.14	.00	.67	.27	.15	.00	.84
Sargent	14,324	.14	.13	.00	.62	.23	.16	.00	.72
Steele	11,520	.27	.13	.00	.72	.23	.13	.00	.78
Traill	14,327	.25	.13	.00	.69	.27	.13	.00	.75
Walsh	21,006	.27	.16	.00	.74	.09	.08	.00	.54

Source: Authors’ calculations with Cropland Data Layer North Dakota, 1997–2006

Table 3
Goodness of Fit of Planting Regression on Soil Characteristics
(Summary Statistics of R^2s of Individual Regressions for Each of Twelve Counties)

Crop	Mean	Min	Median	Max
Spring wheat	0.335	0.210	0.322	0.562
Soybeans	0.237	0.084	0.246	0.384
Corn	0.182	0.057	0.196	0.393
Dry beans	0.156	0.057	0.154	0.337
Sunflowers	0.108	0.047	0.108	0.222
Barley	0.115	0.032	0.110	0.239
Durum wheat	0.102	0.027	0.054	0.257
Other small grains	0.179	0.055	0.168	0.333
Canola	0.078	0.021	0.044	0.250
Beets	0.152	0.028	0.129	0.360
Other selected crops	0.067	0.020	0.048	0.165
Potatoes	0.119	0.041	0.088	0.318

Note: For each regression, the planting choice of a field is regressed on its own field characteristics, ignoring the characteristics of the neighboring fields.

Table 4
Baseline Model Estimates
(N = 208,220)

	Policy Parameters			Structural Parameters		
Crop	γ_O	γ_A	γ_B	θ_A	θ_B	$\Theta = 2\theta_A + \theta_B$
Spring wheat	0.658	0.128	0.087	0.160	0.070	0.389
	(0.006)	(0.004)	(0.006)	(0.008)	(0.012)	(0.008)
Soybeans	0.632	0.137	0.094	0.174	0.074	0.422
	(0.007)	(0.005)	(0.007)	(0.011)	(0.017)	(0.010)
Corn	0.459	0.199	0.144	0.316	0.040	0.672
	(0.007)	(0.005)	(0.007)	(0.027)	(0.044)	(0.014)
Dry beans	0.530	0.181	0.109	0.278	0.016	0.573
	(0.010)	(0.006)	(0.009)	(0.026)	(0.039)	(0.017)
Sunflowers	0.613	0.147	0.093	0.197	0.056	0.451
	(0.012)	(0.008)	(0.011)	(0.020)	(0.030)	(0.017)
Barley	0.757	0.095	0.053	0.113	0.041	0.267
	(0.011)	(0.007)	(0.010)	(0.011)	(0.017)	(0.014)
Durum wheat	0.721	0.115	0.048	0.147	0.019	0.313
	(0.011)	(0.007)	(0.010)	(0.014)	(0.020)	(0.015)
Other small grains	0.519	0.183	0.114	0.284	0.020	0.588
	(0.008)	(0.005)	(0.008)	(0.023)	(0.035)	(0.015)
Canola	0.602	0.143	0.112	0.180	0.100	0.460
	(0.013)	(0.009)	(0.013)	(0.021)	(0.033)	(0.019)
Beets	0.366	0.225	0.184	0.408	0.002	0.818
	(0.012)	(0.008)	(0.012)	(0.088)	(0.152)	(0.028)
Other selected crops	0.706	0.112	0.070	0.136	0.056	0.329
	(0.015)	(0.009)	(0.014)	(0.018)	(0.027)	(0.020)
Potatoes	0.325	0.234	0.207	.441	0	.882
	(0.010)	(0.002)	(0.005)	(0.008)	(0)	(.0015)

Table 5
Robustness Check

	Baseline Model	By County	Only on Prime Farmland
Crop	Θ	Average Θ	Θ
Spring wheat	0.389 (0.008)	0.447 (0.117)*	0.578 (0.022)
Soybeans	0.422 (0.010)	0.433 (0.147)*	0.598 (0.026)
Corn	0.672 (0.014)	0.527 (0.248)*	0.825 (0.028)
Dry beans	0.573 (0.017)	0.567 (0.237)*	0.503 (0.032)
Sunflowers	0.451 (0.017)	0.551 (0.121)*	0.702 (0.045)
Barley	0.267 (0.014)	0.536 (0.244)*	0.624 (0.062)
Durum wheat	0.313 (0.015)	0.543 (0.250)*	0.818 (0.057)
Other small grains	0.588 (0.015)	0.574 (0.229)*	0.470 (0.032)
Canola	0.460 (0.019)	0.664 (0.243)*	0.870 (0.035)
Beets	0.818 (0.028)	0.759 (0.178)*	0.739 (0.026)
Other selected crops	0.329 (0.020)	0.588 (0.266)*	0.804 (0.066)
Potatoes	.882 (.0015)	0.730 (0.148)*	0.867 (0.024)
N	208,220		111,596

* Standard error calculated from county level Θ estimates.

Table 6. Natural Advantage vs. Density Economies Decomposition

Crop	Natural Advantage Specialization Ratio	Density Economies Specialization Ratio	Natural Advantage Share	Density Economy Share
Spring wheat	0.607	0.694	0.690	0.310
Soybeans	0.597	0.729	0.670	0.330
Corn	0.654	0.913	0.674	0.326
Dry beans	0.631	0.846	0.669	0.331
Sunflowers	0.473	0.754	0.544	0.456
Barley	0.492	0.531	0.646	0.354
Durum wheat	0.651	0.585	0.761	0.239
Other small grains	0.597	0.857	0.634	0.366
Canola	0.758	0.773	0.802	0.198
Beets	0.858	0.972	0.862	0.138
Other selected crops	0.723	0.620	0.808	0.192
Potatoes	0.686	0.988	0.689	0.311

Table 7
Ownership Statistics for Cass County
Mean Match Rates for Adjacent Fields

Border Type	Same County Administrator?	Adjacent Fields Part of Same Legal Parcel	Adjacent Fields Have Owners with Same Name	Number of Adjacent Field Pairs
A		0.87	0.92	27,271
C		0.01	0.29	26,946
A	Yes	0.87	0.93	27,161
A	No	0.34	0.41	110
C	Yes	0.01	0.30	25,944
C	No	0.00	0.06	1,002

Source: Cass county property parcel map, county-level common land unit maps

Table 8
Variation in Soil and Crop Plantings
Across Actual Quarters and Fake Quarters

	x variation		*y* variation	
Crop	Actual	Fake	Actual	Fake
Spring wheat	0.136	0.137	0.230	0.379
Soybeans	0.120	0.121	0.257	0.417
Corn	0.249	0.253	0.391	0.713
Dry beans	0.185	0.188	0.477	0.724
Sunflowers	0.125	0.127	0.481	0.765
Barley	0.132	0.133	0.471	0.698
Durum wheat	0.132	0.134	0.506	0.677
Other small grains	0.244	0.248	0.621	0.837
Canola	0.166	0.166	0.469	0.763
Beets	0.249	0.251	0.450	0.840
Other selected crops	0.168	0.169	0.537	0.857
Potatoes	0.284	0.287	0.616	0.866

Table 9. Further Analysis Estimates

Crop	Baseline Model (from Table 4)	Same as Baseline but Fake Quarters	Policy Function Estimates with Discounting if Different County Administrator			Policy Function Estimates with Discounting if Different Ownership Name (Cass County Only)		
	Θ	Θ	γ_A	γ_B	δ	γ_A	γ_B	δ
Spring wheat	0.389 (0.008)	0.318 (0.008)	0.129 (0.004)	0.085 (0.006)	0.592 (0.041)	0.192 (0.016)	0.089 (0.024)	0.836 (0.028)
Soybeans	0.422 (0.010)	0.368 (0.011)	0.137 (0.005)	0.094 (0.007)	0.703 (0.048)	0.157 (0.014)	0.053 (0.021)	0.773 (0.034)
Corn	0.672 (0.014)	0.558 (0.014)	0.199 (0.005)	0.144 (0.007)	0.509 (0.066)	0.197 (0.016)	0.147 (0.023)	0.606 (0.065)
Dry beans	0.573 (0.017)	0.472 (0.016)	0.182 (0.006)	0.108 (0.009)	0.711 (0.052)	0.246 (0.026)	0.160 (0.038)	0.933 (0.056)
Sunflowers	0.451 (0.017)	0.311 (0.018)	0.148 (0.008)	0.093 (0.011)	0.761 (0.074)	0.234 (0.023)	0.132 (0.035)	0.782 (0.067)
Barley	0.267 (0.014)	0.188 (0.014)	0.097 (0.007)	0.050 (0.010)	0.354 (0.101)	0.241 (0.034)	0.266 (0.052)	0.824 (0.045)
Durum wheat	0.313 (0.015)	0.244 (0.015)	0.118 (0.007)	0.045 (0.010)	0.000 (0.091)	0.184 (0.034)	0.255 (0.050)	0.919 (0.056)
Other small grains	0.588 (0.015)	0.449 (0.013)	0.183 (0.005)	0.114 (0.008)	1.326 (0.064)	0.145 (0.013)	0.013 (0.019)	1.642 (0.184)
Canola	0.460 (0.019)	0.375 (0.019)	0.155 (0.008)	0.152 (0.012)	0.000 (0.061)	0.241 (0.039)	0.152 (0.058)	0.992 (0.081)
Beets	0.818 (0.028)	0.759 (0.026)	0.228 (0.008)	0.181 (0.012)	0.781 (0.043)	0.229 (0.024)	0.182 (0.035)	0.874 (0.060)
Other selected crops	0.329 (0.020)	0.283 (0.020)	0.122 (0.009)	0.088 (0.013)	0.000 (0.101)	0.261 (0.038)	0.171 (0.058)	1.014 (0.157)
Potatoes	0.882 (.0015)	0.862 (0.015)	0.235 (0.008)	0.205 (0.011)	1.091 (0.045)	0.225 (0.045)	0.096 (0.065)	1.164 (0.146)
N	208,220	201,596		208,220			25,124	