

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, UBC Sauder School of Business, UBC Economics, UBC Agricultural Economics, 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, the Federal Reserve System, or the NBER.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2009 by Thomas J. Holmes and Sanghoon Lee. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

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, Revised June 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 The 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 particular location may specialize in a particular activity for two broadly-defined reasons. First, the location might have some underlying characteristic that gives it a natural advantage in the activity. Second, some type of scale economy may be attained by concentrating (or *agglomerating*) production at the location. When we observe specialization, we can ask about the roles 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 natural 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 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*. We observe that the various fields within a quarter section tend to be planted the same way and ask: How much of this specialization is due to natural advantage and how much is due to scale economies?

Obviously, natural conditions like soil quality and topography play key roles in determining what is planted. Indeed, agriculture is the textbook case for the role that natural factors can play in the location of economic activity. Adjacent fields will tend to be similar in attributes like soil quality and topography and for this reason it is no surprise to see adjacent fields planted similarly.

Perhaps more subtly, scale economies also can also lead nearby fields to be planted similarly. When a farmer is out in a field and has just run a particular piece of equipment, the farmer can economize on setup costs by continuing on to the next field, treating it the same way. Potential efficiencies extend beyond day-to-day field operations and include economies involving specialized equipment, like a sugar beet harvester. When a farmer acquires expensive equipment like this, the farmer needs to use the equipment on many acres to justify its expense and ensure it gets sufficient utilization. So if the farmer includes sugar beets in the crop rotation for one field, the farmer will have an incentive to include sugar beets in the rotation of the neighboring fields. (We have more to say about crop rotation in Section 2.) The issue of indivisibility applies to the farmer as well. Farmers can have specialized knowledge that is crop specific. If a field is planted with a certain crop that benefits from a particular kind of farmer knowledge, it will be advantageous for neighboring fields to be planted the same way to fully utilize the farmer's particular knowledge.

In standard concepts of scale economies there is no notion of geography; cost savings are achieved by increasing scale at a particular point. With the scale economies considered here, there is a notion of geography. Expanding a particular activity at one point (i.e., the planting of a particular crop on a particular field) makes it advantageous to expand the activity at neighboring points (the neighboring fields). We use the term *density economies* to distinguish this type of scale economies from the standard kind. We follow the literature in using this terminology, as we explain below.

We are drawn to study the factors underlying agglomeration in agriculture because the features of this industry allow for a particularly clean analysis. Agriculture is a unique industry in terms of the extent to which it is possible to get a handle on the natural land characteristics that determine natural advantage, such as the soil type, the slope of the land, and moisture. 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. 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.<sup>1</sup> Glacial activity determined the characteristics of a field's neighbors long ago.

Before discussing results, we say a little more about our data. We focus on the long-run average 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 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. 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.

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

---

<sup>1</sup>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.

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 between 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 long-run planting levels of the particular crop would typically fall on the order of 40 percent.

Using our estimates we can also quantify the factors leading to crop specialization by quarter sections within counties. That is, why we might see all four fields of one quarter section planted with wheat, and in another quarter section within the same county, all four fields 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” 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 emerges beyond that occurring by chance.

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 degree to which specialization of quarter sections within counties extends 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.

We also address the issue that our results may be driven by correlated effects (in Manski (1993)'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.

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), Rosenthal and Strange (2001), 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.

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. Arzaghi and Henderson (2008) study density economies (networking benefits) of advertising agencies by analyzing the location pattern of new advertising agencies in Manhattan. 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.

## 2 Theory

We develop a variant of the linear-in-means social interactions model expounded in the survey paper of Brock and Durlauf (2001b). 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? Before getting into the details of the model, it is useful to raise four points about the model.

First, while we motivate the existence of density economies by appealing to the existence of indivisibilities in the use of specialized capital and gains from plowing the next row over

the same as the previous one, we don't explicitly model these various details about farming. Instead, we capture these forces by writing down a reduced form profit structure where the profitability of the planting choice on a particular field depends on the choices made on neighboring fields. In the exercises we consider, estimates of the parameters of this profit structure are sufficient for what we do.

Second, farmers typically pick *crop rotations* rather than individual crops. For example, a farmer growing sugar beets would typically only plant beets every four years, rotating in other crops the other years. One possibility would be sugar beets (year one), wheat (year two), barley (year three), and wheat (year four). The basic issues that we are interested in apply equally well when the choice variable is a rotation rather than a crop. For example, a farmer choosing a rotation with sugar beets will need specialized equipment and knowledge for sugar beets that a farmer choosing a two-year wheat/barley rotation will not need. Explicitly modeling the underlying agricultural details that lead to crop rotation is beyond the scope of this paper. Instead, we introduce in a reduced-form way an incentive to cycle in, or out, a particular crop over time.

Third, the various economies we have in mind include those that emerge from daily operations (e.g. from continuously operating a plow on adjacent fields) and those that are longer term in nature (e.g. that involve utilization of long-lived equipment). In our model, we explicitly differentiate these short run and long run considerations. However, it is useful to alert the reader up front that in our estimation, we will be unable to separately identify these two factors. Our identification strategy relies on relating planting decisions to characteristics of adjacent fields like soil quality. If measurable field characteristics varied over time, we could look at the dynamics of planting decisions to separate out short-run and long-run considerations. But measurable field characteristics are constant over time, so we are only able to identify the combined effect of short-run and long-run economies. For the exercises we consider, this is sufficient.

Fourth, we simplify by modeling the planting decision of each crop in isolation. Implicitly, when deciding how much to plant a particular crop, there is an opportunity cost in terms of other crops that is captured in a reduced form way in the profit function for the particular crop. In our view, the limitations of this approach are trumped by the benefits, at least for a first cut. In particular, the approach delivers linear policy rules relating long-run average planting levels to neighboring field characteristics. This makes the analysis quite tractable and enables us to back out a structural interpretation from raw moments of the data.

## 2.1 Details of the Model

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.

There are  $T$  periods. For each field  $j$  and period  $t$ , the farmer chooses  $y_{j,t}$ . We interpret this as the planting level of a particular crop, e.g., the quantity of soybeans planted on field  $j$  in period  $t$ . For each field  $j$  there is a variable  $x_j$  that determines the long-run suitability of growing the particular crop. This reflects the underlying soil characteristics of field  $j$ . We will call this the *permanent soil quality measure*. In each period  $t$ , each field  $j$  has a *transient quality measure*  $\omega_{j,t}$ . As this term can vary over time, it introduces a force in the model to induce the farmer to cycle the planting of the crop, e.g. practice crop rotation. We do not impose any restriction on how the  $\omega_{j,t}$  change over time or on the correlation of the  $\omega_{j,t}$  across fields at a particular time. We do assume, without loss of generality, that the  $\omega_{j,t}$  sum up to 0 over time for each field,

$$\sum_t \omega_{j,t} = 0.$$

Overall quality at a point in time is the sum  $x_j + \omega_{j,t}$  of the permanent and transient components. Since the  $\omega_{j,t}$  sum to zero for each  $j$ , the long-run average equality of field  $j$  equals  $x_j$ .

Given the vector of the  $x_j$  and the matrix of the  $\omega_{j,t}$ , the farmer faces the problem of picking a matrix of the  $y_{j,t}$  of plantings over fields and time to maximize total profit,

$$\max_{y_{j,t}} \sum_{j,t} \pi_{j,t}, \tag{1}$$

where for simplicity we are ignoring discounting.

The specification of the  $\pi_{j,t}$  is a key step. Assume that the profit on field 1 in period  $t$  can be written as

$$\begin{aligned} \pi_{1,t} = & \lambda(x_1 + \omega_{1,t})y_{1,t} - \frac{1}{2}y_{1,t}^2 + \frac{1}{2}\phi^L y_{1,t}\bar{y}_1 \\ & + \frac{1}{2}\theta_A^S y_{1,t}(y_{2,t} + y_{3,t}) + \frac{1}{2}\theta_A^L y_{1,t}(\bar{y}_2 + \bar{y}_3) + \frac{1}{2}\theta_B^S y_{1,t}y_{4,t} + \frac{1}{2}\theta_B^L y_{1,t}\bar{y}_4 \end{aligned} \quad (2)$$

for  $\bar{y}_j = (\sum_t y_{j,t})/T$ . The profit on the other fields is the symmetric equivalent to (2).<sup>2</sup>

Observe first that the profit on field 1 depends upon its own field characteristic and its own plantings: permanent soil quality  $x_1$  and transient quality  $\omega_{1,t}$ , current planting  $y_{1,t}$ , and long run average planting  $\bar{y}_1$ . The interaction of  $y_{1,t}$  with  $\bar{y}_1$  captures any long-run scale economy that might arise from planting the same field the same way over time. If the coefficient  $\phi^L > 0$ , then plantings on the same field in different periods are complements; raising the planting level in one period makes it more profitable to raise planting in all periods.

The profit also depends upon the interactions of its own planting  $y_{1,t}$  with the plantings on the other fields. Fields 2 and 3 are the A neighbors (i.e., directly adjacent) to field 1, and the coefficients on these interactions are  $\theta_A^S$  and  $\theta_A^L$ .  $\theta_A^S$  captures short-run density economies between adjacent fields. If  $\theta_A^S > 0$ , then plantings on adjacent fields in the same period are complements.  $\theta_A^L$  captures long-run density economies between adjacent fields. If  $\theta_A^L > 0$ , then plantings on adjacent fields in different periods are complements. Field 4 is the B neighbor (i.e., diagonal), and the coefficients on this interaction are  $\theta_B^S$  and  $\theta_B^L$ . The parameter  $\theta_B^S$  and  $\theta_B^L$  have the analogous roles for diagonal neighbors. 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. We will refer to  $\theta_A^S, \theta_B^S$  as the *short run density economy parameters* and  $\theta_A^L, \theta_B^L$  as the *long run density economy parameters*.

Profit on field  $j$  does not *directly* depend upon the soil quality characteristic  $x_k$  of a neighboring field  $k$ . So this specification zeroes out what are variously called exogenous or contextual effects.<sup>3</sup> The profit on field  $j$  *indirectly* depends upon the soil qualities of

---

<sup>2</sup>Note that in the choice of  $y_{1,t}$ , the farmer takes into account not only how  $y_{1,t}$  impacts its own field profits  $\{\pi_{1,1}, \pi_{1,2}, \dots, \pi_{1,T}\}$ , but also how  $y_{1,t}$  impacts the profits from the other fields 2, 3, and 4. 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)). Here, it is not sensible to think of the back forty as playing a game with the front forty.

<sup>3</sup>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.

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 (Manski (1993)).

The first-order necessary condition of problem (1) with respect to  $y_{1,t}$  (the other fields are symmetric) yields

$$0 = \lambda(x_1 + \omega_{1,t}) - y_{1,t} + \phi^L \bar{y}_1 + \theta_A^S (y_{2,t} + y_{3,t}) + \theta_A^L (\bar{y}_2 + \bar{y}_3) + \theta_B^S y_{4,t} + \theta_B^L \bar{y}_4.$$

Summing up the first order conditions over time we obtain

$$0 = \lambda' x_1 - \bar{y}_1 + \theta_A (\bar{y}_2 + \bar{y}_3) + \theta_B \bar{y}_4, \quad (3)$$

where

$$\begin{aligned} \theta_i &\equiv \frac{\theta_i^L + \theta_i^S}{1 - \phi^L} \quad (i = A, B) \\ \lambda' &\equiv \frac{\lambda}{1 - \phi^L}. \end{aligned}$$

The parameter  $\theta_i$  combines the long-run and short-run density economies for neighbor type  $i$ . It is this parameter that we will go after in the empirical section as we are unable to separately identify its components.

Solving equation (3) for field 1 and corresponding equations for the other fields, we obtain the policy function that specifies the average planting on each field  $j$  as a function of the vector of permanent field qualities  $(x_1, x_2, x_3, x_4)$ . The policy function for field 1 (the other fields are symmetric) is

$$\bar{y}_1 = \gamma_O x_1 + \gamma_A x_2 + \gamma_A x_3 + \gamma_B x_4, \quad (4)$$

where

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

We will refer to  $\gamma_O$ ,  $\gamma_A$ , and  $\gamma_B$  as the *policy function parameters*. The parameter  $\gamma_O$  is the coefficient on the field's own quality, and  $\gamma_A$  and  $\gamma_B$  are the coefficients on neighboring qualities. If the policy function parameters are known, we can work backward and solve for the three structural parameters  $\lambda'$ ,  $\theta_A$ , and  $\theta_B$  as follows:

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

We summarize the density economies by defining the *composite density parameter*  $\Theta$  as

$$\Theta \equiv 2\theta_A + \theta_B. \tag{5}$$

We provide an interpretation of  $\Theta$  by showing how the parameter summarizes the impact of a hypothetical policy experiment shutting down density economies. Imagine a wall is erected in a particular quarter section that separates all four fields in the quarter section, eliminating all potential for density economies. Formally, after the wall, short-run and long-run density economies are zero for both  $A$  and  $B$  neighbors,<sup>4,5</sup>

$$\theta_A^S = \theta_A^L = \theta_B^S = \theta_B^L = 0.$$

---

<sup>4</sup>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.

<sup>5</sup>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  $\lambda$  parameter (which can be interpreted as output price).

Average plantings across the four fields equals

$$\bar{y} \equiv \frac{\bar{y}_1 + \bar{y}_2 + \bar{y}_3 + \bar{y}_4}{4} = \frac{1}{4} (\gamma_O + 2\gamma_A + \gamma_B) (x_1 + x_2 + x_3 + x_4) \quad (6)$$

$$= \frac{\lambda'}{1 - 2\theta_A - \theta_B} \frac{x_1 + x_2 + x_3 + x_4}{4} \quad (7)$$

$$= \left( \frac{\lambda'}{1 - \Theta} \right) \frac{x_1 + x_2 + x_3 + x_4}{4}$$

$$= \frac{\lambda'}{1 - \Theta} \bar{x},$$

where  $\bar{y}$  is average plantings,  $\bar{x}$  is average quality, and again  $\Theta$  is the composite density parameter in (5). We assume that

$$1 - \Theta > 0, \quad (8)$$

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

$$\frac{\lambda'}{1 - \Theta} = 1. \quad (9)$$

With this normalization, the policy function coefficients sum to 1 ( $\gamma_0 + 2\gamma_A + \gamma_B = 1$ ).

Suppose that the initial situation is that there is no wall. Given the normalization (9), we see from (6) that with no wall, average field planting across the four fields equals average field quality,

$$y^{no-wall} = \bar{x}.$$

When the wall is erected it has the effect of reducing the composite density economies to zero,  $\Theta^{wall} = 0$ , so average planting equals

$$y^{wall} = \lambda' \bar{x} = (1 - \Theta) y^{no-wall}.$$

Thus the composite density parameter has a structural interpretation as the fraction that average plantings decrease on account of a wall. Note to identify the impact of this policy experiment, we need not separately identify the short-run and long-run components of density economies. This follows because both kinds are eliminated by a wall and only the sum matters.

Density economies not only impact *average* planting across the four fields, they also impact the *dispersion* of plantings across the four fields. We use the within quarter section

variance as our dispersion measure,

$$d = \frac{\sum_{j=1}^4 (\bar{y}_j - \bar{y})^2}{4}. \quad (10)$$

It is intuitive that when density economies are substantial, it induces a farmer to plant the four fields of a quarter section the same way. Our formal result is,

*Proposition 1.* Set  $\theta_B = 0$  and vary  $\Theta$  over its range  $[0, 1]$  by varying  $\theta_A$ . As  $\Theta$  varies, rescale  $\lambda$  according to (9) to leave average plantings fixed. The variance measure  $d(\Theta)$  strictly declines in  $\Theta$  and goes to zero as  $\Theta$  approaches its theoretical upper bound of 1.

*Proof.* See the Appendix. ■

Thus as  $\Theta$  goes to its theoretical upper bound, plantings across the four fields are equalized.

## 2.2 The “Within” Specialization Measure

As we will see below, farmers tend to plant the four fields of a quarter section the same way. The main task of this paper is to quantify the roles that density economies and natural advantage play in this specialization. We first define the measure for this specialization and show how we decompose it into a density economy share and natural advantage share. We estimate the specialization measure and the shares in the empirical section.

Our specialization measure is based on the dispersion defined in (10) of plantings across the four fields within a quarter section. Suppose we have a set of  $N$  quarters in an area (let’s say a county) and let each quarter section be indexed by  $i = 1, \dots, N$ . Let the county mean *within dispersion* measure be

$$d^{county} = \frac{\sum_{i=1}^N d_i}{N}$$

where  $d_i$  is the planting dispersion of quarter section  $i$ . The term “within” is included here to emphasize that dispersion is first calculated *within* each quarter section and then averaged.

The mean within dispersion measure  $d^{county}$  is closely related to specialization. When  $d^{county}$  it is very small, fields within a quarter section tend to be planted the same way. As a benchmark with which to compare  $d^{county}$ , consider a hypothetical exercise where there are no density economies and no natural advantages. In other words, (1) the planting of an individual field is arbitrarily set to its field quality  $y_{ij}^{dartboard} = x_{ij}$  and (2) all  $4N$  fields in the county are randomly reshuffled into groups of four that we call *dartboard* quarter sections (as opposed to actual quarter sections). This idea of taking into account random dartboard factors follows Ellison and Glaeser (1997). In the appendix, we show that if we were to

calculate the county dispersion measure for this hypothetical exercise, for large  $N$  it would approximately equal

$$d^{dartboard} = \frac{3 \sum_{i=4}^N \sum_{j=1}^4 (x_{ij} - \bar{x})^2}{4N}.$$

With this benchmark in hand, we define the *Within Specialization Measure* by

$$W = \frac{d^{dartboard} - d^{county}}{d^{dartboard}}.$$

The Within Specialization Measure captures the specialization beyond what would happen with a dartboard. If there are no density economies and if soils are randomly distributed across fields (with no tendency for adjacent fields to be correlated in soil types) then  $d^{dartboard} = d^{county}$  and the measure  $W = 0$ . In this extreme case, there exists zero tendency for fields in the same quarter section to be planted the same way (relative to the way other fields in the county are planted). At the other extreme case, if the fields within each quarter section are planted exactly the same way, then the within specialization measure  $W = 1$ .

We focus on the “within” measure because it is jointly determined by the two forces highlighted in the title of the paper. First, as shown in Proposition 1 above, when density economies are big, the four fields will tend to be planted the same way, meaning the within measure will be big. Second, the natural advantage force will also contribute to making the within measure big. This follows because we expect the soil qualities within a quarter to be more highly correlated than across quarters. To decompose the relative importance of these two factors, define an intermediate case where the natural advantage factor is taken into account but not density economies. In particular, define

$$d^{natural\_advantage} = \frac{\sum_{i=1}^N \sum_{j=1}^4 (x_{ij} - \bar{x}_i)^2}{4N}$$

where  $\bar{x}_i$  is the mean field quality of quarter  $i$  and let

$$W^{natural\_advantage} = \frac{d^{dartboard} - d^{natural\_Advantage}}{d^{dartboard}}.$$

If there are no density economies ( $\Theta = 0$ ), then  $W^{natural\_Advantage}$  will equal our overall within measure  $W$  (since  $y_{ij} = x_{ij}$  in this case). If  $\Theta > 0$ , then density economies also

contribute. The share of the credit that can be attributed to natural advantage is

$$\text{Natural Advantage Share} = \frac{W^{\text{natural\_advantage}}}{W}, \quad (11)$$

while the balance of credit goes to density economies,

$$\text{Density Economy Share} = \frac{W - W^{\text{natural\_advantage}}}{W}. \quad (12)$$

### 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:

$$x_{i,j} = z'_{i,j}\beta + \varepsilon_{ij}, \quad (13)$$

where the weight vector  $\beta$  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  $\varepsilon_i$  are drawn i.i.d. across quarter sections. Make the orthogonality restriction,

$$E[\varepsilon_i z'_i] = 0. \quad (14)$$

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  $\varepsilon_i$  that we miss. And this measurement error  $\varepsilon_i$  is unrelated to the soil characteristics that we do observe.

Inserting (13) into (4) yields

$$\bar{y}_1 = \gamma_O z'_1 \beta + \gamma_A (z'_2 + z'_3) \beta + \gamma_B z'_4 \beta + (\gamma_O \varepsilon_1 + \gamma_A (\varepsilon_2 + \varepsilon_3) + \gamma_B \varepsilon_4). \quad (15)$$

Conceptually, there is no difficulty here because we can consistently estimate  $\beta$  and  $\gamma$  jointly with nonlinear least squares. In practice, this estimation approach is difficult when the dimension of  $\beta$  is large, thereby complicating nonlinear optimization. A convenient feature here is that once we fix  $\gamma \equiv (\gamma_O, \gamma_A, \gamma_B)$ , the equation is linear in  $\beta$  and we can

use the ordinary least squares method to calculate the minimum sum of squared errors conditional on  $\gamma$ . It is then easy to find the  $\gamma$  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  $\bar{y}_1$  were regressed on its own measured soil quality  $z'_1\beta$  and the other planting levels  $\bar{y}_2, \bar{y}_3$ , and  $\bar{y}_4 \dots \bar{y}_J$ , we would expect to see positive coefficients on the neighboring planting levels. But when we regress  $\bar{y}_1$  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  $\bar{y}_{i,j}$  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.<sup>6</sup>

### 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

---

<sup>6</sup>The link to the data can be found at <http://strategy.sauder.ubc.ca/lee/research.html>.

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.<sup>7</sup> 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.

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 and we could, in principle, have allowed for smaller decision making units, e.g. quarter quarter quarter sections of 10 acres. Our motivation for aggregating up to the 40 acre field level is that it makes the analysis 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.<sup>8</sup> (They are also listed as part of Table 2 below.) There are a total of 231,000 fields in the region.<sup>9</sup> 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

---

<sup>7</sup>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.

<sup>8</sup>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.

<sup>9</sup>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.

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.<sup>10</sup> The most common category is “Spring wheat,” 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.<sup>11</sup>

Suppose crops are indexed by  $c$  and grid points indexed by  $g$ . Let  $y_{g,t}^c = 1$  if crop  $c$  is planted at grid point  $g$  in year  $t$  and set it to zero otherwise. Let  $\bar{y}_g^c$  be the mean value of  $y_{g,t}^c$  over the years in the sample. For example, if wheat is planted at grid point  $g$  in five of the ten years, then  $\bar{y}_g^{wheat} = \frac{1}{2}$ . 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  $\bar{y}_g^{wheat} = \frac{1}{2}$ .

To aggregate the grid point crop information to the level of a field, we define  $\bar{y}_{i,j}^c$  to be the mean of  $\bar{y}_g^c$  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  $\bar{y}_{i,j}$  in our econometric model.<sup>12</sup>

---

<sup>10</sup>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 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.

<sup>11</sup>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.

<sup>12</sup>The selected data and programs used in the paper are posted at <http://strategy.sauder.ubc.ca/lee/research.html>.

Table 2 presents the summary statistics for  $\bar{y}_{i,j}^c$  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.

### 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 county 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, plus 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 counties. 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 on neighboring qualities  $\gamma_A$  and  $\gamma_B$  equal zero and it is possible to consistently identify (up to a multiplicative scalar) the field attribute coefficients  $\beta$  in (15) 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.<sup>13</sup>

## 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.

---

<sup>13</sup>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  $\theta_A$  and  $\theta_B$  and the field quality coefficients  $\beta$ . (The field quality coefficients  $\beta$  and  $\lambda'$  are scaled so that (9) holds, and  $\lambda'$  drops out.) We use the nonlinear least squares procedure discussed in Section 3 to estimate the structural parameters of equation (15) for each crop.

Our baseline estimates are obtained by estimating the model on a crop-by-crop basis jointly for all twelve counties together.<sup>14</sup> The density parameters  $\theta_A$ ,  $\theta_B$  are assumed to be the same in each county and the soil characteristics vector  $\beta$  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 (9), 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  $\gamma_O$  would equal one. It is apparent in the table that  $\gamma_O$  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  $\gamma_A$  on the two adjacent fields is .13 each, and the weight  $\gamma_B$  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  $\theta_A$  for adjacent fields is significantly positive. The parameter  $\theta_B$  is substantially smaller in each case. The last column contains  $\Theta$ , the composite density parameter (equal to  $2\theta_A + \theta_B$ ). Recall the interpretation for  $\Theta$  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.

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  $\Theta$  because this is a

---

<sup>14</sup>We use only those quarter sections with four complete fields.

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  $\Theta$ . There is little difference in the result. The average  $\Theta$ s 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 contain 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  $\Theta$  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 defined the within specialization measure and decomposed it into natural advantage share and density economy share. We do not directly observe actual soil qualities. 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 measure and shares. We evaluate means at the level of a county. For example, we calculate  $d^{natural\_advantage}$  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 within specialization measure captures specialization beyond what would happen with a dartboard. The measure would be zero if there were no density economies and soil quality were distributed i.i.d. within each county. The measure becomes one if all the four fields of each quarter section were planted exactly the same way. Table 6 shows that the measure is roughly .9 throughout various crops, which indicates a strong tendency for specialization within quarter sections. We decompose this specialization measure into the natural advantage share and the density economy share. Table 6 shows that roughly two thirds of the specialization is due to natural advantage and one third is due to density

economies.

## 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.<sup>15</sup> 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.<sup>16</sup>

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 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.<sup>17</sup> 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

---

<sup>15</sup>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.

<sup>16</sup>In the rare cases where there are different owner names within the same field, we assign ownership to the modal name.

<sup>17</sup>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.

typically quarter sections, though 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.<sup>18</sup> 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 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 the mean within quarter section deviation of soil quality for each crop, normalized by the mean plant-

---

<sup>18</sup>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.

ings for the crop.<sup>19</sup> 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 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  $\Theta$  from Table 4.<sup>20</sup>

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  $\Theta$  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

---

<sup>19</sup>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.)

<sup>20</sup>We get a slightly different number of observations because we use only the fields that have all three neighbors in the same *fake* squares.

story for why we are getting positive estimates for  $\Theta$  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  $\gamma_A$  and  $\gamma_B$  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  $\delta$  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 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  $\gamma_A$  and  $\gamma_B$  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.

One last thing to note is that planting patterns can change at ownership boundaries because different farmers may have different skills for different crops. We do not regard this point as an alternative explanation of the phenomenon we have identified but rather an instance of what we are emphasizing. If a crop is to be planted on a particular field, it is good to have a farmer to work the field whose skill set is a good match for the crop. But given the indivisibilities involved with farm labor, it will be desirable to have that same farmer work adjacent fields.

## 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.<sup>21</sup> 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 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.

---

<sup>21</sup>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.

# A Appendix

## A.1 Proof of Proposition 1

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

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

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

$$\begin{aligned}\bar{y}_1 - \bar{y} &= \frac{1 - 2\theta^2}{1 + 2\theta}x_1 + \frac{\theta}{1 + 2\theta}x_2 + \frac{\theta}{1 + 2\theta}x_3 + \frac{2\theta^2}{1 + 2\theta}x_4 - \frac{1 + 2\theta}{1 + 2\theta}\bar{x} \\ &= \frac{1}{1 + 2\theta} \left( \begin{array}{l} x_1 - 2\theta^2x_1 + \theta x_2 + \theta x_3 + 2\theta^2x_4 - \frac{x_1}{4} - \frac{x_2}{4} - \frac{x_3}{4} - \frac{x_4}{4} \\ -\frac{\theta}{2}x_1 - \frac{\theta}{2}x_2 - \frac{\theta}{2}x_3 - \frac{\theta}{2}x_4 \end{array} \right) \\ &= \frac{1}{1 + 2\theta} \left( \frac{3}{4}x_1 - 2\theta^2x_1 + \frac{\theta}{2}x_2 + \frac{\theta}{2}x_3 + 2\theta^2x_4 - \frac{x_2}{4} - \frac{x_3}{4} - \frac{x_4}{4} - \frac{\theta}{2}x_1 - \frac{\theta}{2}x_4 \right) \\ &= \frac{1}{1 + 2\theta} \left( x_1 - \bar{x} - 2\theta^2(x_1 - x_4) + \frac{\theta}{2}\Delta \right) \\ &= \frac{1}{1 + 2\theta}B_1\end{aligned}$$

for  $B_1$  and  $\Delta$  defined by

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

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

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

decreases in  $\theta$ . Clearly, the effect of  $\theta$  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 lines, we get

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

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

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

Adding these together yields

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

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 Calculation of the Dartboard Dispersion Measure

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

$$\begin{aligned} d^{\text{dartboard}} &= E[x_1 - \bar{x}]^2 \\ &= E \left[ \left( \frac{3}{4}x_1 - \frac{x_2}{4} - \frac{x_3}{4} - \frac{x_4}{4} \right)^2 \right] \\ &= \frac{12}{16} E[x_1^2] - \frac{12}{16} E[x_1]^2 \\ &= \frac{3}{4} \text{var}_x \\ &= \frac{3 \sum_{i=1}^N \sum_{j=1}^4 (x_{ij} - \bar{x})^2}{4N}, \end{aligned}$$

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 an easy way to calculate the standard errors (and covariance matrix) of the estimates  $(\gamma, \beta)$  using any statistical package. (See Davidson and MacKinnon (2004) for a more detailed explanation.) We express the estimation equation (15) as the following:

$$\bar{y}_1 = \mathbf{x}(\gamma, \beta) + u,$$

where  $\mathbf{x}(\gamma, \beta) = \gamma_O z'_1 \beta + \gamma_A (z'_2 + z'_3) \beta + \gamma_B z'_4 \beta$  and  $u = \gamma_O \varepsilon_1 + \gamma_A (\varepsilon_2 + \varepsilon_3) + \gamma_B \varepsilon_4$ .

Let  $(\hat{\gamma}, \hat{\beta})$  be the nonlinear least squares estimates and let  $\mathbf{X}(\gamma, \beta) = [\partial \mathbf{x}(\gamma, \beta) / \partial \gamma \quad \partial \mathbf{x}(\gamma, \beta) / \partial \beta]$ . The consistent estimator of the covariance matrix of the nonlinear estimator  $(\gamma, \beta)$  is

$$\widehat{Var}(\hat{\gamma}, \hat{\beta}) = s^2 (\hat{\mathbf{X}}' \hat{\mathbf{X}})^{-1}, \quad (16)$$

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

$$s^2 \equiv \frac{1}{n-k} \sum \hat{u}^2 = \frac{1}{n-k} \sum \left( y_1 - \mathbf{x}(\hat{\gamma}, \hat{\beta}) \right)^2. \quad (17)$$

An easy way to calculate the covariance matrix (16) is to regress  $\bar{y}_1 - \mathbf{x}(\hat{\gamma}, \hat{\beta})$  on  $\hat{\mathbf{X}} \equiv \mathbf{X}(\hat{\gamma}, \hat{\beta})$  and use the resulting covariance matrix.

$$\bar{y}_1 - \mathbf{x}(\hat{\gamma}, \hat{\beta}) = \hat{\mathbf{X}}\mathbf{b} + \text{residuals}$$

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

$$Var(\hat{\mathbf{b}}) = (s')^2 (\hat{\mathbf{X}}' \hat{\mathbf{X}})^{-1}, \quad (18)$$

where

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

Since the regressor does not have any explanatory power ( $\mathbf{b} = \mathbf{0}$ ),  $(s')^2$  is equal to  $s^2$  in equation (17). Therefore, it holds that

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

We can regress  $\bar{y}_1 - \mathbf{x}(\hat{\gamma}, \hat{\beta})$  on  $\mathbf{X}(\hat{\gamma}, \hat{\beta})$  in any statistical package and easily calculate  $Var(\hat{\mathbf{b}})$  and thus  $\widehat{Var}(\hat{\gamma}, \hat{\beta})$ .

## References

- Arzaghi, Mohammad and J. Vernon Henderson (2008). "Networking off Madison Avenue," *Review of Economic Studies*, vol. 75(4), pages 1011-1038, October.
- 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.casscountynynd.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 Great West*, New York: Norton.
- Davidson, Russell, and James G. MacKinnon (2004), *Econometric Theory and Methods*, New York: Oxford University Press.
- Durantón, 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 (2001), “The Determinants of Agglomeration,” *Journal of Urban Economics* 50(2), 191–229.
- 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

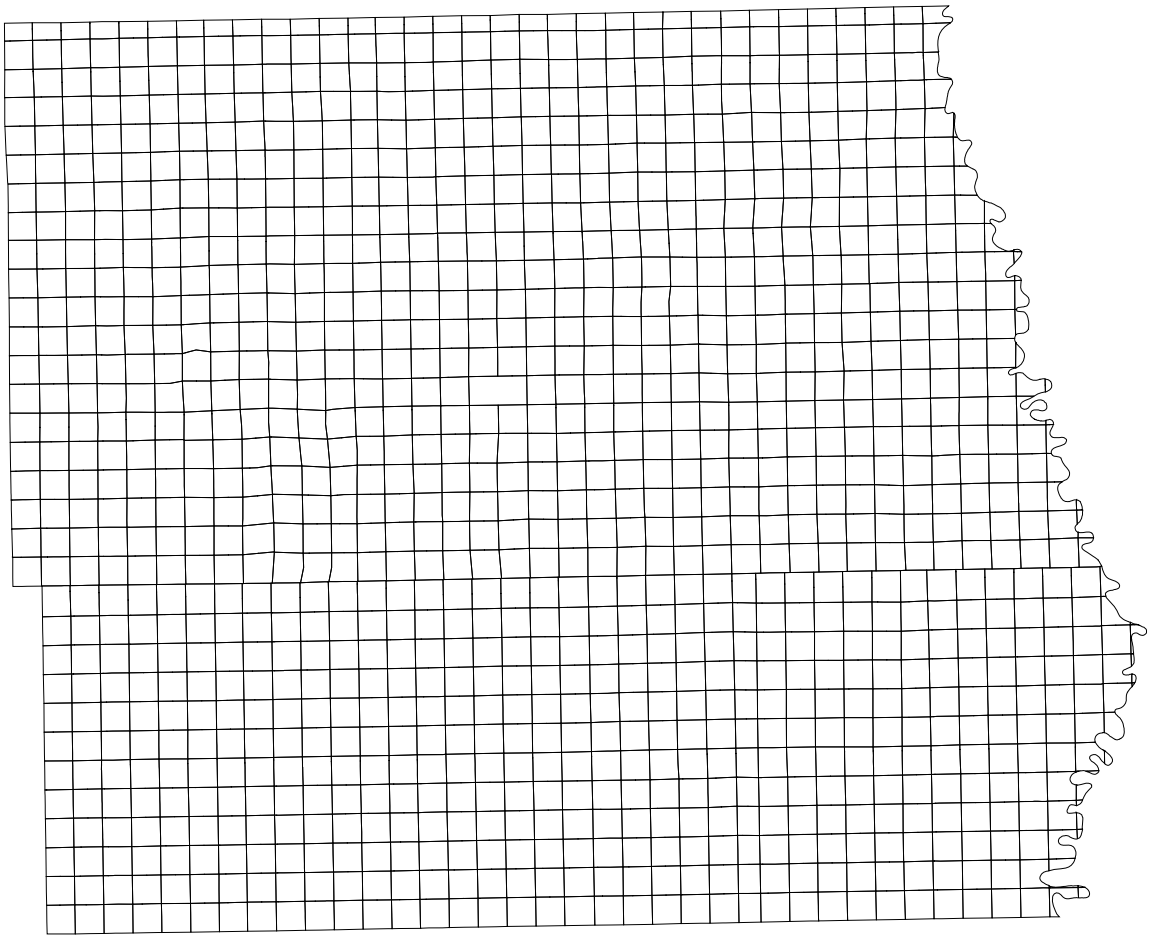


Figure 2. Counties Used in Our Analysis  
(North Dakota Red River Region)



Figure 3. Example Fields and Grid Points  
from Pembina County

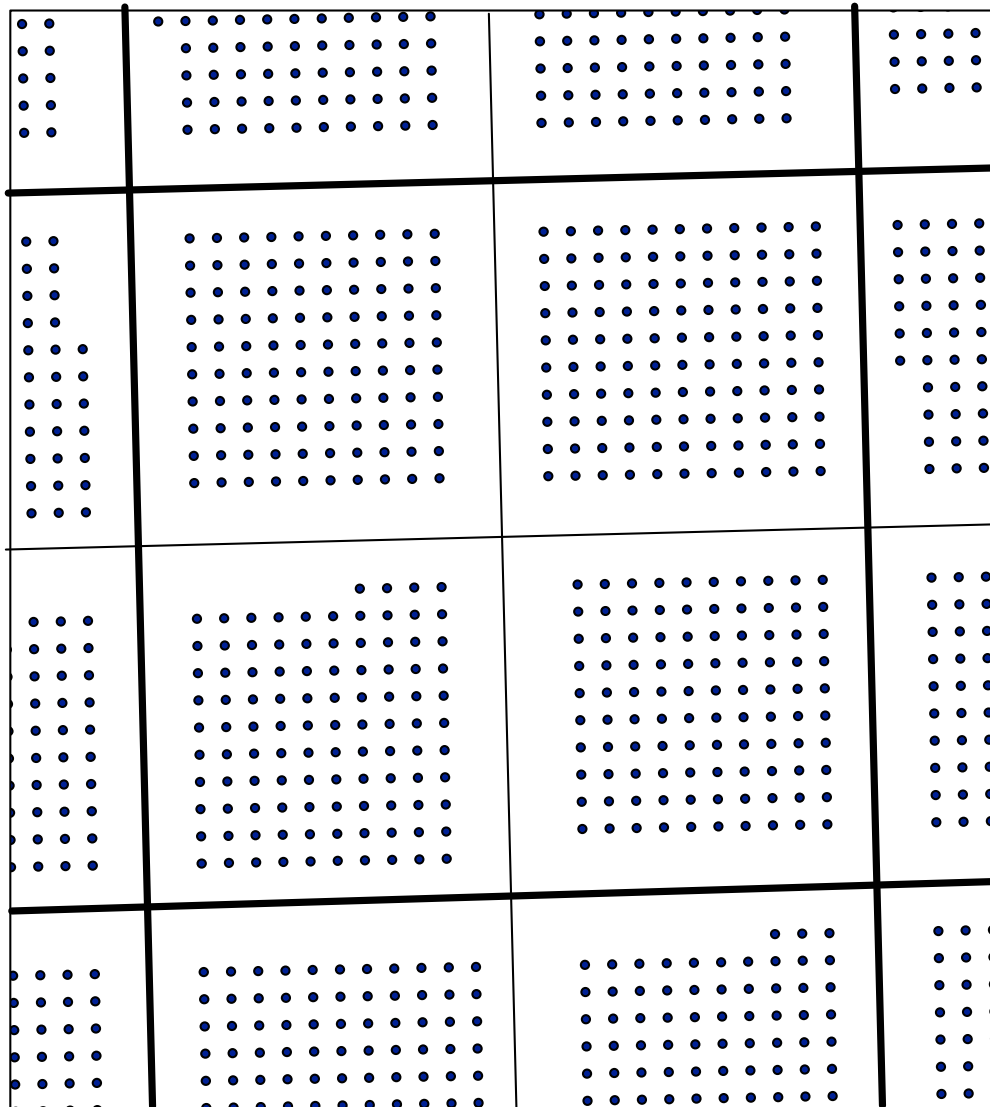


Figure 4a. Extended Neighborhoods

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

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

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

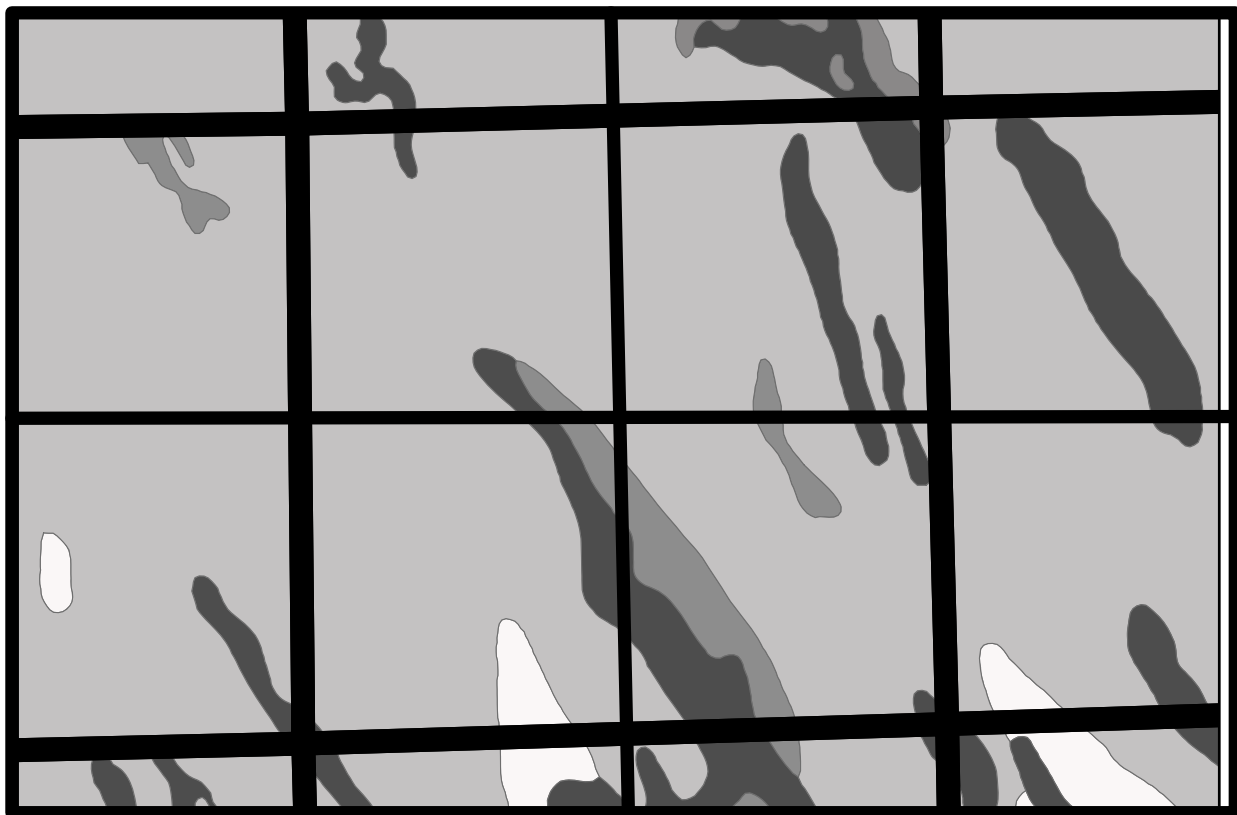


Figure 5b. Example Crop Map in a Quarter Section

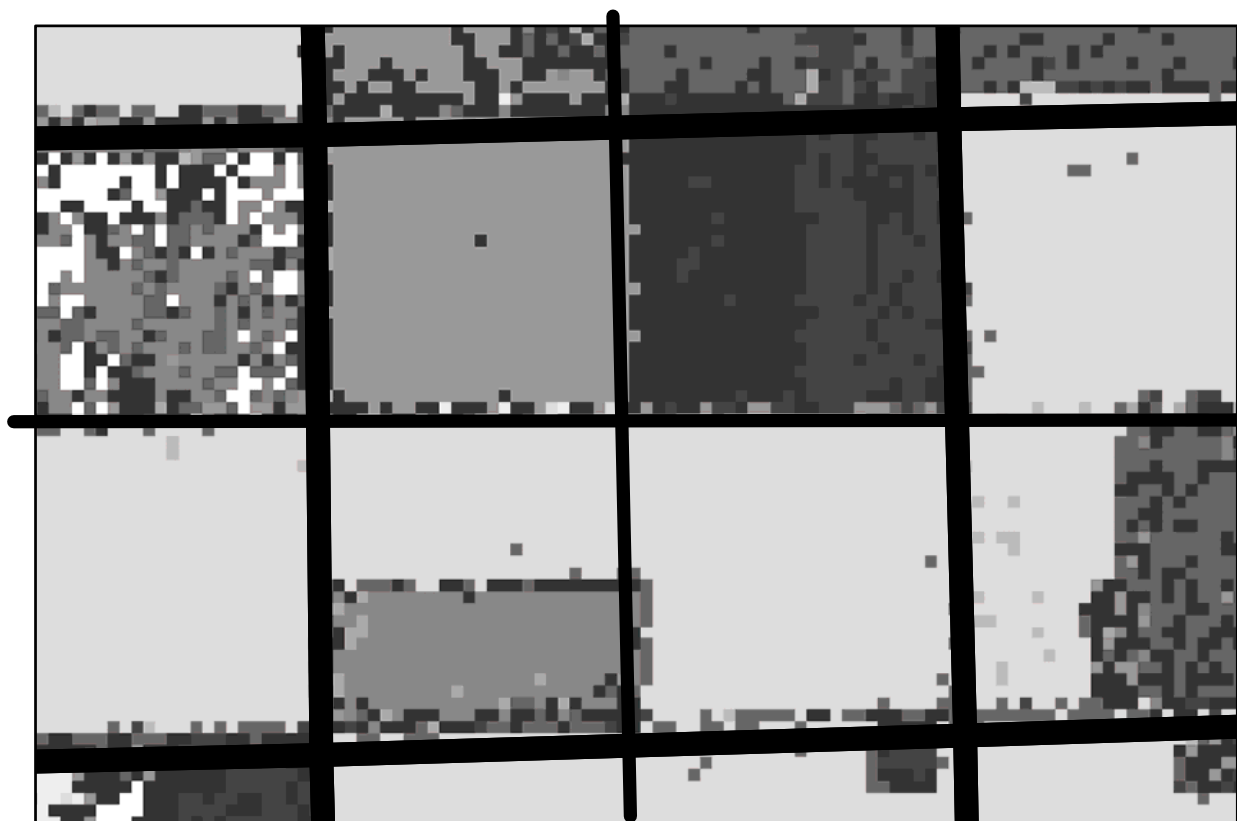


Table 1  
Land Use in North Dakota Red River Region  
Averages over 1997-2006 Period

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 and 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

County	Number of Fields	Spring Wheat				Soybean			
		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<sup>2</sup>s of Individual Regressions for Each of Twelve Counties)

Crop	Mean	Min	Median	Max
Spring Wheat	.335	.210	.322	.562
Soybeans	.237	.084	.246	.384
Corn	.182	.057	.196	.393
Dry Beans	.156	.057	.154	.337
Sunflowers	.108	.047	.108	.222
Barley	.115	.032	.110	.239
Durum Wheat	.102	.027	.054	.257
Other Small Grains	.179	.055	.168	.333
Canola	.078	.021	.044	.250
Beets	.152	.028	.129	.360
Other Selected Crops	.067	.020	.048	.165
Potatoes	.119	.041	.088	.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)

Crop	Policy Parameters			Structural Parameters		
	$\gamma_O$	$\gamma_A$	$\gamma_B$	$\theta_A$	$\theta_B$	$\Theta = 2\theta_A + \theta_B$
Spring Wheat	.658 (.006)	.128 (.004)	.087 (.006)	.160 (.008)	.070 (.012)	.389 (.008)
Soybeans	.632 (.007)	.137 (.005)	.094 (.007)	.174 (.011)	.074 (.017)	.422 (.010)
Corn	.459 (.007)	.199 (.005)	.144 (.007)	.316 (.027)	.040 (.044)	.672 (.014)
Dry Beans	.530 (.010)	.181 (.006)	.109 (.009)	.278 (.026)	.016 (.039)	.573 (.017)
Sunflowers	.613 (.012)	.147 (.008)	.093 (.011)	.197 (.020)	.056 (.030)	.451 (.017)
Barley	.757 (.011)	.095 (.007)	.053 (.010)	.113 (.011)	.041 (.017)	.267 (.014)
Durum Wheat	.721 (.011)	.115 (.007)	.048 (.010)	.147 (.014)	.019 (.020)	.313 (.015)
Other Small Grains	.519 (.008)	.183 (.005)	.114 (.008)	.284 (.023)	.020 (.035)	.588 (.015)
Canola	.602 (.013)	.143 (.009)	.112 (.013)	.180 (.021)	.100 (.033)	.460 (.019)
Beets	.366 (.012)	.225 (.008)	.184 (.012)	.408 (.088)	.002 (.152)	.818 (.028)
Other Selected Crops	.706 (.015)	.112 (.009)	.070 (.014)	.136 (.018)	.056 (.027)	.329 (.020)
Potatoes	.325 (.010)	.234 (.002)	.207 (.005)	.441 (.008)	0 (0)	.882 (.0015)

Table 5  
Robustness Check

Crop	Baseline Model	By County	Only on Prime Farm Land
	$\Theta$	Average $\Theta$	$\Theta$
Spring Wheat	.389 (.008)	.447 (.117)*	.578 (.022)
Soybeans	.422 (.010)	.433 (.147)*	.598 (.026)
Corn	.672 (.014)	.527 (.248)*	.825 (.028)
Dry Beans	.573 (.017)	.567 (.237)*	.503 (.032)
Sunflowers	.451 (.017)	.551 (.121)*	.702 (.045)
Barley	.267 (.014)	.536 (.244)*	.624 (.062)
Durum Wheat	.313 (.015)	.543 (.250)*	.818 (.057)
Other Small Grains	.588 (.015)	.574 (.229)*	.470 (.032)
Canola	.460 (.019)	.664 (.243)*	.870 (.035)
Beets	.818 (.028)	.759 (.178)*	.739 (.026)
Other Selected Crops	.329 (.020)	.588 (.266)*	.804 (.066)
Potatoes	.882 (.0015)	.730 (.148)*	.867 (.024)
N	208,220		111,596

\* Standard error calculated from county level  $\Theta$  estimates

Table 6. Natural Advantage vs. Density Economies Decomposition

Crop	Within Specialization Measure	Natural Advantage Share	Density Economy Share
Spring Wheat	.880	.690	.310
Soybeans	.891	.670	.330
Corn	.970	.674	.326
Dry Beans	.943	.669	.331
Sunflowers	.869	.544	.456
Barley	.762	.646	.354
Durum Wheat	.855	.761	.239
Other Small Grains	.942	.634	.366
Canola	.945	.802	.198
Beets	.995	.862	.138
Other Selected Crops	.895	.808	.192
Potatoes	.996	.689	.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		.87	.92	27,271
C		.01	.29	26,946
A	Yes	.87	.93	27,161
A	No	.34	.41	110
C	Yes	.01	.30	25,944
C	No	.00	.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

Crop	x variation		y variation	
	Actual	Fake	Actual	Fake
Spring Wheat	.136	.137	.230	.379
Soybeans	.120	.121	.257	.417
Corn	.249	.253	.391	.713
Dry Beans	.185	.188	.477	.724
Sunflowers	.125	.127	.481	.765
Barley	.132	.133	.471	.698
Durum Wheat	.132	.134	.506	.677
Other Small Grains	.244	.248	.621	.837
Canola	.166	.166	.469	.763
Beets	.249	.251	.450	.840
Other Selected Crops	.168	.169	.537	.857
Potatoes	.284	.287	.616	.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)		
	$\Theta$	$\Theta$	$\gamma_A$	$\gamma_B$	$\delta$	$\gamma_A$	$\gamma_B$	$\delta$
Spring Wheat	.389 (.008)	.318 (.008)	.129 (.004)	.085 (.006)	.592 (.041)	.192 (.016)	.089 (.024)	.836 (.028)
Soybeans	.422 (.010)	.368 (.011)	.137 (.005)	.094 (.007)	.703 (.048)	.157 (.014)	.053 (.021)	.773 (.034)
Corn	.672 (.014)	.558 (.014)	.199 (.005)	.144 (.007)	.509 (.066)	.197 (.016)	.147 (.023)	.606 (.065)
Dry Beans	.573 (.017)	.472 (.016)	.182 (.006)	.108 (.009)	.711 (.052)	.246 (.026)	.160 (.038)	.933 (.056)
Sunflowers	.451 (.017)	.311 (.018)	.148 (.008)	.093 (.011)	.761 (.074)	.234 (.023)	.132 (.035)	.782 (.067)
Barley	.267 (.014)	.188 (.014)	.097 (.007)	.050 (.010)	.354 (.101)	.241 (.034)	.266 (.052)	.824 (.045)
Durum Wheat	.313 (.015)	.244 (.015)	.118 (.007)	.045 (.010)	.000 (.091)	.184 (.034)	.255 (.050)	.919 (.056)
Other Small Grains	.588 (.015)	.449 (.013)	.183 (.005)	.114 (.008)	1.326 (.064)	.145 (.013)	.013 (.019)	1.642 (.184)
Canola	.460 (.019)	.375 (.019)	.155 (.008)	.152 (.012)	.000 (.061)	.241 (.039)	.152 (.058)	.992 (.081)
Beets	.818 (.028)	.759 (.026)	.228 (.008)	.181 (.012)	.781 (.043)	.229 (.024)	.182 (.035)	.874 (.060)
Other Selected Crops	.329 (.020)	.283 (.020)	.122 (.009)	.088 (.013)	.000 (.101)	.261 (.038)	.171 (.058)	1.014 (.157)
Potatoes	.888 (.032)	.862 (.015)	.235 (.008)	.205 (.011)	1.091 (.045)	.225 (.045)	.096 (.065)	1.164 (.146)
N	208,220	201,596		208,220			25,124	