The Review *of* Economics *and* Statistics

VOL. XCIV FEBRUARY 2012 NUMBER 1

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

Thomas J. Holmes and Sanghoon Lee*

*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 the same). Using rich geographic data from North Dakota, including new data on crop choice collected by satellite, we estimate a model of how a farmer plants adjacent fields under the farmer's control. We find planting decisions on a field are heavily dependent on the soil characteristics of adjacent fields. Through this relationship, we back out the structural parameters of economies of density.

## I. 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 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 scale economies? (See Cronon, 1991, on this issue.) Los Angeles specializes in making movies. Again, 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 scale economies?

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 in 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 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 subtle, scale economies 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 in the same way. Potential efficiencies extend beyond day-to-day field operations and include economies involving specialized equipment, like a sugar beet harvester. A farmer who acquires expensive equipment like this needs to use the equipment on many acres to justify its expense and ensure that it gets sufficient use. So the farmer who includes sugar beets in the crop rotation for one field 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 III.) 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 in the same way to fully use 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 (that is, planting 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 economy from the standard kind. We follow the literature in using this terminology, as we explain below.

We are drawn to study the factors underlying geographic concentration in agriculture because the features of this industry allow 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

Received for publication October 15, 2009. Revision accepted for publication July 22, 2010.

* Holmes: University of Minnesota, Federal Reserve Bank of Minneapolis, and NBER; Lee: Sauder School of Business, University of British Columbia.

T.J.H. is grateful to NSF Grant 0551062 for support of this research. We are grateful for input from the editor and the referees. We have benefited from the comments of seminar participants at the NBER agglomeration conference, University of British Columbia, the University of Minnesota, and the RSAI North American meetings. We are grateful to Julia Thornton Snider for research assistance. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

*The Review of Economics and Statistics,* February 2012, 94(1): 1–19

determine natural advantage, such as 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 on 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.[1] 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 these 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 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 40 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 40-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%.

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 independent 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 that 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 a violation of our exclusion restriction. This restriction is that unobserved exogenous characteristics are uncorrelated with observed exogenous characteristics (soil). We show that a violation of this assumption cannot be driving all of our results through a boundary analysis. We find that the link between a field's planting decision and its neighbor's

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

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 are 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 (for example, knowledge spillovers), leaving natural advantage in the background (see Rosenthal & 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 papers take 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 difficult 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, say, 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 (2011) provides a recent analysis of economies of density in WalMart's store location problem. The cost saving that WalMart 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 the 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 the cost savings achieved when farmers operate land parcels that are close together rather than dispersed.

## II. A Stripped-Down Model for Illustration

The full model includes a number of features and complications needed for estimating the model in the specific empirical context. To illustrate the economics of our approach, this section presents a stripped-down version of the model that highlights essential features. We first lay out the model and then provide a discussion.

### A. *The Model*

Consider a set of land parcels, and suppose that each parcel is divided into two adjacent fields. We index parcels by $i$ and the two fields of each parcel by $j \in \{1, 2\}$. For now, it is convenient to leave $i$ implicit and focus on the two fields of a particular parcel.

We model the planting decision of a particular crop and treat the alternatives to this crop as some sort of composite alternative, the details of which are left in the background. Formally, let $y_j$ be the quantity of the particular crop planted on field $j$. We think of the choice $y_j$ as a continuous variable, that is, the fraction of the land allocated to the particular crop, rather than a discrete variable of planting the whole field one way or another. In the full model of the next section, we will be examining long-run averages of crop planting, which matters because of crop rotations. If a particular crop is planted on field $j$ once every two years, then $y_j = \frac{1}{2}$; if twice over a three-year cycle, then $y_j = \frac{2}{3}$; and so on. Moreover, even within a particular year, sometimes only part of a field is planted with a particular crop, and our data are sufficiently fine for measuring this detail. Given the wide variety of possibilities in the choice set, we think it is a sensible modeling strategy to treat the choice $y_j$ as continuous because it facilitates tractability.

Next, let $q_j$ denote the productivity of field $j$, and let $\lambda$ be the output price. The revenue on field $j$ from planting the particular crop at level $y_j$ equals

$$revenue_j = \lambda q_j y_j.$$

The productivity of field $j$ is the sum of four components:

$$q_j = z_j + r + u_j + c.$$

The variable $z_j$ is the measured soil quality of field $j$ for the particular crop. Assume this is a scalar that is directly observed in the data. The terms $r + u_j$ together denote the unmeasured quality for field $j$. Note that $r$ is common across both fields. We could subsume $r$ into $u_j$, but we keep it separate for expositional purposes. While $r$ is constant for the two adjacent fields of a given parcel, in general it varies across parcels.

The final term $c$ of field productivity is a choice variable that is selected to maximize farm profit. The choice is required to be the same across both fields, so there is no $j$ subscript. The variable $c$ represents an investment, and we assume the cost is quadratic:

$$investment\_cost(c) = \frac{1}{2\rho}c^2. \tag{1}$$

The investment $c$ can be interpreted as a choice of a particular kind of equipment that is specific to the crop. It can also be interpreted as a kind of endogenous farmer quality, as we discuss further below. A key point is that there is a scale economy. The investment in $c$ gets paid once but is used in both fields. The investment is nonrivalrous in that using it on field 1 does not take away from its contribution to productivity on field 2.

We assume quadratic planting costs for each field:

$$planting\_cost(y_j) = \frac{\xi}{2}y_j^2.$$

The cost includes the opportunity cost of using the land for the particular crop under discussion as opposed to the composite alternative.

Total profit across the two fields equals revenue minus the planting cost on each field minus the investment cost:

$$profit = \lambda q_1 y_1 - \frac{\xi}{2}y_1^2 + \lambda q_2 y_2 - \frac{\xi}{2}y_2^2 - \frac{1}{2\rho}c^2. \tag{2}$$

The plantings $y_1$ and $y_2$ and the investment $c$ are chosen jointly to maximize profit. To highlight the economics, we proceed in a two-stage fashion. We first maximize out the investment choice $c$, given planting choices $y_1$ and $y_2$. Next we pick plantings. The first-order necessary condition for the subproblem of maximizing with respect to $c$ yields

$$c = \lambda\rho(y_1 + y_2).$$

Substituting this choice into equation (2) yields maximized profit given plantings,

$$profit^*(y_1, y_2) = \left[\lambda x_1 y_1 - \frac{1}{2}y_1^2 + \frac{1}{2}\theta y_1 y_2\right] + \left[\lambda x_2 y_2 - \frac{1}{2}y_2^2 + \frac{1}{2}\theta y_1 y_2\right], \tag{3}$$

for

$$x_j \equiv z_j + u_j + r$$
$$1 \equiv \xi - \rho\lambda^2$$
$$\theta \equiv \rho\lambda^2.$$

The profit, equation (3), that maximizes out investment $c$ is written as the sum of profits on each field, including a term $\frac{1}{2}\theta y_1 y_2$ for each field that we denote the *economies of density*. The parameter $\theta$ governs the extent of the economies of density. The variable $x_j$ is the sum of measured and unmeasured field quality. The condition $1 \equiv \xi - \rho\lambda^2$ is a normalization that we make without loss of generality.

Next we maximize equation (3) over the choice of $y_1$ and $y_2$. From the first-order condition for the choice of $y_1$, we get

$$y_1 = \lambda x_1 + \theta y_2. \tag{4}$$

This equation highlights the connection between the plantings on the two fields. Holding field 1 quality $x_1$ fixed, planting on field 1 is higher with more planting $y_2$ on field 2. In the background as $y_2$ is increased, the investment choice $c$ increases. In turn, the marginal product of plantings on field 1 rises. We can take the analogous first-order condition for $y_2$ and solve the two equations for the optimal choice of $y_1$ as a function of the quality of each field,

$$y_1 = \frac{\lambda}{1-\theta^2}x_1 + \frac{\lambda\theta}{1-\theta^2}x_2 \tag{5}$$
$$= \frac{1}{1+\theta}x_1 + \frac{\theta}{1+\theta}x_2, \tag{6}$$

and derive the analogous condition for $y_2$. We note that we require the density parameter to satisfy $\theta < 1$ for the solution to be well defined. The second line imposes a normalization $\lambda = 1 - \theta$ on the output price, which is without loss of generality.

We now bring back the subscript $i$ for the different parcels, break down total field quality into observed and unobserved components, $x_{i,j} = z_{i,j} + r_i + u_{i,j}$, and rewrite equation (5) as

$$y_{i1} = \frac{1}{1+\theta}z_{i,1} + \frac{\theta}{1+\theta}z_{i,2} + \varepsilon_{i,1} \tag{7}$$

for

$$\varepsilon_{i,1} \equiv \frac{1}{1+\theta}\left(r_i + u_{i,1}\right) + \frac{\theta}{1+\theta}\left(r_i + u_{i,2}\right).$$

The error term $\varepsilon_{i,1}$ in equation (7) is a function of the unobserved quality components of both fields. Assume the error term $\varepsilon_{i,1}$ is orthogonal to the observed quality attributes,

$$E[\varepsilon_{i,1} z_{i,j}] = 0, \text{ for } j \in \{1, 2\}, \tag{8}$$

and make the analogous assumption on $\varepsilon_{i,2}$. We make no assumption on the correlation of the error term across fields. In particular, because of the presence of the common $r_i$ of both error terms, we expect

$$E[\varepsilon_{i,1}\varepsilon_{i,2}] > 0. \tag{9}$$

We run an OLS regression of $y_{i,1}$ on $z_{i,1}$ and $z_{i,2}$ (and the analogous regression of $y_{i,2}$ on $z_{i,2}$ and $z_{i,1}$), and under mean independence, equation (8), OLS provides consistent estimates of the slopes $\frac{1}{1+\theta}$ and $\frac{\theta}{1+\theta}$ in equation (7), from which we solve for an estimate of $\theta$.

*B. Discussion*

The model is a variant of the linear-in-means social interactions model exposited in the survey paper of Brock and Durlauf (2001b).[2] 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? One thing to note is that we are studying a planner's problem of maximizing joint profit across neighboring fields. This is different from a typical social interactions setup where the choice made for a particular unit does not take into account the externality on neighboring units. In that literature, Nash equilibrium, rather than joint-profit maximization, is the appropriate concept (Brock & Durlauf, 2001a; Bajari et al., 2010). At the level of our analysis, neighboring fields are generally part of the same farm operation, so joint-profit maximization is the appropriate concept. And even if neighboring fields are operated by different firms, it is plausible to model the neighbors as cooperating to maximize joint surplus, for example, by sharing equipment.

To relate our model to this literature, it is useful to recall formula (4) for how the optimal planting of field 1 depends on the quality of field 1 and the planting choice on field 2,

$$y_{i,1} = \lambda z_{i,1} + \theta y_{i,2} + \lambda(r_i + u_{i,1}), \tag{10}$$

where we now separate out field quality into its observed component $z_{i,1}$ and its unobserved component $r_i + u_{i,1}$. The case of $\theta > 0$ where there are density economies is an endogenous effect in the literature (Manski, 1993). Optimal planting on field 1 is higher the higher the planting on field 2, which itself is an endogenous variable. Note that if we took equation (10) and regressed field 1's planting on its own measured quality $z_{i,1}$ and field 2's planting $y_{i,2}$, the estimated coefficient on $y_{i,2}$ would be an inconsistent estimate of $\theta$. This follows because the error term in equation (10) is correlated with the choice $y_{i,2}$. That is, the unobserved quality components of the two fields are correlated, and this implies that planting choices of the neighboring fields would tend to be similar even if $\theta = 0$. That is why our strategy is to use the neighbors' exogenous characteristics rather than the neighbors' choices to identify $\theta$. This strategy is well understood in the literature.

We have ruled out what Manski (1993) calls "exogenous effects." These exist if planting on field 1 in equation (10) depends not only on the planting choice on field 2, but also directly on field 2 characteristics. It is possible to think of stories where exogenous effects might matter.[3] However, these stories strike us as second order in nature for our application. Because they complicate identification, we zero out exogenous effects a priori.

We have allowed for what Manski (1993) calls "correlated effects," which here is the correlation in unobservable quality of neighboring fields, condition (9). While our soil data are extremely rich, obviously we do not observe all aspects of quality, and there are good reasons to expect neighboring fields to be similar in unobserved characteristics. (And we have built this in with the common component $r_i$ across fields for parcel $i$.) Relatedly, we get consistent estimates if we allow for measurement error that is correlated across neighboring fields in the observed planting level $y_{i,j}$. This happens, for example, when clouds block the satellite view of neighboring fields.

If measured quality is the same on both fields, $z_{i,1} = z_{i,2}$, the regressors in equation (7) are collinear and $\theta$ is not identified. This case boils down to the "reflection problem" of Manski, where endogenous effects ($\theta$) and correlated effects cannot be separately identified. In our data, however, there are differences in measured soil quality across fields in the same parcel. This fact, along with the exclusion restrictions that we have made, permits us to identify $\theta$.

Having presented the model and linked it to the literature, we turn now to interpretations. The role that soil quality plays in the analysis is straightfoward to see, but how should we interpret the role of the farmer? In particular, there may be variation across farmers in their suitability for particular crops (for example, because of differences in training or access to capital markets). To the extent that differences across land parcels in farmer talent are exogenous (for example, from inheritance), this is captured by the unobserved quality term $r_i$ that is the same for both fields. To the extent that the differences are endogenous—farmers with different talents select different fields—this is captured by the investment $c$ described in the model. The lumpiness in farmer talent leads to precisely the kind of density economies we intend to capture in our model. If field 1 is particularly suitable for the crop in question, it is efficient to find a farmer for field 1 who is a good match for the crop by setting a high level of investment $c$. But because the farmer for field 1 is the same as for field 2, it will be efficient to have high levels of planting on field 2 as well. Thus, the selection of farmer quality is conceptually the same as interpreting the investment $c$ as crop-specific equipment or crop-specific human capital.

One final issue is why the existence of crop-specific equipment creates a spatial effect. If equipment could be rented in a frictionless market with zero transportation costs, there would be no spatial effect. One reason a spatial effect emerges is that farm equipment is in fact bulky and hard to move around. In particular, once the fixed cost is paid of hauling in a piece of machinery from somewhere else, the machine can be used on a neighboring field, without hauling it in a second time. A second reason is that rental markets are not frictionless, so farmers tend to own equipment. If two adjacent fields are farmed by the same person (and it is easy to see the advantage of contiguous management) and if that person owns his or her own equipment, then the equipment available for use on the adjacent fields is thereby linked.

[2] Actually, in this stripped-down model, "means" across neighbors are degenerate because each field has only one neighbor. In the full model, fields will have more than one neighbor over which to take means.

[3] 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.

## III. The Full Model

In order to take the model to the data, we enrich the stripped-down model in four ways.

First, we generalize the model so that there are four fields instead of two. In particular, the fields form the four quadrants of a square. We explicitly take into account that the relationship between diagonal fields may be different from adjacent fields, and this detail adds to the complexity.

Second, we take into account that farmers typically pick crop rotations rather than individual crops. For example, a farmer growing sugar beets typically plants only beets every four years, rotating in other crops during the intervening years. One possibility would be sugar beets (year 1), wheat (year 2), barley (year 3), and wheat (year 4). 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 and 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 (say, from continuously operating a plow on adjacent fields) and those that are longer term in nature (that involve investment in long-lived equipment or investment in high farmer match quality, for example). In our model, we explicitly differentiate these short-run and long-run considerations. However, it is useful to alert readers upfront that in our estimation, we will be unable to separately identify these two factors. As we explained in section II, our identification strategy relies on relating planting decisions to the soil quality characteristics of adjacent fields. 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 able to identify only the combined effect of short-run and long-run economies. For the exercises we consider, this is sufficient.

Fourth, in the stripped-down model, measured soil quality is a scalar. In our data, we observe a vector of soil characteristics. We show with the full model how to take this into account.

Section III A presents the enrichments to the basic model of the previous section. Section III B develops an interpretation of the estimate of density economies in terms of what would happen to plantings if "walls" between fields were erected to block density economies. And section III C shows how the estimates can be used to separate out the roles of natural advantage and density economies in accounting for observed specialization.

### A. *The Enrichments to the Model*

A square land parcel has four quadrants (figure 1). We refer to the quadrants as fields and index them by $j \in \{1, 2, 3, 4\}$.

FIGURE 1.—SQUARE LAND PARCEL

| | |
|---|---|
| 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, and a parcel is a 160-acre quarter section.

There are $T$ periods. For each field $j$ and period $t$, the farmer chooses a planting level $y_{j,t}$ for the particular crop. As before, let $x_j$ be the field quality (measured and unmeasured) for growing the particular crop at field $j$. Here, we call this the *permanent quality measure*. In each period $t$, there is also a *transient quality measure* $\omega_{j,t}$ for field $j$. Because this term can vary over time, it introduces a force in the model to induce the farmer to cycle the planting of the crop by practicing 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 then the sum $x_j + \omega_{j,t}$ of the permanent and transient components. Since the $\omega_{j,t}$ sum to 0 for each $j$, the long-run average quality of field $j$ equals $x_j$.

In the previous section, we developed a micromodel of an investment $c$ and then maximized this choice out to obtain the reduced-form profit function, equation (3). The density parameter $\theta$ that emerged in this profit function is our key parameter of interest, and it is possible to start with the profit function as the primitive. Here, for simplicity, we skip the step of an underlying model of investment, going directly to the reduced-form profit that has the maximized investments in the background. In particular, the analog of equation (3) here is the sum of profit over the four fields and over all the time periods,

$$profit^* = \sum_{t=1}^{T} (\pi_{1,t} + \pi_{2,t} + \pi_{3,t} + \pi_{4,t}), \tag{11}$$

where, for simplicity, we are ignoring discounting and where the profit $\pi_{1,t}$ on field 1 in period $t$ can be written as

$$\begin{aligned}\pi_{1,t} = {} & \lambda(x_1 + \omega_{1,t})y_{1,t} - \frac{1}{2}y_{1,t}^2 + \frac{1}{2}\phi^L y_{1,t}\bar{y}_1 \\ & + \frac{1}{2}\theta_A^S y_{1,t}(y_{2,t} + y_{3,t}) + \frac{1}{2}\theta_A^L y_{1,t}(\bar{y}_2 + \bar{y}_3) \\ & + \frac{1}{2}\theta_B^S y_{1,t} y_{4,t} + \frac{1}{2}\theta_B^L y_1 \bar{y}_4 \end{aligned} \tag{12}$$

for $\bar{y}_j = (\sum_t y_{j,t})/T$. The profit on the other fields is the symmetric equivalent to equation (12).

Observe first that the profit on field 1 depends on 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 on the interactions of its own planting $y_{1,t}$ with the plantings on the other fields. Fields 2 and 3 are the A neighbors (directly adjacent) to field 1, and the coefficients on these interactions are $\theta_A^S$ and $\theta_A^L$. The parameter $\theta_A^S$ captures short-run density economies between adjacent fields. If $\theta_A^S > 0$, then plantings on adjacent fields in the same period are complements. The parameter $\theta_A^L$ 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 (on the diagonal), and the coefficients on this interaction are $\theta_B^S$ and $\theta_B^L$. The parameters $\theta_B^S$ and $\theta_B^L$ have the analogous roles for diagonal neighbors. Note that 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*.

Maximizing equation (11) with respect to $y_{1,t}$ yields the first-order necessary condition (the other fields are symmetric):

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

Summing up the first-order conditions over time, we obtain

$$0 = \lambda' x_1 - \bar{y}_1 + \theta_A (\bar{y}_2 + \bar{y}_3) + \theta_B \bar{y}_4, \tag{13}$$

where

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

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 (13) 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

$$\bar{y}_1 = \gamma_O x_1 + \gamma_A x_2 + \gamma_A x_3 + \gamma_B x_4, \tag{14}$$

where

$$\gamma_O = \frac{\lambda'\left(1 - 2\theta_A^2 - \theta_B\right)}{(1 - 2\theta_A - \theta_B)(1 + 2\theta_A - \theta_B)(1 + \theta_B)},$$
$$\gamma_A = \frac{\lambda' \theta_A}{(1 - 2\theta_A - \theta_B)(1 + 2\theta_A - \theta_B)},$$
$$\gamma_B = \frac{\lambda'\left(2\theta_A^2 + \theta_B - \theta_B^2\right)}{(1 - 2\theta_A - \theta_B)(1 + 2\theta_A - \theta_B)(1 + \theta_B)}.$$

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:

$$\lambda' = \frac{(\gamma_O - \gamma_B)\left(\gamma_O^2 - 4\gamma_A^2 + 2\gamma_O\gamma_B + \gamma_B^2\right)}{\gamma_O^2 - 2\gamma_A^2 + \gamma_O\gamma_B},$$
$$\theta_A = \frac{\gamma_A(\gamma_O - \gamma_B)}{\gamma_O^2 - 2\gamma_A^2 + \gamma_O\gamma_B},$$
$$\theta_B = \frac{-2\gamma_A^2 + \gamma_B(\gamma_O + \gamma_B)}{\gamma_O^2 - 2\gamma_A^2 + \gamma_O\gamma_B}.$$

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

$$\Theta \equiv 2\theta_A + \theta_B. \tag{15}$$

The last enrichment to the model is to take into account that rather than observing a scalar quality index $x_{i,j}$, we observe a vector of characteristics $z_{i,j}$ with $K$ elements for each field $j$ on each parcel $i$. 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

$$x_{i,j} = z_{i,j}'\beta + u_{i,j}, \tag{16}$$

where the weight vector $\beta$ on characteristics is unknown and where we subsume the earlier variable $r_i$ into the $u_{i,j}$. Let $u_i \equiv (u_{i,1}, u_{i,2}, u_{i,3}, u_{i,4})$ be the vector of unobserved quality components for parcel $i$, and, analogously, let $z_i \equiv (z_{i,1}, z_{i,2}, z_{i,3}, z_{i,4})$. Make the orthogonality restriction,

$$E[u_i z_i'] = 0. \tag{17}$$

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 $u_i$ that we miss. And the error $u_i$ is unrelated to the soil characteristics that we do observe.

Inserting equation (16) into equation (14) yields

$$\bar{y}_1 = \gamma_O z_1'\beta + \gamma_A\left(z_2' + z_3'\right)\beta + \gamma_B z_4'\beta + (\gamma_O u_1 + \gamma_A(u_2 + u_3) + \gamma_B u_4). \tag{18}$$

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 becomes linear in $\beta$ and we can run an OLS regression 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.

### B. *The Impact of a Wall on Plantings*

We provide an interpretation of the composite density parameter $\Theta$ by showing how it summarizes the impact of a hypothetical policy experiment shutting down density economies. Imagine a wall is erected in a particular parcel that separates all four fields of the parcel, eliminating all potential for density economies. Formally, after the wall, short-run and long-run density economies are 0 for both $A$ and $B$ neighbors:[4]

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

Average planting across the four fields equals

$$\begin{aligned}\bar{y} &\equiv \frac{\bar{y}_1 + \bar{y}_2 + \bar{y}_3 + \bar{y}_4}{4} \\ &= \frac{1}{4}(\gamma_O + 2\gamma_A + \gamma_B)(x_1 + x_2 + x_3 + x_4) \\ &= \frac{\lambda'}{1 - 2\theta_A - \theta_B} \frac{x_1 + x_2 + x_3 + x_4}{4} \\ &= \left(\frac{\lambda'}{1 - \Theta}\right) \frac{x_1 + x_2 + x_3 + x_4}{4} \\ &= \frac{\lambda'}{1 - \Theta}\bar{x}, \end{aligned} \tag{19}$$

where $\bar{y}$ is average planting, $\bar{x}$ is average quality, and again $\Theta$ is the composite density parameter in equation (15). We assume that

$$1 - \Theta > 0, \tag{20}$$

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

$$\frac{\lambda'}{1 - \Theta} = 1. \tag{21}$$

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 equation (21), we see from (19) 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 0, $\Theta^{wall} = 0$, 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 that 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 have an impact on not only average planting across the four fields, but also the dispersion of plantings across the four fields. We use the within-parcel variance as our dispersion measure:

$$d = \frac{\sum_{j=1}^{4}(\bar{y}_j - \bar{y})^2}{4}. \tag{22}$$

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

**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 equation (21) to leave average plantings fixed. The variance measure $d(\Theta)$ strictly declines in $\Theta$ and goes to 0 as $\Theta$ approaches its theoretical upper bound of 1.

**Proof.** See the supplementary web appendix. Available online at http://www.mitpressjournals.org/doi/suppl/10.1162/REST_a_00149.

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

### C. *The Within Specialization Measure*

As we will see below, farmers tend to plant the four fields of a parcel the same way. (A parcel is a quarter section.) 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 equation (22) of plantings across the four fields within a parcel. Suppose we have a set of $N$ parcels in an area (say, a county), with each parcel indexed by $i = 1, \ldots, N$. Let the county mean within dispersion measure be

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

[4] 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. For this discussion, we set $\Theta = 0$ for one particular parcel, holding it fixed in other parcels. Otherwise, this aggregate change in technology might affect prices and ultimately the $\lambda$ parameter.

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

The mean within dispersion measure $d^{county}$ is closely related to specialization. When the measure is very small, fields within a parcel tend to be planted in 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, the planting of an individual field is arbitrarily set to its field quality $y_{ij}^{dartboard} = x_{ij}$, and all $4N$ fields in the county are randomly reshuffled into groups of four that we call dartboard parcels (as opposed to actual parcels). This idea of taking into account random dartboard factors follows Ellison and Glaeser (1997). In the Supplementary Appendix on the Web, we show that if we were to calculate the county dispersion measure for this hypothetical exercise, for large $N$ it would approximately equal

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

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 parcel 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 parcel 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, 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 parcel to be more highly correlated than across parcels. 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 parcel $i$, 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

$$Natural\ advantage\ share = \frac{W^{natural\_advantage}}{W}, \tag{23}$$

while the balance of credit goes to density economies,

$$Density\ economy\ share = \frac{W - W^{natural\_advantage}}{W}. \tag{24}$$

## IV. Data

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

### *A. Fields*

The Public Land Survey System imposed a grid of squares on 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). The grid consists of a hierarchy of different size squares. There are four quarter sections (1/2 mile a side) in a section (1 mile a side). There are 36 sections in a township (6 miles a side). Figure 2 illustrates the section grid for Pembina County, one of the counties in North Dakota included in our study.[6] The eastern boundary of the county is irregular, following the Red River. The northern boundary meets Canada, so the top row is not full height. Otherwise the section grid is a virtually perfect system of 1- by 1-mile squares.

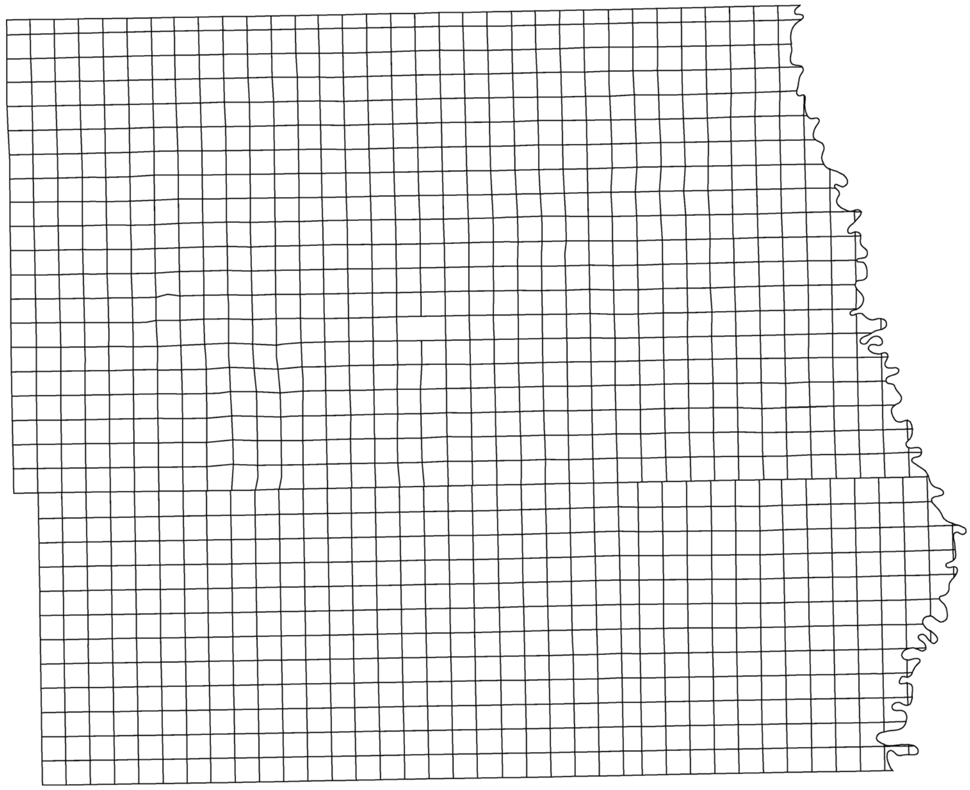
We will analyze the farmer's problem at the quarter section level: a quarter section corresponds to a parcel in the model, and fields are quarter quarters that are 40 acres, or 1/16 of a square mile. Our crop and soil data are at a higher resolution than the field level, and we could, in principle, have allowed for smaller decision-making units, for example, 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, defined 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

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

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

FIGURE 2.—SECTION BOUNDARIES, PEMBINA COUNTY, NORTH DAKOTA



illustrated in figure 3.[7] (They are also listed as part of table 2.) There are 231,000 fields in the region.[8] Equivalently, there are $14,400 \approx 231,000/16$ square miles, since there are 16 fields in each square mile section.

### B. *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 on the cooperation of state agencies. The availability of many years of data with which we can determine long-run average land use is an important reason that we picked North Dakota over other states.

The resolution of the crop data is at the level of 30 meter by 30 meter squares (approximately 4 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 to 2006 sample period.[9] 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% 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 4 illustrates some fields in Pembina County and the grid points they contain. The dark lines are the quarter section boundaries, and the lighter lines are field boundaries within a quarter section. As can be seen in

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

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

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

FIGURE 3.—COUNTIES USED IN OUR ANALYSIS, NORTH DAKOTA RED RIVER REGION


TABLE 1.—LAND USE IN NORTH DAKOTA RED RIVER REGION AVERAGES, 1997–2006

| 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 and hay (oats, millet, rye and winter wheat, alfalfa and 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 |

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

figure 4, we trim off the points near the border of each field and use only interior points. We want to be careful not to misclassify a point near a border as being in an adjacent field. Each field side is .25 miles (400 meters). We trim the points that are .03 miles (48 meters) on each side of the border. Each field has approximately 100 points in the interior.[10]

Suppose crops are indexed by $c$ and grid points indexed by $g$. Let $y^c_{g,t} = 1$ if crop $c$ is planted at grid point $g$ in year $t$ and set it to 0 otherwise. Let $\bar{y}^c_g$ be the mean value of $y^c_{g,t}$ over the years in the sample. For example, if wheat is planted at grid point $g$ in five of the ten years, then $\bar{y}^{wheat}_g = \frac{1}{2}$. Farmers in the region commonly practice crop rotation, and one such rotation is to alternate between wheat and soybeans. If this rotation is practiced at point $g$, then $\bar{y}^{wheat}_g = \frac{1}{2}$.

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

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

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

FIGURE 4.—EXAMPLE FIELDS AND GRID POINTS FROM PEMBINA COUNTY

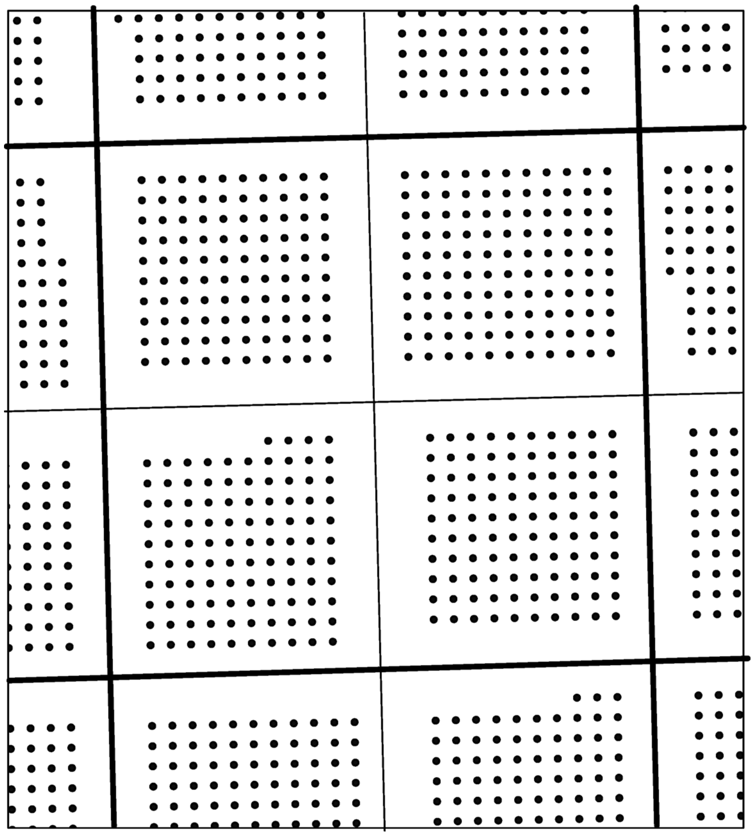


TABLE 2.—SUMMARY STATISTICS OF CROP PLANTING SHARE AT THE FIELD LEVEL FOR TWO MAJOR CROPS IN NORTH DAKOTA RED RIVER REGION

| | | Spring Wheat | | | | Soybean | | | |
|---|---|---|---|---|---|---|---|---|---|
| County | Number of Fields | Mean | S.D. | Minimum | Maximum | Mean | S.D. | Minimum | Maximum |
| 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 |

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

(.23 in Sargent, .27 in Richland). There is also substantial variation across fields within each county.

### *C. Soil*

Our soil data are taken from the Soil Survey Geographic (SSURGO) database maintained by the U.S. Department of Agriculture. A single 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 grid that is 30 meters by 30 meters 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 12 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 0, and it is possible to consistently identify (up to a multiplicative scalar) the field attribute coefficients $\beta$ in equation (18) through an 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 12 counties is .34 for spring wheat and .24 for soybeans. The $R^2$ is less for the smaller crops, but it is still nonnegligible. 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 affect 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 affect 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 lead to erosion and result in changes in near-surface soil properties. However, in the Red River Valley region, farmers have tended to use similar, proven practices on this valuable land. Most of the soils in the Red River Valley did not suffer the Dust Bowl erosion problems in the 1930s, as areas farther west and south did. According to Mike Ulmer, the USDA-NRCS senior regional soil scientist for the Northern Great Plains, most significant variations in soil variables for the Red River Valley region are due to natural soil genesis rather than human behavior.[12]

Table 3.—Goodness of Fit of Planting Regression on Soil Characteristics
Summary Statistics of $R^2$s of Individual Regressions for Each of 12 Counties

| Crop | Mean | Minimum | Median | Maximum |
|---|---|---|---|---|
| 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 |

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

## V. Basic Analysis

This section conducts the basic empirical analysis. Section V A estimates the structural model parameters. Section V B uses the model estimates to examine the contributions of density economies and comparative advantage to specialization.

### A. *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 equation (21) holds and $\lambda'$ drops out.)

Our baseline estimates are obtained by estimating the model on a crop-by-crop basis jointly for all 12 counties together.[13] The density parameters $\theta_A$, $\theta_B$ are assumed to be the same across all counties 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.)

While equation (14) specifies $\bar{y}_1$ as the regressor, we use the analogous equations for the plantings on the other fields

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

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

TABLE 4.—BASELINE MODEL ESTIMATES FOR POLICY AND STRUCTURAL PARAMETERS ($N = 208,220$)

| | Policy Parameters | | | Structural Parameters | | |
|---|---|---|---|---|---|---|
| Crop | $\gamma_O$ | $\gamma_A$ | $\gamma_B$ | $\theta_A$ | $\theta_B$ | $\Theta = 2\theta_A + \theta_B$ |
| Spring wheat | .658 | .128 | .087 | .160 | .070 | .389 |
| | (.030) | (.009) | (.012) | (.014) | (.013) | (.038) |
| Soybeans | .632 | .137 | .094 | .174 | .074 | .422 |
| | (.022) | (.008) | (.009) | (.013) | (.012) | (.029) |
| Corn | .459 | .199 | .144 | .316 | .040 | .672 |
| | (.017) | (.005) | (.009) | (.017) | (.020) | (.026) |
| Dry beans | .530 | .181 | .109 | .278 | .016 | .573 |
| | (.025) | (.008) | (.012) | (.024) | (.029) | (.036) |
| Sunflowers | .613 | .147 | .093 | .197 | .056 | .451 |
| | (.031) | (.008) | (.016) | (.015) | (.023) | (.040) |
| Barley | .757 | .095 | .053 | .113 | .041 | .267 |
| | (.031) | (.009) | (.015) | (.012) | (.017) | (.036) |
| Durum wheat | .721 | .115 | .048 | .147 | .019 | .313 |
| | (.027) | (.009) | (.011) | (.015) | (.014) | (.033) |
| Other small grains | .519 | .183 | .114 | .284 | .020 | .588 |
| | (.029) | (.010) | (.011) | (.031) | (.027) | (.044) |
| Canola | .602 | .143 | .112 | .180 | .100 | .460 |
| | (.029) | (.011) | (.010) | (.019) | (.016) | (.039) |
| Beets | .366 | .225 | .184 | .408 | .002 | .818 |
| | (.017) | (.005) | (.008) | (.033) | (.043) | (.029) |
| Other selected crops | .706 | .112 | .070 | .136 | .056 | .329 |
| | (.043) | (.015) | (.016) | (.022) | (.018) | (.054) |
| Potatoes | .325 | .234 | .204 | .474 | 0 | .882 |
| | (.009) | (.005) | (.005) | (.084) | (.146) | (.023) |

All standard errors are clustered by county.

TABLE 5.—ROBUSTNESS CHECK OF ESTIMATES OF COMPOSITE STRUCTURAL PARAMETER $\Theta$

| | Baseline | By County | Only on Prime Farmland |
|---|---|---|---|
| Crop | Model $\Theta$ | Average $\Theta$ | $\Theta$ |
| Spring wheat | .389 | .447 | .615 |
| | (.038) | (.117)* | (.049) |
| Soybeans | .422 | .433 | .607 |
| | (.029) | (.147)* | (.036) |
| Corn | .672 | .527 | .825 |
| | (.026) | (.248)* | (.027) |
| Dry beans | .573 | .567 | .503 |
| | (.036) | (.237)* | (.021) |
| Sunflowers | .451 | .551 | .712 |
| | (.040) | (.121)* | (.071) |
| Barley | .267 | .536 | .624 |
| | (.036) | (.244)* | (.071) |
| Durum wheat | .313 | .543 | .856 |
| | (.033) | (.250)* | (.114) |
| Other small grains | .588 | .574 | .515 |
| | (.044) | (.229)* | (.076) |
| Canola | .460 | .664 | .951 |
| | (.039) | (.243)* | (.040) |
| Beets | .818 | .759 | .762 |
| | (.029) | (.178)* | (.060) |
| Other selected crops | .329 | .588 | .804 |
| | (.054) | (.266)* | (.059) |
| Potatoes | .882 | .730 | .906 |
| | (.023) | (.148)* | (.080) |
| $N$ | 208,220 | | 111,596 |

Standard errors are calculated from county-level $\Theta$ estimates: *Significant at the 10% level.

$\bar{y}_2$, $\bar{y}_3$, and $\bar{y}_4$, so that we have four regression observations for each quarter section. From equation (18), it is clear that the error terms across the four fields within a quarter section are correlated. To account for these correlations and conditional heteroskedasticity, we use heteroskedastic robust standard errors that cluster within counties. We also tried clustering standard errors at different levels—townships, sections, and quarter sections—and they all generate lower standard errors.

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, equation (21), the policy parameters sum to 1 ($\gamma_0 + 2\gamma_A + \gamma_B = 1$). If there were no density economies, then the own quality coefficient $\gamma_O$ would equal 1. It is apparent in the table that $\gamma_O$ is substantially less than 1 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 on 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 useful summary statistic. The baseline specification imposes that the structural parameters, including the soil coefficients, are constant across all counties. Our first robustness check is to estimate the model separately for each of the 12 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

TABLE 6.—NATURAL ADVANTAGE VERSUS DENSITY ECONOMIES DECOMPOSITION

| Crop | Within Specialization Measure | Natural Advantage Share | Density Economy Share |
|---|---|---|---|
| Spring wheat | .901 | .747 | .253 |
| Soybeans | .903 | .708 | .292 |
| Corn | .969 | .658 | .342 |
| Dry beans | .948 | .694 | .306 |
| Sunflowers | .907 | .683 | .317 |
| Barley | .813 | .737 | .263 |
| Durum wheat | .855 | .753 | .247 |
| Other small grains | .944 | .641 | .359 |
| Canola | .931 | .746 | .254 |
| Beets | .994 | .784 | .216 |
| Other selected crops | .878 | .767 | .233 |
| Potatoes | .996 | .680 | .320 |

restrict attention to quarter sections where all four fields are 100% prime farmland, we eliminate 46% 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 .615 for wheat, .422 to .607 for soybeans, and .672 to .825 for corn.

### B. *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 0 if there were no density economies and soil quality were distributed i.i.d. within each county. The measure becomes 1 if all four fields of each quarter section were planted in 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.

## VI. 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 measurement error in soil classification. Perhaps a soil scientist made a mistake in classifying one field but got the classification 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 the boundaries are relatively more likely to be those between farm operations. We expect the potential for density economies to be less when adjacent fields are managed by different operations.

Section VI A introduces additional data on farm operations. Section VI B presents simple descriptive evidence to make our point. And in Section VI C, we reestimate our model with some of the additional data brought in.

### A. *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 that 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 5a, a field has two additional adjacent neighbors in different quarter sections. Call these the $C$ neighbors of a field.

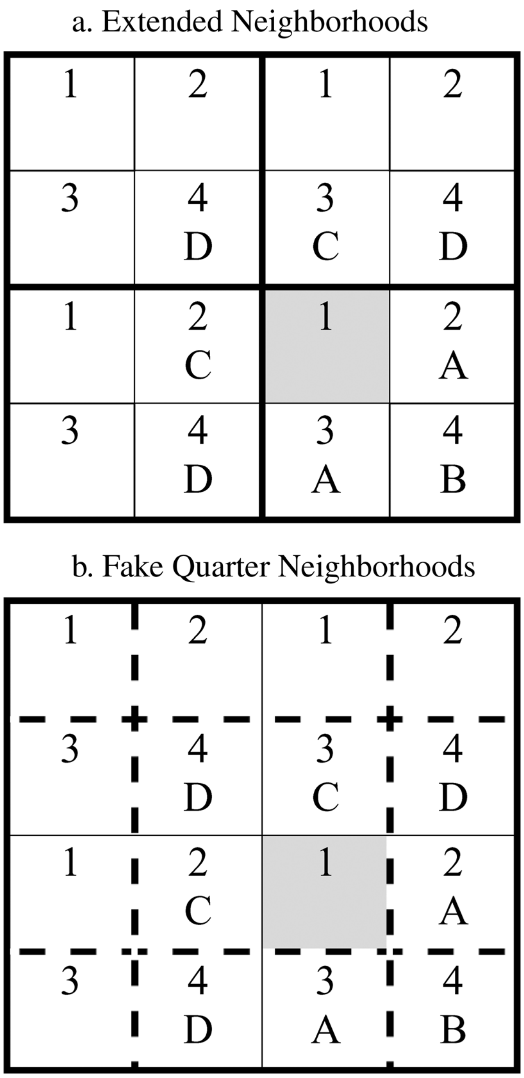
For one of the counties of our study, Cass County, we have obtained a file containing all land parcels in the county and the name of the owner of each parcel.[14] We take each point in our grid of 30 meters by 30 meters and map it to our parcel information. We then aggregate up to the field level.[15]

In table 7, we examine differences in ownership at field boundaries. For this analysis, we exclude fields categorized as urban from the soil file information (approximately 2% of the observations) and fields where one of the owner names

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

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

FIGURE 5.—ILLUSTRATION OF ADJACENT FIELD NEIGHBORS




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

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 |

Data are from the Cass County property parcel map and county-level common land unit maps.

is blank at the field boundaries (slightly more than 1% 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. Table 7 also shows the fraction of cases where owner name is identical for the adjacent fields.[16] At *A* borders, the match rate goes from .87 for parcels to .92 for owner name. At *C* borders, the match rate goes from .01 to .29. So we see that ownership commonly crosses quarter section boundaries. Nevertheless, in well more than half of the cases, adjacent fields that cross quarter section boundaries are held by owners with different names.

Even when ownership names differ across quarter section boundaries, the fields may be operated as part of the same operation. Fields held within the same family can be listed under different names (a wife's or grandmother's maiden name, for example). And farmers can operate land that they lease. The U.S. 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 were fortunate to get the data when we did.) The data are geospatial information on common land units (CLUs), the reporting units for government subsidy programs. CLUs are typically quarter sections, though there is much variation. The public release of the CLU data did not disclose the individual operators. Nevertheless, the data were published in such a way that we were able to manipulate them to determine which county office administers the federal farm programs for each field.[17] Typically a field is administered by the office in the same county as the field. But there are cases where farm operations cross county boundaries, and in such cases it is typically convenient for the farmer to work with a single administrative office. In such a case, a field can be administered by an office in a different county from where the field is located.

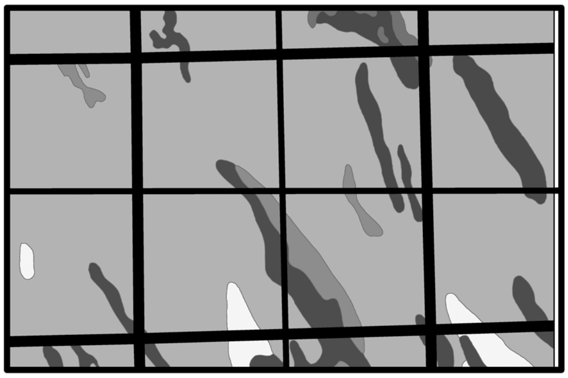
The bottom part of table 7 shows our results for the county administrator variable in Cass County. Adjacent fields being administered by different counties is relatively rare. Out of about 54,000 adjacent field pairs, this happens only 1,002 times. When this does happen, it is 10 times as likely to occur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

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

[17] The county-level CLU data contain all the CLUs in the county plus CLUs outside the county that are administered by the county.

FIGURE 6.—SOIL AND CROP MAPS

a. Example Soil Map in a Quarter Section



b. Example Crop Map in a Quarter Section

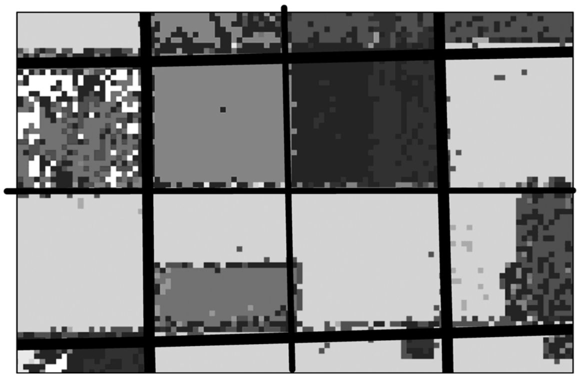


TABLE 8.—VARIATION IN SOIL AND CROP PLANTINGS ACROSS ACTUAL QUARTERS AND FAKE QUARTERS

| | $x$ Variation | | $y$ Variation | |
|---|---|---|---|---|
| Crop | Actual | Fake | Actual | Fake |
| Spring wheat | .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 |

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 the legal parcel information just for Cass County. Below, we use both variables.

### B. *Evidence of Planting Discontinuities at Quarter Section Boundaries*

This section makes the point that soil quality does not change discontinuously at quarter section boundaries, but planting does. We begin with a graphical illustration. Figure 6a provides a map of soil boundaries for a particular area in our sample and an overlay of the quarter section boundaries. It illustrates that there is heterogeneity in soils within a field. Given the arbitrary way in which quarter section boundaries were drawn in the early 1800s, we expect to see no connection with soil boundaries, and no connection is evident here. Figure 6b 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 plantings for the crop.[18] The statistic reported is like a coefficient of variation. Note that the variation of soil quality within quarter sections is significant, the statistic ranging from about .12 to .24 throughout the various crops. The existence of this within quarter section soil heterogeneity is a key part of our identification strategy.

To explain the second column, we introduce the concept of a *fake quarter section*. As illustrated in figure 5b, 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 of table 8, 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 6a, 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 6b.

### C. *Extended Model Estimates*

We reestimate our earlier model in three different ways and show how taking into account boundary considerations affects 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.[19]

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

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

TABLE 9.—FURTHER ANALYSIS ESTIMATES

| | 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) | | |
|---|---|---|---|---|---|---|---|---|
| Crop | $\Theta$ | $\Theta$ | $\gamma_A$ | $\gamma_B$ | $\delta$ | $\gamma_A$ | $\gamma_B$ | $\delta$ |
| Spring wheat | .389 | .318 | .129 | .085 | .592 | .192 | .089 | .835 |
| | (.038) | (.038) | (.009) | (.012) | (.103) | (.013) | (.021) | (.065) |
| Soybeans | .422 | .368 | .137 | .094 | .703 | .157 | .053 | .773 |
| | (.029) | (.022) | (.008) | (.009) | (.106) | (.014) | (.023) | (.070) |
| Corn | .672 | .558 | .199 | .144 | .509 | .197 | .147 | .606 |
| | (.026) | (.040) | (.005) | (.009) | (.175) | (.013) | (.016) | (.183) |
| Dry beans | .573 | .472 | .182 | .108 | .711 | .246 | .160 | .933 |
| | (.036) | (.048) | (.008) | (.013) | (.139) | (.011) | (.020) | (.181) |
| Sunflowers | .451 | .311 | .148 | .093 | .761 | .234 | .132 | .782 |
| | (.040) | (.043) | (.009) | (.016) | (.164) | (.008) | (.023) | (.097) |
| Barley | .267 | .188 | .097 | .050 | .354 | .241 | .266 | .824 |
| | (.036) | (.037) | (.010) | (.013) | (.203) | (.018) | (.036) | (.067) |
| Durum wheat | .313 | .244 | .118 | .045 | .000 | .184 | .255 | .919 |
| | (.033) | (.051) | (.010) | (.010) | (.172) | (.017) | (.037) | (.093) |
| Other small grains | .588 | .449 | .183 | .114 | 1.326 | .145 | .013 | 1.642 |
| | (.044) | (.041) | (.011) | (.011) | (.145) | (.023) | (.025) | (.338) |
| Beets | .818 | .759 | .228 | .181 | .781 | .229 | .182 | .874 |
| | (.029) | (.048) | (.006) | (.007) | (.281) | (.008) | (.019) | (.129) |
| Potatoes | .882 | .862 | .235 | .205 | 1.091 | .225 | .096 | 1.164 |
| | (.023) | (.024) | (.005) | (.005) | (.047) | (.027) | (.054) | (.219) |
| *N* | 208,220 | 201,596 | | 208,220 | | | 25,124 | |

Standard errors are clustered by townships instead of counties. We do not report estimates for canola and other selected crops because we encountered numerical difficulties.

The first exercise reestimates the model exactly as we did in table 4, except we use the fake quarters rather than the actual ones. 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, and corn .67 to .56. 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 section VII. Our main point is that a measurement error 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 in 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 a 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, and .51. These are substantially below 1.

The third exercise is analogous to the second one. But now we discount when the name is different rather than when the county administrator is different. The only data available for us to use are from Cass County. For all but the negligible crops at the bottom of the list where there are few data, the pattern of discounting is clear. For example, for spring wheat, soybeans, and corn the estimated discount factors are .84, .77, and .61. Again, these are well less than 1, but not 0. We do not expect these to come out to 0 because farm operation boundaries clearly can cross ownership name boundaries through rental markets and 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 point 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. (See the discussion of this point in section II.)

## VII. 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 the natural advantage factor in agriculture is overwhelming, we are able to get extremely detailed data at a narrow geographic level on land characteristics and choice, and in the early 1800s, the U.S. government drew an arbitrary square grid of quarter sections in the landscape, and we make heavy use of this grid.

We believe the major limitation of this paper is that it does not take into account density economies that extend beyond the quarter section level. The average farm size in North Dakota from recent Census figures is eight quarter sections.[20] Of course, planting decisions at the individual farm operation level 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 cross individual farm operations, the assumption of joint-profit maximization that we have used becomes an issue. Neighboring farmers cooperate to some extent, for example, by sharing equipment and creating agricultural cooperatives. But cooperation is unlikely to be perfect, in which case we need to bring in various game-theoretic coordination 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.

This paper presents results for a particular part of North Dakota. Some of the results use data from a single county. It is natural to ask what kind of general knowledge can be obtained from such an analysis. Given the broad regularities we found across a variety of crops, we conjecture that if we were to run a similar exercise for other agricultural regions in the United States, we would likely get qualitatively similar results. We have no reason to believe North Dakota is unique in its ability to enjoy density economies in agriculture.

But the broader knowledge we gain is not about agricultural analysis at all. Rather it is about the potential role of this kind of analysis as a stepping-stone for more sophisticated analyses of geographic concentration of general economic activity. We have taken ideas from the social interactions literature, putting them to work in an innovative way to examine the sources of geographic concentration across a four-point quadrant. Generalizing our approach to a larger number of geographic points is conceptually straightforward. We have modeled the level of activity $y_j$ at a point $j$, with $y_j$ being plantings in our case. In principle, $y_j$ could be other kinds of activity. For example, $y_j$ could be population at a point $j$ that is a plot of land 1 mile by 1 mile. While the analogy is clear, admittedly there is a big jump between thinking about sugar beet plantings on neighboring 40-acre fields in North Dakota and the population levels of neighboring square miles in Manhattan. One obvious difference is that use of a social planner problem is less natural for the latter case. This is likely something that can be finessed. More daunting is finding variables for a population analysis that play the role of soil in our agricultural analysis. We hope our paper at least stimulates thinking along these lines. In short, perhaps lessons learned from hunting mice may be of service in the hunt for elephants.

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

## REFERENCES


Arzaghi, Mohammad, and J. Vernon Henderson, "Networking off Madison Avenue," *Review of Economic Studies* 75:4 (2008), 1011–1038.

Bajari, Patrick, Han Hong, John Krainer, and Denis Nekipelov, "Estimating Static Models of Strategic Interactions," *Journal of Business and Economic Statistics* 28 (2010), 469–482.

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

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

——— "Interactions-Based Models," in James J. Heckman and Edward E. Leamer, (Eds.), *Handbook of Econometrics* (Amsterdam: Elsevier, 2001b).

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

Cass County GIS.ed. Cass County Property Parcels [computer file]. http://www.casscountynd.gov/departments/gis/Download.htm (accessed August 2008).

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

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

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

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

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

Ellison, Glenn, Edward L. Glaeser, and William Kerr, "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," NBER working paper no. 13068 (April 2007).

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

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

——— "The Diffusion of Wal-Mart and Economies of Density," *Econometrica* 79 (2011), 253–302.

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

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

Rosenthal, Stuart S., and William C. Strange, "The Determinants of Agglomeration," *Journal of Urban Economics* 50:2 (2001), 191–229.

——— "Evidence on the Nature and Sources of Agglomeration Economies" (pp. 2119–2172), in J. Vernon Henderson and Jacques-François Thisse (Eds.), *Handbook of Regional and Urban Economics*, Vol. 4, *Cities and Geography* (Amsterdam: Elsevier, 2004).

U.S. Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer for North Dakota 1997–98, 1999–00, 2001–02, 2003–04; Midwestern states 2005 and 2006. DVD. http://www.nass.usda.gov/research/Cropland/cdorderform.htm.

——— Natural Resources Conservation Service, Soil Survey Staff, Soil Survey Geographic (SSURGO) Database for North Dakota [computer file]. http://soildatamart.nrcs.usda.gov (accessed 2007).