Fueling Local Water Pollution: Ethanol Refineries, Land Use, and Nitrate Runoff

Andrew Stevens

Department of Agricultural & Resource Economics
University of California – Berkeley
stevens@berkeley.edu

Selected Paper prepared for presentation at the 2015 Agricultural & Applied Economics Association and Western Agricultural Economics Association Annual Meeting, San Francisco, CA, July 26-28

Fueling Local Water Pollution: Ethanol Refineries, Land Use, and Nitrate Runoff*

Andrew Stevens†

May 27, 2015

Abstract

Ethanol production in the United States, driven by federal renewable fuel policy, has exploded over the past two decades and has prompted the construction of many ethanol refineries throughout the US Corn Belt. These refineries have introduced a new inelastic demand for corn in the areas where they were built, reducing basis for nearby farmers and effectively subsidizing local corn production. In this paper, I explore whether and to what extent the construction of new ethanol refineries has actually increased local corn acreage. I also explore some environmental effects of this acreage increase. Using a thirteen year panel of over two million field-level observations in Illinois, Indiana, Iowa, and Nebraska, I estimate a net increase of nearly 300,000 acres of corn in 2014 relative to 2002 that can be attributed to the placements of new ethanol refineries. This increase comprises approximately 0.75% of the total 2014 corn acreage within my dataset. Furthermore, this effect is separate from the general equilibrium effect of ethanol policy increasing aggregate demand for corn. Back-of-the-envelope calculations suggest that over 21,000 tons of the nitrogen applied to fields in my sample in 2014 can be attributed to refinery location effects. Essentially all of these observed effects occur only in areas within 30 miles of an ethanol refinery, suggesting that refineries have meaningful localized impacts on land use and environmental quality such as nitrate runoff. JEL codes: Q15, Q16, Q53

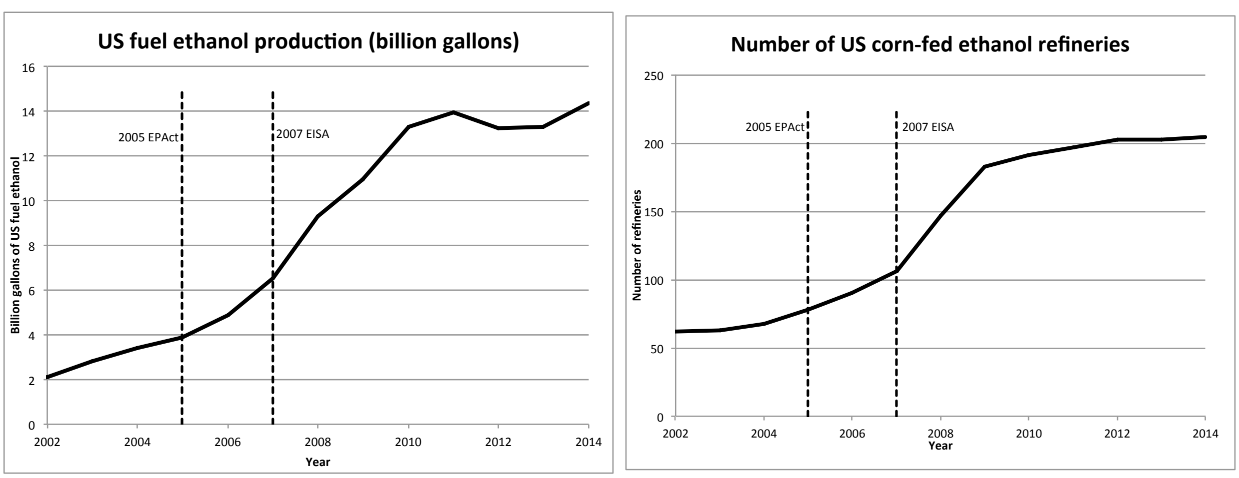
*This project has benefited greatly from comments by Peter Berck, Kendon Bell, and Keith Knapp. I also thank Nathan Hendricks for his help in acquiring data for this paper. Any remaining errors are my own.

†Department of Agricultural & Resource Economics, University of California – Berkeley; 207 Giannini Hall; Berkeley, CA 94720-3310. E-mail: stevens@berkeley.edu.

1 Introduction

Since the early 2000s, US ethanol production has exploded in response to federal policies incentivizing the production of renewable fuels. In 2005, Congress passed the Energy Policy Act (EPAct) introducing a Renewable Fuel Standard (RFS) mandating that 2.78% of gasoline sold in the US be from renewable sources. In 2007, Congress passed the Energy Independence and Security Act (EISA) setting annual renewable fuel mandates for US production with an ultimate goal of 36 billion gallons by 2022. Of these 36 billion gallons, 15 billion are to be conventional biofuels – corn-based ethanol in particular.

The US ethanol industry has clearly responded to the Renewable Fuel Standards established in the EPAct and EISA. Between 2002 and 2014, US ethanol production has increased from just over 2 billion gallons per year to over 14 billion gallons per year (Figure 1a). In order to produce such quantities of ethanol, the number of corn ethanol refineries in the US has increased from 62 in 2002 to 204 in 2014 (Figure 1b).



(a) US Ethanol Production

(b) US Ethanol Refineries

Figure 1: Growth of US Ethanol Production and Refineries, 2002-2014. Source: Renewable Fuels Association.

The striking increase in US corn ethanol production has raised several important questions about its unintended consequences. One strand of research has explored how increased demand for ethanol has affected land use in the US corn belt as aggregate demand for corn increases (Fatal & Thurman, 2014; Miao, 2013; Feng & Babcock, 2010). Another strand of research has been more concerned about the environmental externalities of changing agricultural patterns, particularly focused on nitrate runoff and water pollution (Donner & Kucharik, 2008; Thomas *et al*, 2009).

In this project, I explore both the land use change effects and environmental effects of expanding ethanol production. In particular, I study the geospatial effect of ethanol refineries' placement on nearby land use change and use my results to estimate environmental consequences. I am specifically interested in how the *location* of ethanol refineries spatially affects agricultural land, and I do not attempt to identify the full general equilibrium effect of the 14 billion gallon US corn ethanol industry. Put another way, I study how the distribution of ethanol refineries differentially affects different agricultural areas net of the ethanol industry's aggregate effect on corn prices.

I find that within a population of almost 114 million acres of agricultural land in Illinois, Indiana, Iowa, and Nebraska, nearly 300,000 more acres of corn were grown in 2014 than in 2002 due merely to ethanol refinery location effects. This represents approximately 21,000 tons of nitrogen applied as fertilizer. Almost all the 300,000 acres of increased corn acreage exist within 30 miles of an ethanol refinery, suggesting that these refineries have strong local effects on land use change and nitrogen use.

There is clear economic intuition for why ethanol refineries would differentially affect nearby and faraway agricultural land. When a corn-fed ethanol refinery is built, it represents a new terminal market for corn. Since refineries operate continuously, they have an inelastic demand for this input. And since transportation costs are significant for grains, one would expect an ethanol refinery to source its corn from the nearest producers. Thus, by reducing transportation costs for nearby producers (reducing basis), ethanol refineries essentially subsidize corn production for nearby farmers. On the margin, this subsidy incentivizes farmers to grow more corn – or grow corn more often – than they otherwise would. As corn production increases, so will nitrogen fertilizer use. Corn requires higher levels of nitrogen fertilizer than other Corn Belt crops, and particularly high levels of fertilizer when grown successively corn-after-corn. Thus, economic intuition suggests ethanol refineries would have a localized effect increasing corn production and nitrogen fertilizer use. Consequently, these refineries would also have an effect on localized nitrate runoff due to the increased nitrogen fertilizer use.

Researchers have previously addressed different components of the ethanol industry's effects on land use change and nitrate runoff. One line of research has explored whether the hypothesized local corn subsidy provided by nearby ethanol refineries actually exists. In a frequently cited paper, McNew & Griffith (2005) find that corn prices at an ethanol refinery are 12.5¢ higher than average, that the effect is slightly stronger for "upstream" refineries than for "downstream" refineries, and that price effects can be detected up to 68 miles from a refinery. However, Katchova (2009) and O'Brien (2009) both fail to find such a subsidy. Gallagher *et al.* (2005) highlight that locally-owned and non-locally-owned refineries have different effects on corn prices: the authors find that corn prices are increased by proximity to a non-locally-owned refinery, but not by proximity to a locally-owned refinery. Finally, Lewis (2010) finds different results in different states: ethanol refineries in Michigan and Kansas affect local corn prices, but refineries in Iowa and Indiana do not.

Other authors have explored whether ethanol refineries have an effect on land use. Fatal & Thurman (2014) use county-level data to estimate the corn acreage effect of ethanol refineries. They find that a typical ethanol refinery increases corn acreage in its home county by over 500 acres and has effects that can persist for up to 300 miles. Miao (2013) also uses county-level data and finds a significant effect of ethanol refineries on corn acreage, as well as a differential effect between locally-owned and non-locally-owned refineries. Turnquist *et al.* (2008), in contrast to more recent studies, fail to find any significant agricultural land conversion in areas near Wisconsin ethanol refineries. Finally, Feng & Babcock (2010) explore the full general equilibrium effect of increased ethanol production and find an unambiguous increase in corn acreage.

Several researchers have focused on how ethanol production affects water quality and nitrate runoff. Donner & Kucharik (2008) highlight how the aggregate impact of the EISA will likely make achieving nitrate level goals in the Mississippi impossible. Thomas *et al.* (2009) use hydrologic models to estimate the water quality impacts of corn production caused by increased demand due to biofuel mandates. They find significant negative results.

While it is likely true that "refineries cause corn," it is also likely true that "corn causes refineries." Ethanol refineries are not located at random, and several researchers have explored the topic of ethanol refinery placement. A series of papers have shown, unsurprisingly, that ethanol refineries are more likely to locate near areas with large corn production, near transportation infrastructure, and not near existing ethanol refineries (Sarmiento *et al.*, 2012; Haddad *et al.*, 2010; Lambert *et al.*, 2008). This finding is important because it highlights that ethanol refinery placement cannot be treated and random in econometric analyses.

My project improves upon previous work by leveraging new sources of field-level land use data and exploiting a finer-scaled panel of observations than previous authors. I exploit both the Cropland Data Layer (CDL) and Common Land Unit (CLU) to create annual observations of field-level land use. These agricultural micro-data allow for much more nuanced econometric estimation than in previous studies. Other authors have exploited similar micro-data in agricultural research to great effect (Livingston *et al.*, 2015; Hendricks *et al.*, 2014; Wright & Wimberly, 2013). I also highlight the locality effect of ethanol refineries rather than the general equilibrium effect, focusing on small-scale heterogeneous effects that have not been well identified in previous work.

The remainder of this paper is divided into a theoretical framework (model), a summary of my data, an overview of my econometric methods, a discussion of my results, and a conclusion.

2 Model

Consider a farmer maximizing expected profits from an agricultural field. I assume the farmer is not forward-looking, and maximizes only the current year's expected profits.[1] The farmer observes input prices, expected output prices, the locations of terminal markets for possible crops, and the field's planting history. Then, the farmer chooses to plant either corn (C), soy (S), or another crop (O) to maximize:

$$\begin{aligned}\mathbb{E}\left[\Pi_i\right] &= \max\left\{\mathbb{E}\left[\Pi_i(C)\right], \mathbb{E}\left[\Pi_i(S)\right], \mathbb{E}\left[\Pi_i(O)\right]\right\} \\ &= \max_{x\in\{C,S,O\}}\left\{\mathbb{E}\left[\left(p_x - b_{i,x}\right) f_x\left(\mathbf{v}_x|x_{i,-1}\right) - \mathbf{z}_x \cdot \mathbf{v}_x\right]\right\} \qquad (1)\end{aligned}$$

where Π_i are profits for field i, p_x is the output price for crop x, $b_{i,x}$ is the basis for crop x on field i, f_x is the production function for crop x, $\mathbf{v}_x$ is a vector of inputs to produce crop x, $x_{i,-1}$ is the crop planted on field i in the previous period, and $\mathbf{z}_x$ is a vector of input prices for inputs $\mathbf{v}_x$. Basis $b_{i,x}$ is assumed to be a linear function of distance from field i to the nearest terminal market for crop x, and input quantities $\mathbf{v}_x$ are determined my maximizing Π_i conditional on x and $x_{i,-1}$.

Given the problem outlined above, a farmer's optimal decision is deterministic. However, for an econometrician who does not observe all relevant data and production functions, the above problem gives rise to a probability that field i will be planted crop x given previous planting decisions: $P_i(x|x_{i,-1})$. Summing across different possible planting histories, this gives rise to an unconditional probability that field i will be planted to crop x: $P_i(x)$. In this project, I am interested in how ethanol refineries affect the probability a field will be planted to corn: $P_i(C)$.

[1] Of course, we expect farmers to be forward-looking and dynamically optimizing their cropping decisions. Livingston *et al.* (2015) provide an excellent treatment of how these dynamics affect farmers' optimal choices. These authors find relatively little difference in the optimal behavior of a myopic farmer compared to that of an infinitely-forward-looking farmer, suggesting my own model is an acceptable approximation of reality. However, one could easily incorporate Livingston *et al.*'s (2015) dynamics into a more complicated version of the model I present here.

Now consider the introduction of a new corn-fed ethanol refinery. There are two interesting cases. First, suppose the new refinery is closer to field i than any existing refineries, but further away than field i's current terminal market for corn. In this case, field i's distance-to-nearest-refinery changes, but its distance-to-nearest-corn-market remains unchanged and its basis for corn, $b_{i,C}$ remains the same as before. In the second case, suppose the new refinery is closer to field i than field i's current terminal market for corn. In this case, field i's basis for corn, $b_{i,C}$ gets smaller while its bases for soy and other crops, $b_{i,S}$ and$b_{i,O}$, remain unchanged.[2]

Within the second case outlined above, the specific placement of the ethanol refinery matters. Since I assume basis is linear in distance to terminal market, $b_{i,C}$ will increase with distance to the new ethanol refinery as long as the refinery is closer to field i than the next-nearest terminal market.

These assumptions give rise to two predictions about the effect of corn-fed ethanol refineries on the probability corn is planted on any field i. First, regardless of planting history, if the new refinery is further away than the current terminal market for corn, there will be no impact on field i's probability of being planted to corn $P_i(C)$. Second, regardless of planting history, if the new refinery is closer than the current terminal market for corn, the new refinery will increase $P_i(C)$ linearly in distance. Figure 2 gives a graphical representation of these two predictions.

Unfortunately, in reality, I only observe a field's distance to its nearest ethanol refinery. I do not observe its distance to all nearest terminal markets for corn or any other crop. If I were able to observe locations for all grain elevators or other terminal markets, I might be able to more explicitly test for both predictions summarized in Figure 2. As it is, I can only hope to observe an effect that has the general shape outlined in Figure 2.

An important point to note is that the effect of distance to the nearest ethanol refinery on

[2]This model makes the implicit assumption that a field's nearest terminal market for any crop x is able to absorb the product of all fields for whom it is the nearest terminal market. In the case of ethanol refineries, this may not be true. Rather, refinery production capacities may impose additional constraints or relax other constraints. I leave this issue to future work.

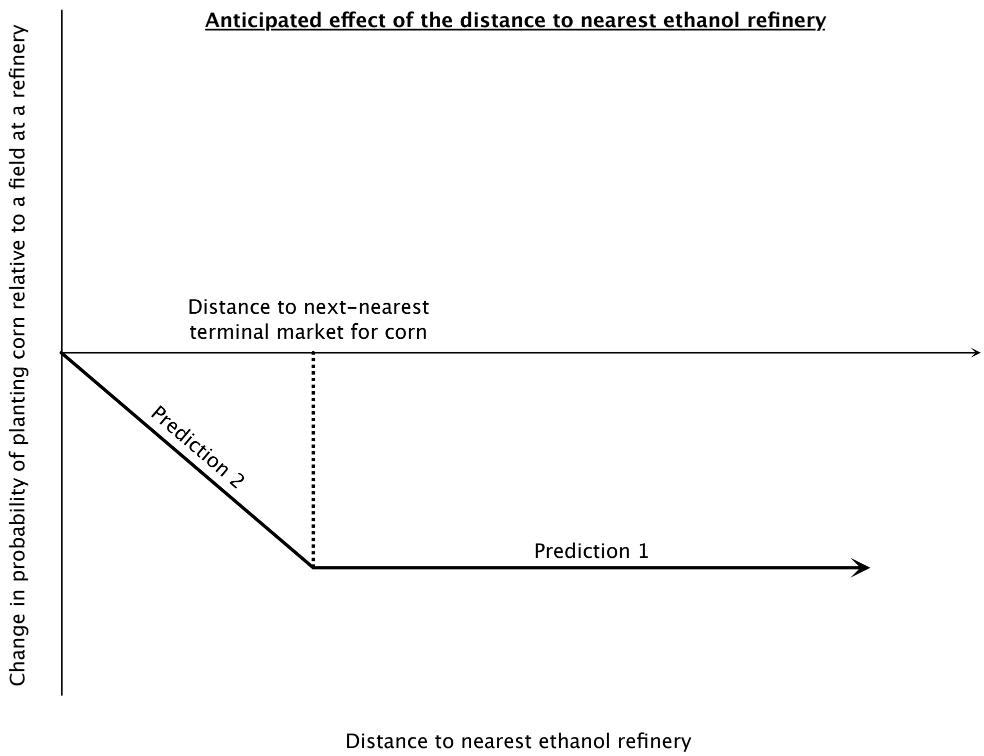


Figure 2: Model prediction of the effect of distance to nearest ethanol refinery on probability a field is planted to corn.

the probability of planting corn is not strictly linear; it is piecewise-linear. This suggests that running any regression with simply a linear "distance to nearest ethanol refinery" covariate on the right-hand-side will obfuscate the true underlying relationship. A non-linear specification is required.

3 Data

To conduct my analysis, I construct a balanced panel of annual crop choices for 2,145,848 agricultural fields in Illinois, Indiana, Iowa, and Nebraska over thirteen years from 2002 to 2014. In each year, I also calculate the distance from each field to the nearest ethanol refinery. I rely on three data sources to create my panel: the Cropland Data Layer, Common Land Unit, and ethanol refinery locations.

3.1 Cropland Data Layer

The Cropland Data Layer (CDL) is a raster dataset of landcover in the United States collected and maintained by the National Agricultural Statistics Service (NASS) of the USDA. A satellite records the electro-magnetic wavelengths of light reflected from different points on the earth's surface and uses a ground-tested algorithm to assign each pixel a single land-cover type for the year. Pixels measure 30 meters by 30 meters, except for years 2006-2009 when pixels measured 56 meters by 56 meters.[3] The CDL provides remarkably high-resolution land cover data and is able to distinguish between many different types of vegetation. Figure 3 displays the CDL for Illinois, Indiana, Iowa, and Nebraska in 2014. Yellow pixels represent corn, dark green pixels represent soy, and light green pixels represent other grassland-like land covers. Red dots mark ethanol refineries operating in 2014.

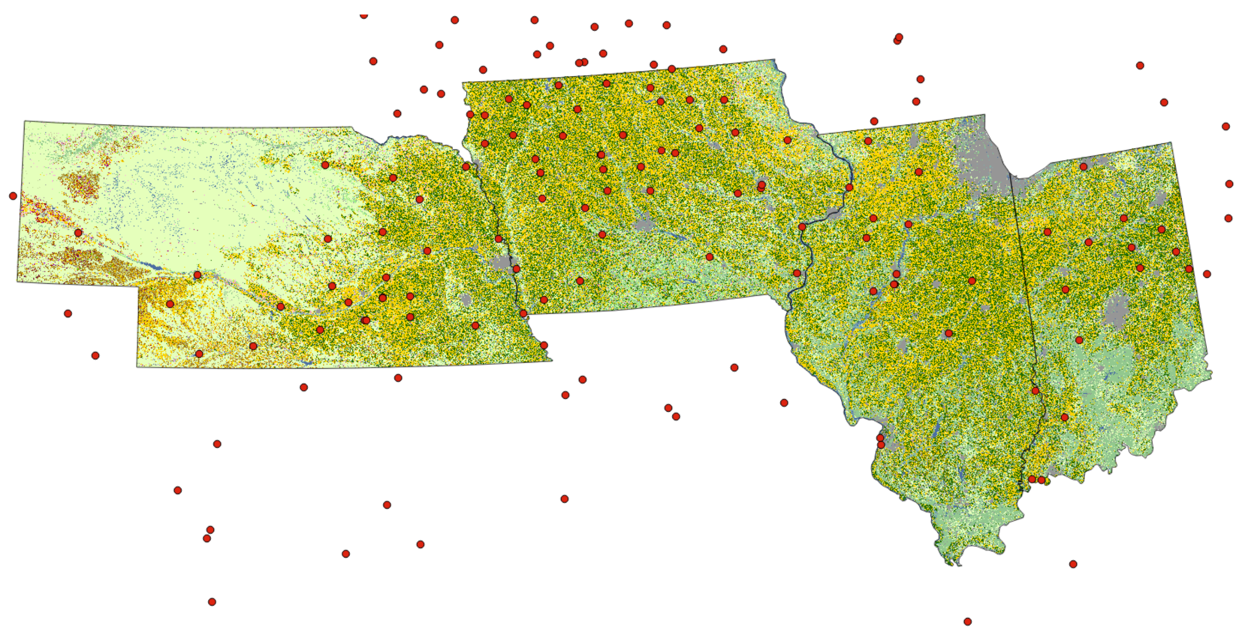


Figure 3: 2014 Cropland Data Layer (CDL) with ethanol refinery locations. Sources: NASS & Renewable Fuels Association.

The CDL identifies different crops with different accuracies. For major row crops, CDL

[3]Data collection for the CDL began in the late 1990s in only three states. By 2008, all 48 contiguous states had been included in the CDL. Changes in CDL technical specifications – such as different pixel sizes in different years – can be attributed to a growing data collection program with evolving hardware and software resources.

accuracy is usually between 85% and 95% (Boryan *et al.*, 2011). For corn and soy, pixel accuracy is particularly high. For example, according to the 2014 CDL metadata, accuracy for both corn and soy in Iowa was over 97%. However, the CDL is considerably less accurate at distinguishing between more similar land covers, such as between alfalfa, rangeland, and grassland. This fact has complicated research that explores extensive land use change in the Western Corn Belt (Wright & Wimberly, 2013), but is not a concern for research focused on corn and soy.

One problem with using raw CDL data is that a 30 meter by 30 meter pixel is likely not the appropriate unit of analysis. Rather, economists are more interested in observing field-level crop choices. Additionally, while CDL data are quite accurate for primary row crops, it is apparent that individual pixels are frequently mis-measured. For instance, upon visual inspection of a CDL image, it is not uncommon to observe what is clearly a large field of more than 100 pixels planted to soybeans, with one or two pixels somewhere in the field reported as corn. If analysis is conducted at the pixel level rather than the field level, such mis-measurements become a large concern. To address this concern, I exploit Common Land Unit data to construct field-level crop cover observations.

3.2 Common Land Unit

According to the Farm Service Agency (FSA) of the USDA, a Common Land Unit (CLU) is "an individual contiguous farming parcel, which is the smallest unit of land that has a permanent, contiguous boundary, common land cover and land management, a common owner, and/or a common producer association" (FSA, 2012). Practically, a CLU represents a single agricultural field. Polygon shapefiles of CLUs are maintained by the FSA, but are not currently publicly available.

I obtain CLU data for Illinois, Indiana, Iowa, and Nebraska from the website GeoCommunity (`http://www.geocomm.com`). These data contain shapefiles from the mid 2000s, before CLU data were removed from the public domain. In this research, I implicitly assume that individual CLUs do not change over time: a reasonable assumption given the FSA definition. In reality, the FSA does adjust individual CLU definitions on a case-by-case basis if necessary, but I assume these adjustments to be negligible as in previous similar studies (Hendricks *et al.*, 2014).

Using the geospatial software ArcGIS, I overlay the CDL raster data with CLU polygons as shown in Figure 4. Upon visual inspection, the fit is quite good: CLU boundaries line up with crop changes in the CDL, roads appear clearly in both datasets, and geographical features such as waterways and elevation changes are visible. One concern is that many CLUs are quite small and appear to outline geographical features such as gullies, rather than larger constituent fields. This is particularly pronounced in areas near urban sprawl. Therefore, to maintain confidence that the fields I study are actually "fields" in the way we think of them, I drop all CLUs from my dataset with areas of less than 10 acres. I also drop CLUs with areas of greater than 10,000 acres, based on an assumption that these CLUs are incorrect.[4]

To assign each CLU a single crop cover, I calculate the modal value of the raster pixels contained within each CLU polygon. I then assign that modal value to the entire CLU. This procedure enforces the assumption that each field (CLU) is planted to a single crop – an assumption strongly supported by a visual examination of the data. To my knowledge, this is the first instance of using modal statistics to interact the CDL and CLU datasets. Previous research (Hendricks *et al.*, 2014) has used an off-center centroid to sample a single point of the underlying raster data. My procedure is preferable in that it reduces the chance of idiosyncratic mis-measurement of the field's true land cover.

Finally, for each CLU polygon, I construct a centroid for the field. This centroid is constrained to exist within the boundaries of its parent CLU polygon. I then use these CLU centroids to calculate distances from each field to its nearest ethanol refinery in each year.

[4] I drop 3,201,933 fields that have areas under 10 acres. The aggregate area dropped is 10,043,985 acres. I drop 108 fields that have areas over 10,000 acres. The aggregate area dropped is 1,965,324 acres. The total acreage dropped is less than 10% of the 125,987,632 acres in my initial dataset.

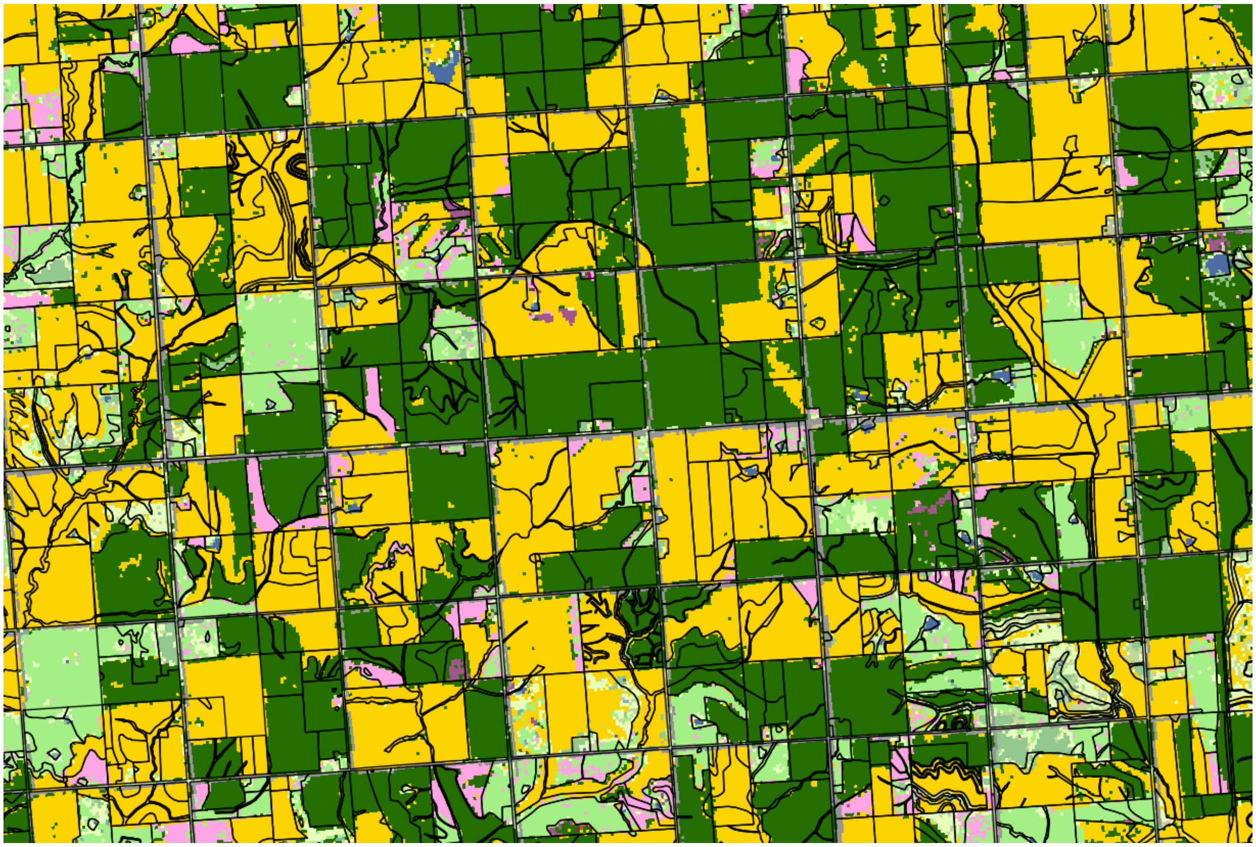


Figure 4: Detail of Cropland Data Layer (CDL) and Common Land Unit (CLU) data: Iowa, 2014. Yellow pixels represent corn, dark green pixels represent soy, light green pixels represent grassland, and pink pixels represent alfalfa. Black lines are CLU borders. Sources: NASS & FSA.

3.3 Ethanol Refineries

I obtain data on ethanol refinery location, capacity, and opening date from the Renewable Fuels Association (RFA). The RFA has comprehensive data on ethanol refineries each year starting in 2002.[5] Using these data, I geo-code the locations of over 200 ethanol refineries in the US. Since new ethanol refineries open each year, I create a separate dataset of operating ethanol refineries for each year from 2002-2014. I only include refineries that can use corn as an input and omit refineries that only accept cellulosic biomass or non-corn inputs. Figure 5 displays the geographic expansion of corn-fed ethanol refineries between 2002 and 2014.

[5]Data on ethanol refineries is unavailable for 2013. In my analysis, I assume all "new" ethanol refineries in 2014 opened in 2014, even though some of them may have opened in 2013.

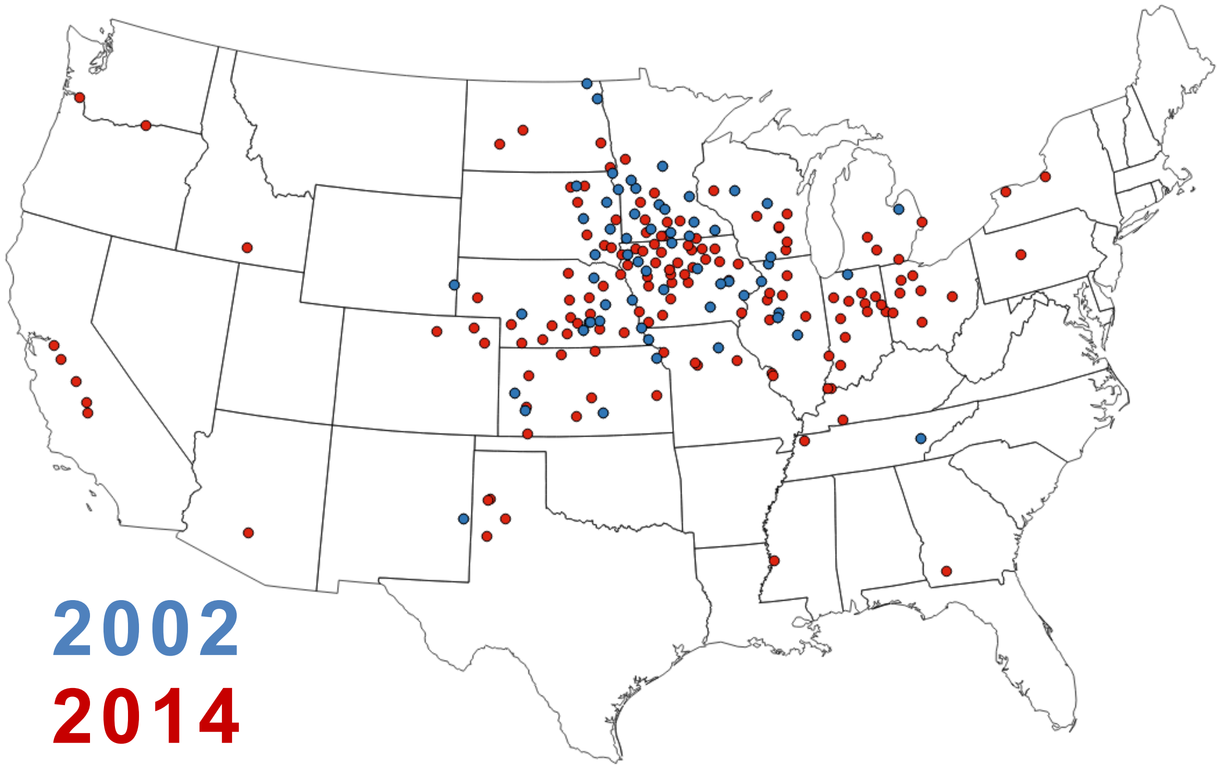


Figure 5: Ethanol refinery locations in 2002 and 2014. Source: Renewable Fuels Association.

Using ArcGIS, I calculate the distance from each CLU centroid to the nearest ethanol refinery for each year from 2002 to 2014. As new ethanol refineries are constructed, this distance will decrease for nearby fields. This change in distance-to-nearest-refinery is the variation I will use to identify my econometric analysis. Figure 6 displays the change in distribution of nearest-distances from 2002 to 2014. Distributions for years 2003-2013 are omitted for clarity, but the distribution skews more and more to the left in each year.

In the current project, I do not incorporate ethanol refinery production capacity into my analysis. Rather, I treat each refinery as identical. Thus, there is no analytic difference between a field 30 miles from a 10 million-gallons-per-year (mgy) refinery and a field 30 miles from a 100 mgy refinery. I plan to exploit production capacity information in future work.

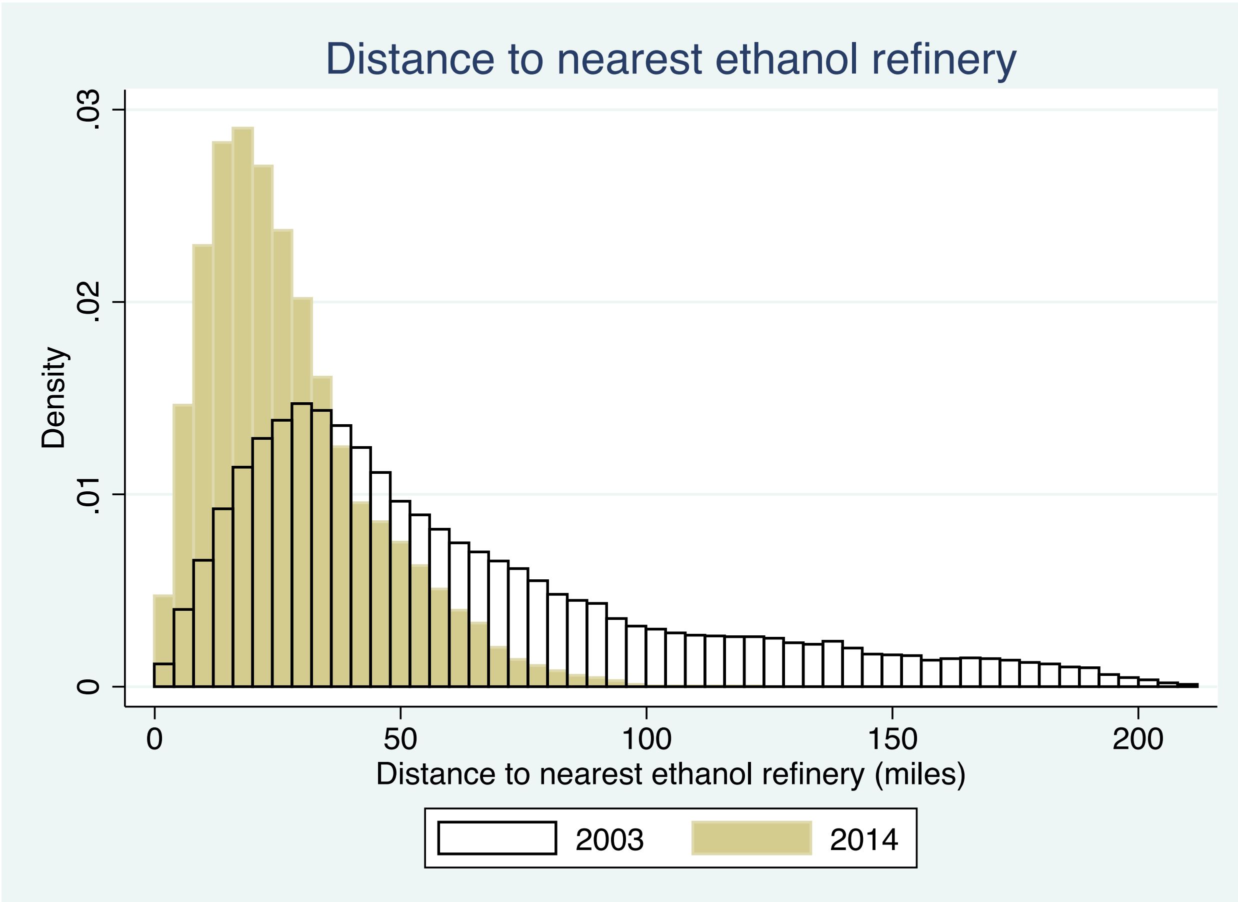


Figure 6: Distributions of field-level distance-to-nearest-refinery in 2003 and 2014. Each distribution represents 2,145,848 agricultural fields. Source: author's calculations.

3.4 Summary Statistics

The final dataset I use for my analysis consists of a balanced panel of 2,145,853 agricultural fields over 13 years. Table 1 presents summary statistics for the data. The average field is 53.12 acres in area, is 62.14 miles away from the nearest ethanol refinery in 2002, and 27.52 miles away from the nearest refinery in 2014.

In any particular year, an average of 32.85% of all fields are planted to corn representing 33.27% of all acreage. Similarly, 25.28% of fields are planted to soy representing 23.51% of all acreage. Remaining fields and acreage are left to other crops.

Table 1: Summary Statistics

VARIABLE	Mean	Median	Standard Dev.	Minimum	Maximum	Observations
Field area (acres)	53.12	30.15	123.60	10	9,934.94	2,145,853
Fields of corn	704,992	699,640	54,088	627,912	785,925	13
% of fields corn	32.85	32.60	2.52	29.26	36.63	13
Acreage of corn	37,915,466	38,217,232	2,819,787	33,887,520	41,720,688	13
% of acreage corn	33.27	33.53	2.47	29.73	36.60	13
Fields of soy	542,552	545,443	39,075	434,344	593,006	13
% of fields soy	25.28	25.42	1.82	20.24	27.63	13
Acreage of soy	26,792,631	27,205,006	1,809,211	21,987,108	29,220,192	13
% of acreage soy	23.51	23.87	1.59	19.29	25.64	13
Fields of other crops	898,309	901,122	58,856	816,503	981,631	13
% of fields other crops	41.86	41.99	2.74	38.05	45.75	13
Acreage of other crops	49,270,226	49,726,160	2,207,439	46,356,128	52,840,092	13
% of acreage other crops	43.23	43.63	1.94	40.67	46.36	13
Dist. to nearest refinery, 2002	62.14	48.95	44.05	0.01	211.84	2,145,853
Dist. to nearest refinery, 2003	61.60	47.96	44.19	0.01	211.84	2,145,853
Dist. to nearest refinery, 2004	61.58	47.96	44.19	0.01	211.84	2,145,853
Dist. to nearest refinery, 2005	49.38	41.66	30.99	0.01	157.32	2,145,853
Dist. to nearest refinery, 2006	48.55	40.89	31.40	0.01	157.32	2,145,853
Dist. to nearest refinery, 2007	46.77	38.44	31.19	0.01	157.32	2,145,853
Dist. to nearest refinery, 2008	34.15	28.36	22.86	0.01	131.05	2,145,853
Dist. to nearest refinery, 2009	30.07	25.57	19.54	0.01	128.41	2,145,853
Dist. to nearest refinery, 2010	29.44	24.81	19.48	0.01	128.41	2,145,853
Dist. to nearest refinery, 2011	28.91	24.59	18.87	0.01	126.30	2,145,853
Dist. to nearest refinery, 2012	27.52	23.75	17.07	0.01	126.30	2,145,853
Dist. to nearest refinery, 2013*	27.52	23.75	17.07	0.01	126.30	2,145,853
Dist. to nearest refinery, 2014**	27.52	23.75	17.07	0.01	126.30	2,145,853

Notes: Total number of fields: 2,145,853. Total acreage: 113,978,323 acres. Variables with 2,145,853 observations are measured at the field level. Variables with 13 observations are measured at the year level. All distances measured in miles. *The Renewable Fuels Association did not publish data on ethanol refineries in 2013, so there is no change in distance in my data between 2012 and 2013. **Although the mean and median distances for 2014 appear identical to those for 2013, this is only due to rounding. The values for 2014 are in fact smaller.

4 Econometric Methods

Using a balanced panel of 2,145,848 agricultural fields over 13 years, I estimate a linear probability model (LPM) of the probability that an individual field is planted to corn. I include as independent variables a field-level fixed effect to capture time-invariant unobserved characteristics of individual fields, state-by-year fixed effects to capture input and output prices, and distance-bin dummy variables measuring distance to the nearest ethanol refinery. These distance-bin variables allow me to observe a non-linear relationship between a field's distance to its nearest ethanol refinery and the probability that field is planted to corn. The distance-bins are the covariates of interest. In particular, I estimate equation 2:

$$P_{it}(C) = \beta_0 + \beta_1 \text{bin10}_{it} + ... + \beta_{21} \text{bin210}_{it} + \gamma_{st} + \alpha_i + \varepsilon_{it} \tag{2}$$

where i indexes field, t indexes year, $P_{it}(C)$ is the probability of field i growing corn in year t, $\text{bin10}, ..., \text{bin210}$ are sequential distance bin dummies beginning at the 10 mile mark and each representing a range of 10 miles (the omitted bin in 0-10 miles), γ_{st} is a state-by-year fixed effect, α_i is a field-level fixed effect, and ε_{it} is an error term. I cluster standard errors at the field level to control for heteroskedasticity and correlation over time.

Equation 2 has several desirable qualities. First, and most importantly, an LPM allows me to control for field-level fixed effects by exploiting the within-transformation. This allows me to ignore any characteristics of a field that do not change over time, such as soil quality, average weather patterns, and field slope. Second, an LPM allows me to easily interpret the coefficients $\beta_1, ...\beta_{21}$ as marginal effects. For instance, a field 25 miles away from its nearest ethanol plant is $100 \times \beta_2$ percent more likely to grow corn than a field 0-10 miles away from its nearest ethanol plant, *ceteris paribus*.

Linear probability models have drawbacks as well. LPMs can result in coefficients that will predict outcomes outside the [0,1] interval, as opposed to discrete-choice models such as logit or probit. The reason I use an LPM rather than a discrete-choice model is that fixed-effects are difficult to impossible to incorporate in such a framework. Additionally, with large sample sizes, LPMs often perform quite similarly to discrete choice models.

In addition to equation 2, I estimate similar equations conditioning on a field's land use during the previous year. In particular, I estimate $P_{it}\left(C|C_{-1}\right)$ (the probability of growing corn given that corn was grown last year), $P_{it}\left(C|S_{-1}\right)$ (the probability of growing corn given that soy was grown last year), and $P_{it}\left(C|O_{-1}\right)$ (the probability of growing corn given something other than corn or soy was grown last year). These specifications allow me to explore how ethanol refineries affect specific crop rotation dynamics rather than merely an aggregate effect.

It is important to note that one cannot interpret the coefficients from equation 2, $(\beta_1, ..., \beta_{21})$ as purely causal. This is because ethanol refinery placement is non-random: refineries locate in particular places due to the presence of corn supply, access to transportation infrastructure, and distance from other ethanol refineries, among other factors (Sarmiento *et al.*, 2012; Haddad *et al.*, 2010; Lambert *et al.*, 2008). Instead, one should interpret the coefficients as the differential effect of distance to the nearest ethanol plant conditional on some unobserved characteristics driving ethanol refinery placement. In this context, these coefficients have meaningful and valid interpretations for constructing regional counterfactuals about corn acreage changes.

5 Results

I estimate equation 2 using the `reghdfe` command in `stata`. This command optimizes the estimation of high-dimensional fixed effects models and runs considerably faster than `xtreg`. However, the `reghdfe` command subsumes the constant term β_0 which must be reconstructed after the regression has been estimated. Thus, the constants reported in Table 2 do not include standard errors.

Table 2 presents my results for four different econometric specifications. In all cases, the dependent variable is the probability of a field being planted to corn. In specification (1), this probability in unconditional. In specifications (2), (3), and (4), the probability is conditional upon corn, soy, or neither (respectively) being grown on the field in the previous year.

Table 2: Probability of a field being planted to corn

VARIABLES	(1) $P_{it}(C)$	(2) $P_{it}(C\|C_{-1})$	(3) $P_{it}(C\|S_{-1})$	(4) $P_{it}(C\|O_{-1})$
Constant (recovered)	0.3346	0.3249	0.7679	0.0825
Distance bin: 10-20 miles	-0.0004	-0.0047***	0.0026**	0.0004
	(0.0006)	(0.0011)	(0.0011)	(0.0008)
Distance bin: 20-30 miles	-0.0001	-0.0149***	0.0103***	0.0062***
	(0.0006)	(0.0011)	(0.0012)	(0.0008)
Distance bin: 30-40 miles	-0.0077***	-0.0266***	0.0098***	0.0001
	(0.0006)	(0.0012)	(0.0012)	(0.0008)
Distance bin: 40-50 miles	-0.0088***	-0.0316***	0.0154***	-0.0021***
	(0.0006)	(0.0013)	(0.0013)	(0.0008)
Distance bin: 50-60 miles	-0.0080***	-0.0171***	0.0125***	-0.0057***
	(0.0006)	(0.0014)	(0.0014)	(0.0008)
Distance bin: 60-70 miles	-0.0048***	-0.0166***	0.0138***	-0.0013
	(0.0007)	(0.0015)	(0.0016)	(0.0008)
Distance bin: 70-80 miles	-0.0062***	-0.0166***	0.0060***	-0.0035***
	(0.0007)	(0.0016)	(0.0017)	(0.0008)
Distance bin: 80-90 miles	-0.0074***	-0.0145***	0.0060***	-0.0067***
	(0.0007)	(0.0017)	(0.0018)	(0.0009)
Distance bin: 90-100 miles	-0.0049***	0.0003	-0.0010	-0.0046***
	(0.0008)	(0.0018)	(0.0020)	(0.0009)
Distance bin: 100-110 miles	-0.0098***	0.0160***	-0.031***	-0.0121***
	(0.0008)	(0.0019)	(0.0022)	(0.0010)
Distance bin: 110-120 miles	-0.0144***	0.0339***	-0.0415***	-0.0295***
	(0.0009)	(0.0021)	(0.0024)	(0.0012)
Distance bin: 120-130 miles	-0.0084***	0.0211***	-0.0254***	-0.0244***
	(0.0010)	(0.0022)	(0.0026)	(0.0013)
Distance bin: 130-140 miles	-0.0058***	0.0089***	-0.0454***	-0.0150***
	(0.0011)	(0.0026)	(0.0030)	(0.0013)
Distance bin: 140-150 miles	-0.0112***	0.0045	-0.0804***	-0.0193***
	(0.0012)	(0.0033)	(0.0036)	(0.0015)
Distance bin: 150-160 miles	-0.0113***	0.0141***	-0.1125***	-0.0031*
	(0.0014)	(0.0040)	(0.0044)	(0.0016)
Distance bin: 160-170 miles	-0.0108***	0.0246***	-0.0999***	-0.0081***
	(0.0015)	(0.0046)	(0.0049)	(0.0015)
Distance bin: 170-180 miles	-0.0077***	0.0198***	-0.1071***	-0.0129***
	(0.0015)	(0.0050)	(0.0052)	(0.0016)
Distance bin: 180-190 miles	0.0046***	0.0273***	-0.1079***	-0.0089***
	(0.0016)	(0.0053)	(0.0060)	(0.0017)
Distance bin: 190-200 miles	0.0051**	0.0368***	-0.0837***	-0.0192***
	(0.0020)	(0.0092)	(0.0073)	(0.0021)
Distance bin: 200-210 miles	0.0206***	0.0388**	-0.0630***	-0.0167***
	(0.0027)	(0.0155)	(0.0144)	(0.0028)
Distance bin: 210-220 miles	0.0193	0.0412	-0.3913***	-0.0090
	(0.0104)	(0.0896)	(0.0690)	(0.0090)
Field FE	YES	YES	YES	YES
State-by-Year FE	YES	YES	YES	YES
Observations	25,750,236	8,424,413	6,477,454	10,848,369
Number of fields	2,145,853	1,549,958	1,451,343	1,414,126
R-squared	0.3324	0.4834	0.5052	0.5021

Notes: In each specification, the dependent variable is the probability of a field being planted to corn. In specification (1), the probability is unconditional. In specifications (2), (3), and (4), the probability is conditional upon corn, soy, or neither (respectively) being grown on the field in the previous year. Distance bins are dummy variables for 10-mile ranges of distance to the nearest ethanol refinery. The omitted bin is 0-10 miles, so the constant term represents the probability of a field 0-10 miles from the nearest ethanol plant being planted to corn. Coefficients on the distance bins are interpreted as marginal effects relative to the constant term. Standard errors clustered at the field level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

The unconditional probability of a field 0-10 miles from its nearest ethanol refinery being planted to corn is 33.46%. This is only slightly higher than the 32.49% probability of being planted to corn after being planted to corn last year. However, the same probability after being planted to soy is more than twice as large at 76.79%. Corn is relatively unlikely to be grown after crops other than corn or soy, with a probability of only 8.25%.

It is easiest to interpret the coefficients on each of the distance bins by plotting them on a graph. Figures 7, 8, 9, and 10 correspond to specifications (1), (2), (3), and (4), respectively. In each case, vertical distance on the graph measures changes in the probability that a field is planted to corn relative to fields 0-10 miles from their nearest ethanol refineries. Also, recall that by 2014, the mean distance to a field's nearest ethanol plant is 27.52 miles, the median distance is 23.75 miles, and the maximum distance is 126.30 miles. I highlight these facts to focus readers' attention on the areas of the following graphs most relevant to the underlying population of fields.

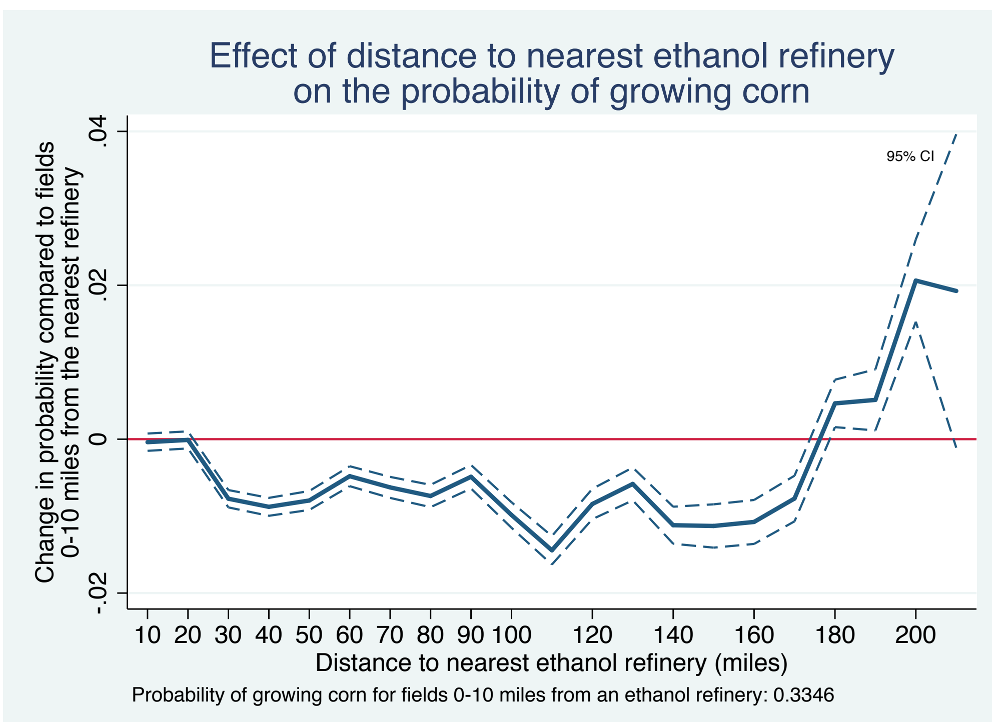


Figure 7: Unconditional probability of growing corn: specification (1).

Consider Figure 7. The first result to note is an overall shape between the 20 and 100 mile markers that looks remarkably similar to the shape of the piecewise-linear curve in Figure 2. If we rely heavily on our theoretical framework, Figure 7 may suggest the average field is approximately 30-40 miles away from its nearest non-ethanol-refinery terminal market for corn. Unlike my model's prediction, however, Figure 7 displays no statistically significant effects of distance in the 10 and 20 mile bins. Broadly, specification (1) suggests that fields 0-10 miles from their nearest ethanol refineries are approximately 1% more likely to grow corn in any given year than fields 40-170 miles away from their nearest refinery.

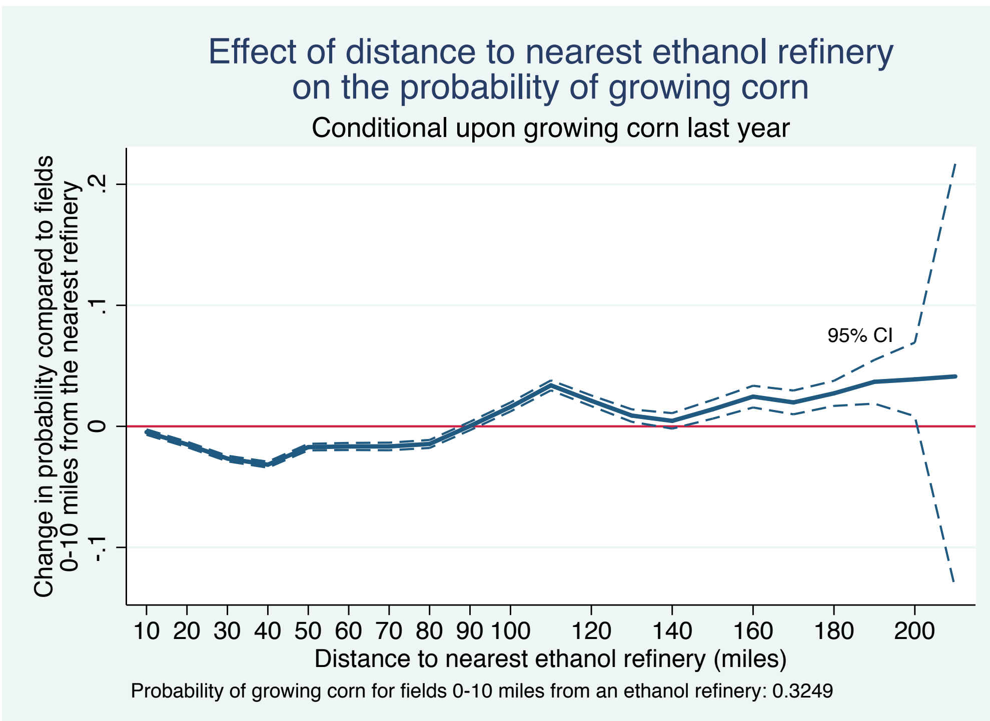


Figure 8: Probability of growing corn conditional on growing corn in the prior year: specification (2).

Figure 8 presents results conditioning on corn being planted in the previous year (specification (2)). Note that the scale of the vertical axis in Figure 8 is larger than in Figure 7. In this case, we see a stronger and more pronounced effect of distance to nearest ethanol refinery on the probability of planting corn in the region 10-40 miles away from a refinery. Here we more clearly see the sloped portion of the curve predicted in Figure 2. The interpretation of this regression is that as ethanol refineries are constructed, they strongly incentivize nearby fields to grow corn-after-corn relative to fields further away. This behavior particularly exacerbates any negative externalities of nitrogen fertilizer use.

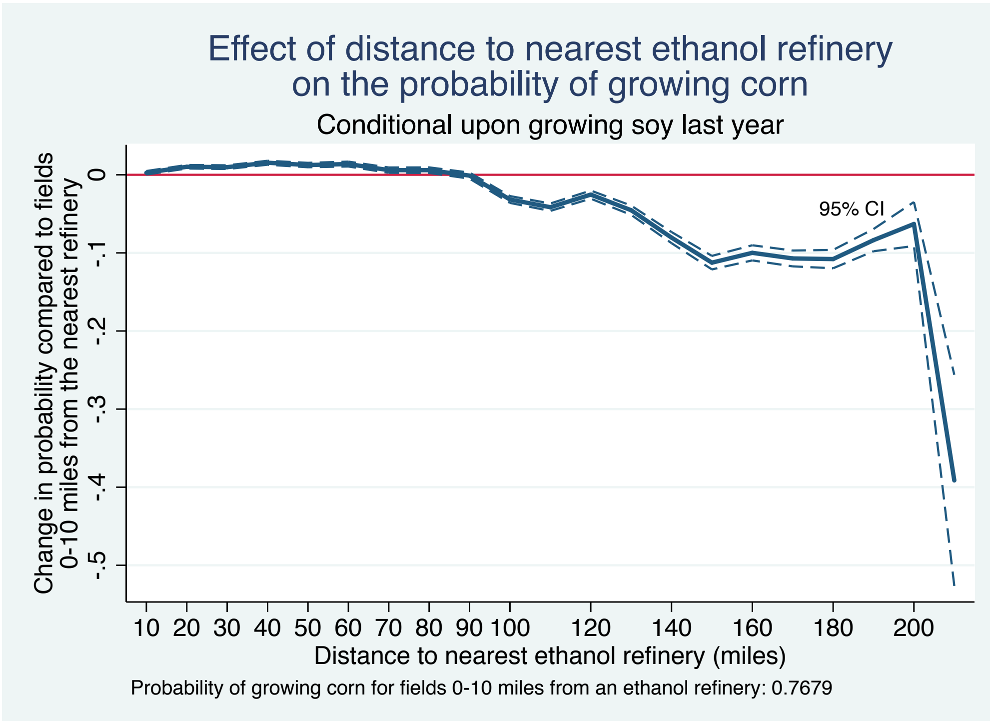


Figure 9: Probability of growing corn conditional on growing soy in the prior year: specification (3).

Figure 9 presents results conditioning on soy being planted in the previous year. This figure displays a puzzling relationship: within the range of 10-70 miles, fields close to ethanol refineries are *less* likely to grow corn-after-soy than are fields further away. This result is contrary to the prediction developed in Figure 2, and has no obvious explanation.

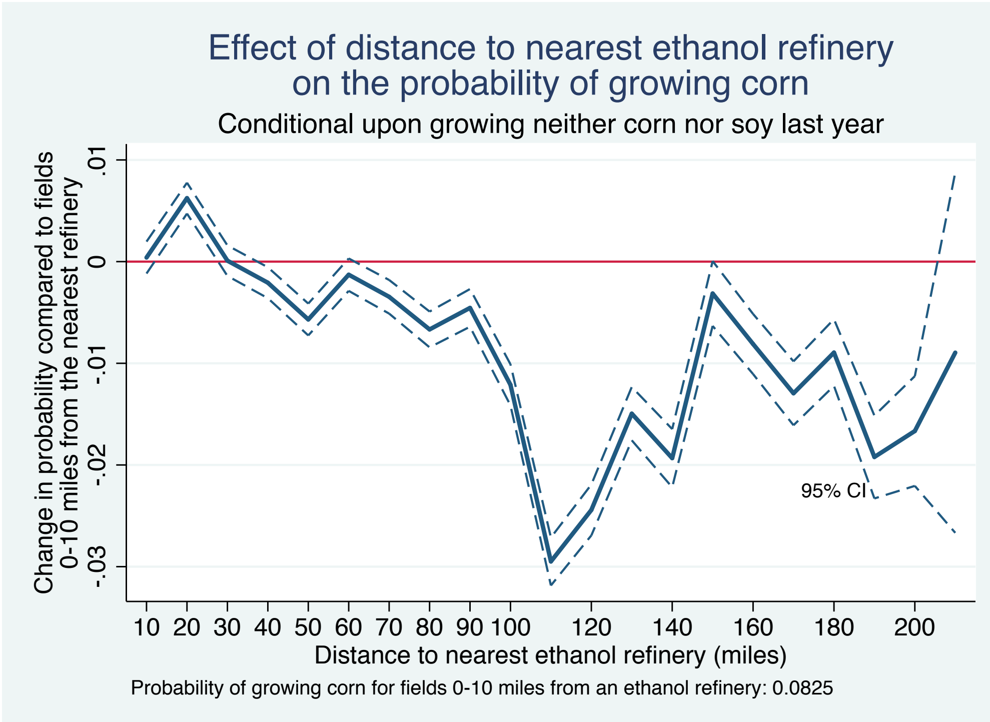


Figure 10: Probability of growing corn conditional on growing something other than corn or soy in the prior year: specification (4).

Figure 10 presents results conditioning on something other than corn or soy being planted in the previous year. While this relationship is much more volatile than those in Figures 8 and 9, there remains a strong downward trend between 20 and 110 miles from the nearest ethanol refinery.

Taken together, and relying primarily on the unconditional results shown in Figure 7, I conclude that the distance to a field's nearest ethanol plant does affect that field's probability of being planted to corn in a non-linear way. If anything, results from Figure 8 suggest that farmers are realizing this effect in a way that maximizes strain on crop rotations and that exacerbates the negative externalities associated with nitrogen fertilizer use.

Next, I use the results from specification (1) to determine how the entry of ethanol refineries between 2002 and 2014 affected corn acreage in my population of fields. For each of the 2,145,853 fields in my dataset, I use their distance to the nearest ethanol refinery in 2002 and the relevant coefficient from specification (1) to approximate the unconditional probability of that field being planted to corn in 2002. I then repeat this process using each field's distance to nearest ethanol refinery in 2014. Subtracting the former probability from the latter, I construct the change in unconditional probability between 2002 and 2014 of each field growing corn in any particular year. Note that this change in probability is entirely attributable to the distance-to-nearest-refinery effect, and does not depend on level-shifts in corn demand between 2002 and 2014. I then multiply each field's change in probability by its acreage and sum across all fields to find the total net change in acreage between 2002 and 2014. I find a total increase of 298,718 acres in my population of 113,978,323 acres. Figure 11 presents this change in acreage divided into distance-bins of 10 miles each.

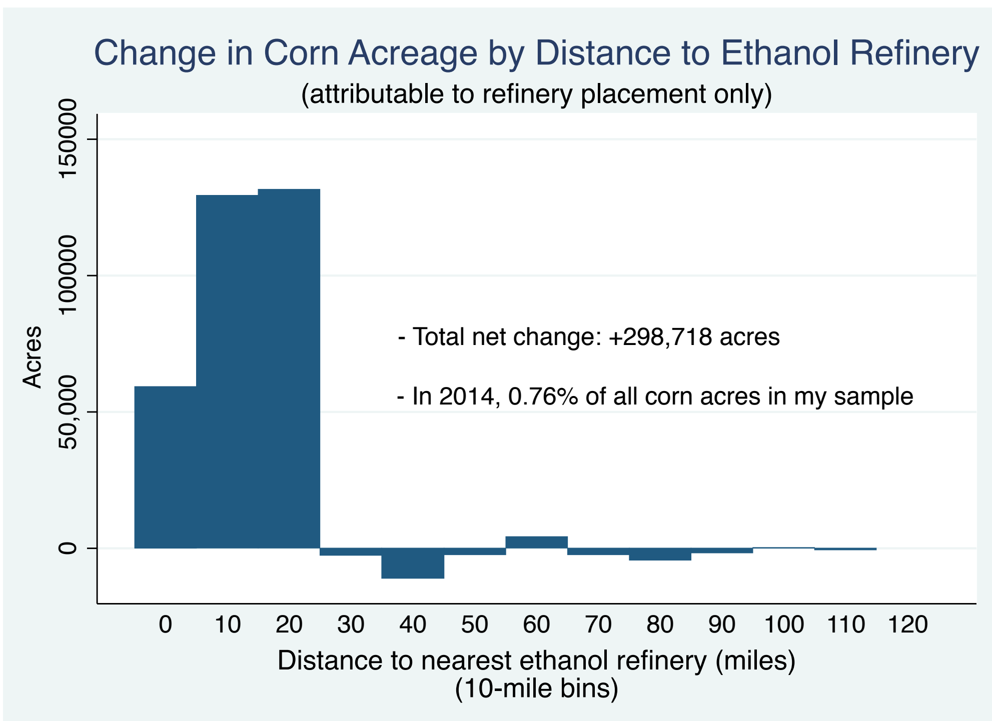


Figure 11: Effect of distance to nearest ethanol refinery on corn acreage. Derived from coefficients estimated in specification (1).

The net increase in corn acreage of 298,718 that I find is only 0.26% of the 113,978,323 acres in my population, but it is 0.76% of all corn acreage in my population in 2014. This is a significant number given that it can be attributed to only the distance-to-nearest-refinery effect. In other words, the effect of new ethanol refineries since 2002 on lowering transportation costs (and *not* the general-equilibrium effect of ethanol increasing aggregate demand for corn) can explain almost 300,000 acres of the corn grown in a subset of the fields across Illinois, Indiana, Iowa, and Nebraska.

Figure 11 highlights that the entirety of this acreage effect is captured by fields less than 30 miles from the nearest ethanol refinery. This result matches incredibly well with the predictions outlined in Figure 2. It also demonstrates that any spatial externalities associated with increased corn cultivation due to ethanol refinery location occur entirely within 30 miles of ethanol refineries. This suggests highly localized effects.

What do the acreage increases highlighted in Figure 11 mean for nitrogen application? A 2007 Iowa State University Extension publication suggests that optimal nitrogen application for corn-after-soy is 125 lb N/acre, and optimal application for corn-after-corn is 175 lb N/acre (Sawyer, 2007). Taking a middle value of 140 lb N/acre, (recall that most corn is grown after soy), the 298,718 acres of increased corn acreage estimated in Figure 11 represent 41,820,520 lbs, or almost 21,000 tons of extra nitrogen.

These 21,000 tons of additional nitrogen that are attributable to the distance effect of ethanol refinery placement are essentially all applied to areas within 30 miles of an ethanol refinery. While this number is relatively small relative to the total application of nitrogen in the US Corn Belt, there is cause for concern about localized geographic effects. Nitrate runoff into local water sources is harmful to water ecosystems, animals, and humans, and has been a growing problem in the US Corn Belt (Donner & Kucharik, 2008; Mueller & Helsel, 1996). Local water quality data from the USGS could be used in future research to look for an effect of ethanol refineries on nitrate levels directly.

6 Conclusion

In this paper, I have demonstrated that ethanol refineries exert a statistically significant effect on the land use of surrounding fields. Increases in corn acreage and nitrogen application occur within 30 miles of ethanol refineries, suggesting a highly localized effect. These findings are consistent with a model of ethanol refineries lowering corn basis for nearby farmers. Within a sample of almost 114 million acres, I find nearly 300,000 acres of the corn grown in 2014 can be attributed to ethanol placement effects accumulated over the years between 2002 and 2014.

This project makes several important contributions to the existing literature and improves upon previous research. Most importantly, I leverage field-level observations of land use to create a thirteen year panel of over two million observations. This allows me to estimate a highly nonlinear relationship between distance to a field's nearest ethanol refinery and that field's probability of growing corn. My panel also allows me to include field-level fixed effects that control for time-invariant characteristics of each field such as soil type.

In three econometric specifications that condition on the previous year's land use, I find interesting patterns. Ethanol refineries seem to strongly incentivize nearby fields to grow corn-after-corn, while the effect appears opposite for corn-after-soy. The result for corn-after-soy is puzzling and has not been explained by theory. Future work may attempt to better understand this result. Nonetheless, the net effect of these two individual effects is that farmers appear to be growing more corn near ethanol refineries in the way the most stresses crop rotations and most exacerbates the use of nitrate-producing fertilizer.

There is considerable room for further work on these questions. First, the results of this project must be carefully interpreted as the locations of ethanol refineries are not themselves random. Second, this analysis treats all refineries as identical. In reality, different refineries have different production capacities and may have heterogeneous effects on surrounding land. Third, there is room to explore a wider range of econometric specifications beyond the linear probability model estimated in this paper. Finally, future work should explore data from the US Geological Survey to test whether water nitrate levels directly reflect the effect derived in the current project.

While the findings of this paper appear relatively small in the context of the entire US

Corn Belt, they are strongly statistically significant and demonstrate a real and important localized effect of ethanol refinery placement. My results are useful for anyone interested in a fuller understanding of the spatial forces driving land use change and nitrogen application in agriculture.

References

Boryan, C., Yang, Z., Mueller, R., and Craig, M. (2011). Monitoring US agriculture: the US department of agriculture, national agricultural statistics service, cropland data layer program. *Geocarto International*, 26(5):341–358.

Donner, S. D. and Kucharik, C. J. (2008). Corn-based ethanol production compromises goal of reducing nitrogen export by the Mississippi river. *Proceedings of the National Academy of Sciences*, 105(11):4513–4518.

Fatal, Y. S. and Thurman, W. N. (2014). The response of corn acreage to ethanol plant siting. *Journal of Agricultural and Applied Economics*, 46(02).

Feng, H. and Babcock, B. A. (2010). Impacts of ethanol on planted acreage in market equilibrium. *American Journal of Agricultural Economics*, 92(3):789–802.

FSA (2012). Common Land Unit (CLU): Information sheet. Technical report, USDA.

Gallagher, P., Wisner, R., and Brubacker, H. (2005). Price relationships in processors' input market areas: Testing theories for corn prices near ethanol plants. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*, 53(2-3):117–139.

Haddad, M. A., Taylor, G., and Owusu, F. (2010). Locational choices of the ethanol industry in the midwest corn belt. *Economic Development Quarterly*, 24(1):74–86.

Hendricks, N. P., Smith, A., and Sumner, D. A. (2014). Crop supply dynamics and the illusion of partial adjustment. *American Journal of Agricultural Economics*, 96(5):1469–1491.

Katchova, A. L. (2009). The spatial effect of ethanol biorefinery locations on local corn prices. In *Proceedings of the Agricultural and Applied Economics Association AAEA and ACCI Joint Annual Meeting*, Milwaukee, WI.

Lambert, D., Wilcox, M., English, A., and Stewart, L. (2008). Ethanol plant location determinants and county comparative advantage. *Journal of Agricultural and Applied Economics*, 40:117–135.

Lewis, K. E. (2010). The impact of ethanol on corn market relationships and corn price basis levels. Masters thesis, Michigan State University.

Livingston, M., Roberts, M. J., and Zhang, Y. (2015). Optimal sequential plantings of corn and soybeans under price uncertainty. *American Journal of Agricultural Economics*, 97(3):855–878.

McNew, K. and Griffith, D. (2005). Measuring the impact of ethanol plants on local grain prices. *Applied Economic Perspectives and Policy*, 27(2):164–180.

Miao, R. (2013). Impact of ethanol plants on local land use change. *Agricultural and Resource Economics Review*, 42(2).

Mueller, D. K. and Helsel, D. R. (1996). Nutrients in the nation's waters–too much of a good thing? Technical Report 1136, US Geological Survey.

O'Brien, D. M. (2009). The effects of the micro-market structure for kansas grain elevators on spatial grain price differentials. In *Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*, St. Louis, MO.

Sarmiento, C., Wilson, W. W., and Dahl, B. (2012). Spatial competition and ethanol plant location decisions. *Agribusiness*, 28(3):260–273.

Sawyer, J. (2007). Nitrogen fertilization for continuous and rotated corn. Technical Report AEP 0200f, Iowa State University Extension.

Thomas, M. A., Engel, B. A., and Chaubey, I. (2009). Water quality impacts of corn production to meet biofuel demands. *Journal of Environmental Engineering*, 135(11):1123–1135.

Turnquist, A., Fortenbery, T. R., and Foltz, J. (2008). Progress or devastation? The effects of ethanol plant location on local land use. In *Proceedings of the American Agricultural Economics Association Annual Meeting*, Orlando, FL.

Wright, C. K. and Wimberly, M. C. (2013). Recent land use change in the western corn belt threatens grasslands and wetlands. *Proceedings of the National Academy of Sciences*, 110(10):4134–4139.