Corn Ethanol Expansion and the Evolution of US Crop Patterns

T. Jake Smith and GianCarlo Moschini

Working Paper 23-WP 656
December 2023

Center for Agricultural and Rural Development
Iowa State University
Ames, Iowa 50011-1070
www.card.iastate.edu

T. Jake Smith is Postdoctoral Research Fellow, Center for Agricultural and Rural Development, Iowa State University, Ames, Iowa, 50011. E-mail: jpsmith@iastate.edu.

GianCarlo Moschini is Charles F. Curtiss Distinguished Professor and Pioneer Endowed Chair in Science and Technology Policy, Department of Economics, Iowa State University, Ames, Iowa, 50011. E-mail: moschini@iastate.edu.

For questions or comments about the contents of this paper, please contact GianCarlo Moschini, moschini@iastate.edu.

Corn Ethanol Expansion and the Evolution of U.S. Crop Patterns

T. Jake Smith and GianCarlo Moschini *

Abstract. The 2007 expansion of the U.S. Renewable Fuel Standard (the so-called RFS2) has had significant, and controversial, impacts on U.S. agriculture. We combine rich spatial data on observed crop choices with two metrics of corn demand by the ethanol industry to estimate the impact of the RFS2 on crop patterns. Crop choices are observed at the field level using two recent extensions of the Cropland Data Layer database. The time path of the RFS2 mandate captures its corn demand expansion at the aggregate level, and the heterogeneous impact of this biofuel policy across space is measured by a novel local ethanol demand index computed at the field level. This framework permits us to characterize both the local and overall impacts of the RFS2. We find that the ceteris paribus local presence of RFS2-induced ethanol plants, for the 13 Midwest states in the analysis, increased total corn area by 1.6%, increased the area of corn following corn by 4.3%, and had a small, spatially heterogeneous impact on soybeans. Of more consequence is the overall impact of the RFS2, which is estimated to have increased total corn area by 9.7% and corn following corn by 16.8%. Interestingly, total soybean area was also significantly impacted by the RFS2, increasing by 6.7%, and corn-soybean rotations increased by about 10%. These results indicate a major impact of the RFS2 on the geography of corn and soybean cultivation—intensification in the heart of the Corn Belt, and significant expansion near the periphery of the Corn Belt—which included the conversion of 13.9% of land previously used for other crops and/or conservation to corn/soybean production.

Keywords: Biofuels, Crop choice, Ethanol, Land use changes, Markov chains, Renewable Fuel Standard, Rotations

This draft: 19 December 2023

* T. Jake Smith is a Postdoctoral Research Fellow with Center for Agricultural and Rural Development, Iowa State University. GianCarlo Moschini is C.F. Curtiss Distinguished Professor and the Pioneer Chair in Science & Technology Policy, Department of Economics, Iowa State University.

1. Introduction

The expansion of biofuel production is one of the most significant developments affecting agricultural markets over the last two decades. In the United States, the surge of the biofuels industry was largely driven by policies that promoted the use of increasing amounts of biofuels in the U.S. transportation fuel supply. These "mandates," known as the Renewable Fuel Standard (RFS), originated with the Energy Policy Act of 2005, but were significantly expanded by the Energy Independence and Security Act (EISA) of 2007 (Schnepf and Yacobucci 2013). The RFS includes annual requirements for the total use of biofuels, as well as the utilization of specific advanced biofuels such as biodiesel and cellulosic biofuel. Though there is no statutory requirement specifically for conventional (that is, non-advanced) biofuel, in practice the vast majority of the total biofuel mandate—the portion in excess of advanced biofuel mandates—is met with the most common conventional biofuel: domestically-produced ethanol distilled from corn starch.

U.S. biofuel policies have had major impacts. Ethanol production added a sizeable new source of demand for corn, necessarily putting upward pressure on commodity prices (de Gorter, Drabik and Just 2015; Carter, Rausser, and Smith 2017). The implicit subsidy effect of biofuel policies clearly benefits the agricultural production sector (de Gorter and Just 2010), and, by improving the United States' terms of trade in commodity export markets, may increase U.S. aggregate surplus as well (Cui et al., 2011; Moschini, Lapan and Kim 2017). For a fuller welfare assessment of these policies, of course, one must consider the various market failures that matter in this context. An important set of issues concerns the environmental implications of the RFS. Indeed, the objective of reducing greenhouse gas (GHG) emissions was one of the rationalizations proffered in support of biofuel policies—because of their renewable nature, biofuels can reduce carbon emission by cutting the use of fossil fuels (Farrell et al. 2006).

The extent to which biofuels can contribute to reducing GHG emissions remains controversial. Lark et al. (2022) provide an ambitious assessment of the environmental impacts of the RFS, with a focus of the resulting land use changes. Their analysis shows that farmers' responses to the higher prices caused by biofuel policies include increased planting of corn, reduction of

diversity in crop rotations, intensification of production practices and greater use of nitrogen fertilizer, and expansion of cultivated cropland. The net environmental outcomes are sobering: the RFS does not meet its own GHG emission objectives, and biofuels are unlikely to provide net environmental benefits. Such negative conclusions have been challenged by Taheripur et al. (2022), who take issue with several modeling features of the analysis, including the way changes in cropland use are constructed from separately estimated price impacts attributed to the RFS. The mediating role of prices, in the model, was also questioned by Falconi et al. (2022).

In this paper, we provide a novel appraisal of a key element of this controversy: namely, the ceteris paribus impact of the RFS2-induced ethanol production expansion on planted crop patterns. Previous work has inferred the impact of the RFS2 by first estimating the response of equilibrium prices to the policy-driven expansion of corn demand, as well as (and often separately) the supply response of crop choices to such price changes. Identification of the parameters underlying these adjustments is challenging. One aspect of specific interest here is that the local impacts of the biofuel policy have come about via the proliferation of new ethanol plants, the locations of which have disrupted historical price basis patterns. Whereas representative national commodity prices likely to inform and influence farmers' decisions (such as futures prices) are readily available, local prices suitable to accurately characterize local supply response are not. Our strategy, therefore, is to directly relate the exogenous expansion of the RFS ethanol mandate, and the ensuing proliferation of local ethanol processing capacity induced by the policy, to the observed crop choices of farmers. To model the local impacts, we develop a suitable index of ethanol plant density that accounts for the specific location of all ethanol plants, their capacity, and their initial date of operation.

Crop choices are modeled at their most elementary unit, the field level. To this end, we supplement the Cropland Data Layer (CDL) used by other studies in this area (e.g., Motamed, McPhail, and Williams 2016; Y. Wang et al. 2020; Lark et al. 2022) with the Corn-Soy Data Layer (CSDL) published by S. Wang et al. (2020). We aggregate pixels classified by the CSDL and CDL to the field level using the recently released USDA Crop Sequence Boundary (CSB) dataset (Hunt et al. 2023). Similar to previous studies that have relied on CDL data (e.g., Pates and

Hendricks 2021), our model also emphasizes the role of rotation in farmers' planting choices. We make a few additional contributions on the modeling side. First, the Markov chain transition probabilities are embedded in a discrete choice model of crop planting. All transition probabilities are therefore estimated jointly and based on the same data, which provides advantages for estimation and inference. Second, rather than reducing the choice set of farmers to a binary choice (corn and "other crops"), we separately track three distinct choices: corn, soybeans, and other crops.

Having estimated how the diffusion of ethanol plants affected crop choices, the model permits a counterfactual analysis for the impact of U.S. biofuel policies. Our focus here is on the incremental impact of the augmented mandates attributable to the 2007 expansion of the RFS (the so-called RFS2), relative to the 2005 version of the RFS. The latter mandated ethanol volumes for 2006–2012, with the volume rising from 4 billion gallons in 2006 to 7.5 billion gallons in 2012 (Schnepf and Yacobucci 2013; U.S. Environmental Protection Agency 2021). We believe these earlier mandates were essentially non-binding, because ethanol had emerged as the oxygenate of choice to comply with the requirements of the 1990 amendments to the Clean Air Act.[1] The RFS2, however, significantly increased the mandated volume of ethanol above its use as oxygenate. Under the RFS2, the mandated volume of ethanol rose steeply before plateauing at 15 billion gallons per year after 2015. The RFS2 thus effectively doubled the long-run mandated volume of ethanol from the 7.5 billion gallons per year mandated by the RFS to 15 billion gallons per year. Our counterfactual experiment thus asks the question: what if the expanded mandates of the RFS2 were not implemented?

The main research question of this paper, therefore, concerns the extent to which the large increase in corn demand driven by the RFS2 impacted crop patterns in the United States. The model we develop permits us to estimate the local effects of ethanol plants on surrounding crop

[1] The Clean Air Act had introduced a 2% oxygen requirement for gasoline, which was earlier met by using methyl tertiary butyl ether (MTBE). The phaseout of MTBE in the early 2000s, due to concerns about groundwater contamination, resulted in ethanol being the additive of choice to meet the oxygen requirement. Prior to 2011, ethanol production was also supported by a federal subsidy (a blender tax credit) (Moschini, Cui, and Lapan 2012).

patterns in addition to the overall impact of the RFS2. Unlike previous work that has focused exclusively on corn, we characterize the impact of the biofuel policy on both corn and soybean cultivation. There are two reasons why attention to both crops is warranted. First, because corn is typically grown in rotation with soybeans, expansion of corn demand has the potential to affect crop rotation practices, thereby affecting the planted areas for both crops. Second, although the most salient feature of the RFS2 concerned its implied ethanol mandate, EISA also specified biodiesel mandates. Meeting the latter has relied on feedstocks that include (among others) soybean oil, as well as imports. Several considerations, however, suggest that biodiesel production is unlikely to have generated local effects comparable to those of ethanol.[2] At the national level, explicitly accounting for biodiesel would simply reinforce the overall corn-ethanol impact of the RFS2. In this paper, therefore, our focus in on the ethanol component of the RFS2, and on the mediating roles of the spatial distribution of ethanol plants.

To estimate the local effects of ethanol plants on surrounding crop patterns, we analyze the change in crop areas relative to a counterfactual in which the RFS2 increased national corn demand but ethanol production capacity in our study area remained at pre-RFS2 levels. We find that the RFS2-induced expansion of ethanol capacity increased corn area, across the 13 states in the study region, by an average of 1.6% over the 2008–2018 period. The local impacts of ethanol plants on soybean area were spatially heterogeneous, with decreases in soybean area in Iowa and parts of Minnesota and Illinois, as fields shifted from corn-soybean rotation to more continuous corn planting, and increases in soybean area around the periphery of the Corn Belt as fields transitioned from cultivating other crops to corn-soybean rotation.

[2] Standard biodiesel production relies on a transesterification process with organic fats and oils as feedstocks. In the United States soybean oil has accounted for about one-half of the feedstocks (Brown 2020). The location of biodiesel plants tends to reflect availability of feedstocks—for soybean oil, that depends of where the crushing of soybeans takes place—and it is more scattered across the United States than for ethanol (Gerveni, Irwin, and Hubbs 2023). Recent plant developments have privileged renewable diesel, a distinct form of biodiesel that uses a hydrogenation pathway similar to that of oil refining (Brown 2020). Such plants often entail conversion of existing refineries, and their location is chosen so as to benefit from existing petroleum infrastructures.

Of more consequence is the overall impact of the RFS2. We find that the augmented mandate due to the RFS2 increased total corn area by 9.7% and corn following corn by 16.8%. Interestingly, we find that total soybean area was also significantly impacted by the RFS2, with total soybean area increasing by 6.7%, and corn-soybean rotations increasing by about 10%. Thus, our analysis uncovers a major impact of the RFS2 on the geography of corn and soybean cultivation. Throughout the U.S. Midwest corn increased along the extensive margin, with some fields that had previously produced other crops switching to corn production. At the periphery of the Corn Belt, including parts of North Dakota, South Dakota, Nebraska, and Kansas, there was a pronounced transition from other crops to corn-soybean rotation practices, increasing the total area of both corn and soybean cultivation. Meanwhile, corn area also increased along the (intensive) rotational margin in Iowa, southern Minnesota, and northern Illinois, which showed an increase in corn-corn rotation and a decrease in corn-soybean rotation, and Indiana, which showed an increase in corn-soybean rotation and a decrease in soybean-soybean rotation. The total net effect was an increase in corn area throughout the U.S. Midwest, and a relative shift of soybean cultivation from the heart of the Corn Belt toward its western periphery.

2. Data

We combine several sources of rich spatial data to study the impact of the RFS2 on crop dynamics. Together, these datasets allow us to observe the locations and capacities of U.S. ethanol plants as well as high-resolution crop choice data across the U.S. Midwest in each year from 2001 to 2018. We use the data on ethanol plant locations and capacities to construct a novel measure of the demand pressure that ethanol plants place on surrounding farms, which we use to estimate the RFS2's local impacts on crop production.

2.1 Ethanol Plant Data

We collect data on U.S. ethanol plants from the Renewable Fuel Association (RFA) and the Nebraska Department of Environment and Energy (NDEE) (Renewable Fuels Association 2022; Nebraska Department of Environment and Energy 2018). Both sources contain annual data on the capacities of all U.S. ethanol plants as of January 1 of each year, with each source covering

different years. Between the two sources we assemble data on ethanol plant capacities in each year from 2001 to 2018. We use RFA data for 2001–2004 and NDEE data for 2005–2018.

In the data, we observe an ethanol plant's nameplate capacity, operating capacity (indicating whether the plant was operating or idle during a particular year), and capacity under construction, as well as the feedstock used to produce the ethanol. We use operating capacity as the relevant measure to assess ethanol's impact on crop production. We focus our analysis on plants that use corn as a feedstock. We track 203 distinct corn ethanol plants, 178 of which are located in the Midwestern states that are the focus of this paper, over the sample period of 2001–2018. Appendix Figure A.1 shows the locations of the 203 plants we track.

The data we have assembled provide a comprehensive history of all ethanol plants opening, expanding, and closing. To assess the impact of the RFS2, a key piece of our strategy is to construct a counterfactual history for the existence, location, and capacity of ethanol plants that would have occurred absent the enactment of the RFS2. To this end, we identify as best as we can the ethanol plants that would have operated in each year had the RFS2 not been enacted. This requires paying close attention to the timing of the RFS2 and the buildout of U.S. ethanol capacity. The RFS2 was signed into law as part of EISA in December 2007. This legislation, however, was first introduced in the House of Representatives in January 2007, and its passage was clearly anticipated by market operators even before then.

The RFA reports total U.S. ethanol capacity under construction as of January 1 of each year. As shown in Figure A.2 in the Supplementary Appendix, capacity under construction rose steadily from 375 million gallons per year (Mgal/y) in 2002 to 720 Mgal/y in 2005, jumped to about 1700 Mgal/y in 2006, and spiked to an unprecedented level of over 5200 Mgal/y in 2007 (Renewable Fuels Association 2022). We (conservatively) interpret the smaller jump in construction in January 2006 as being driven by the RFS as enacted by the Energy Policy Act in August 2005 (the RFS1). However, we believe the magnitude of the construction spike in January 2007 is best explained as an anticipatory response to the RFS2.

We thus define a *counterfactual plant* as any plant that was either under construction or already operating in January 2006 or earlier. To account for the dynamics of individual plants'

expansion or exit after 2007, we proceed as follows. If an expansion was already under construction by 2006, we assume that it would have occurred even without the RFS2 and include it in our counterfactual history. If an expansion began construction after 2006, we assume that it would not have occurred without the RFS2. We assume that all plant closures that occur after 2006 were due to increased competitive pressure driven by the RFS2, and (again conservatively) would not have occurred in the absence of the RFS2.

Table 1 presents a summary of numbers and capacities of U.S. corn ethanol plants over the period of 2001–2018. In addition to the actual observed history in the first two columns, Table 1 also reports the data for the counterfactual history constructed as per the foregoing description. The next two columns report the (conventional) biofuel mandates as per the 2005 RFS1 and the 2007 RFS2. It is apparent the actual history tracks well the RFS2 mandates, whereas the counterfactual history is consistent with the original RFS1 mandates (specified only up to 2012). The last column of Table 1 reports the total ethanol capacity projected by the USDA in February 2006, before the introduction of the RFS2. The fact that the counterfactual ethanol capacity we have constructed also matches rather well these USDA projections supports the contention that this counterfactual scenario is a reasonable estimate of what would have occurred had the RFS2 not been enacted.

2.2 Crop Choice Data

Crop choice data are assembled from two sources. First, we use the CDL published annually by the USDA National Agricultural Statistics Service (NASS). This dataset uses satellite imagery to map the crops grown across the United States at a 30-meter resolution. Satellite imagery and ground-truth validation are used to classify each pixel as one of over 100 potential crops or other land uses—including, most importantly for this study, corn and soybeans.

CDL data are available from 1997 to 2021; however, the spatial coverage is limited before 2008. CDL data start off only covering North Dakota in 1997, with the spatial coverage gradually expanding each year before covering the entire continental United States from 2008 onward. To fill in these spatial gaps, we complement the CDL with the CSDL published by S. Wang et al. (2020). The CSDL uses archives of satellite imagery recently made public and machine learning

to classify crops in 13 Midwest Corn Belt states from 1999 to 2018.[3] As with the CDL, the CSDL classifies land use at a 30-meter resolution. The CSDL is somewhat less specific in its classification than the CDL—each image pixel is classified as either corn, soybeans, "other" (any other crop), or non-cropland.

We focus our analysis on the 13 Midwest states covered by the CSDL to allow for observation of crop choices during the full study period. We perform our analysis at the Land Resource Region (LRR) level.[4] Eleven LLRs are represented in the area covered by the CSDL, but four of them (J, O, P, and R) have minimal overlap and are ignored in the analysis. Thus, we define the study area as the intersection of the CSDL area and the seven main LRRs involved. Note that the area thus identified encompasses much of the cultivated land in the 13 states listed, and contains the vast majority of U.S. production of corn, soybeans, and ethanol. Figure 1 depicts our study area and the boundaries of the LRRs used in the study.

We aggregate pixels classified by the CSDL and CDL to the field level using the CSB dataset recently released by the USDA (Hunt et al. 2023). The CSB delineates field boundaries based on contiguous areas of common cropping patterns (as observed in the CDL) over several years. We define the fields for our study using the 2018 CSB dataset, which relies on CDL observations from 2011–2018 to determine field boundaries. The CSB dataset includes the field-level crop observations from 2011–2018 and the size of each field in acres. We collect field-level crop observations for 2000–2010 by sampling one point from each field in the CSB dataset and taking the value of the CDL (where available) or CSDL (where the CDL is not available) for each point and each year.

Our approach to field aggregation is similar to that taken by Hendricks, Smith, and Sumner (2014), Pates and Hendricks (2021), and Lark et al. (2022). These studies also proceed by

[3] The 13 states covered by the CSDL are: North Dakota, South Dakota, Nebraska, Kansas, Minnesota, Iowa, Missouri, Wisconsin, Illinois, Michigan, Indiana, Ohio, and Kentucky.

[4] LRRs are areas of similar geography, soils, and climate defined by the USDA. Each LRR is composed of several Major Land Resource Areas (MLRAs), which further delineate areas of similarity.

sampling one pixel inside each field, but rely on field boundaries from the USDA's Common Land Unit (CLU) data from 2008. Though these data were publicly available in 2008, they are no longer in the public domain. The newly-released CSB datasets, on the other hand, are publicly available and created using open-source algorithms.

An alternative approach to pixel aggregation in the literature is grid-cell analysis. Motamed, McPhail, and Williams (2016) and Y. Wang et al. (2020) employ this approach by aggregating the pixel-level crop observations into shares of 10 km × 10 km grid cells. This procedure allows for computationally feasible analysis at a more granular level than a county. However, this approach still only permits aggregate analysis, whereas the methodology of our study permits analysis at the level of individual farmer choices.

We aggregate observed crop choices into three final categories: corn, soybeans, and everything else ("other"). This characterization of crop choices, while simple, is fully adequate for the research question of the paper. In fact, relative to related studies, maintaining the separation between soybeans and "other" crops provides the ability to observe potential changes in corn-soybean rotational practices. Table 2 shows summary statistics for observed fields. We focus our main analysis on large fields because, as noted by Pates and Hendricks (2021), small fields may be more susceptible to measurement error. Our definition of "large" selects fields with an area greater than or equal to 10 acres. From data reported in Table 2, it is apparent that such large fields, over all regions and years considered in this study, account for more than 91% of the corn and soybean area.

3. Ethanol Demand Index

As noted, the expanded demand for corn due to the RFS2 necessarily puts upward pressure on corn prices and, indirectly, on other commodity prices. In addition to this aggregate market effects—the proverbial "tide that lifts all boats"—earlier works note that the specific location of corn ethanol plants affects the local price basis because transportation costs make local supply more desirable to satisfy the local corn demand for ethanol use (McNew and Griffith 2005). Rather than modeling the crop planting response to prices, however, the strategy of this paper is to pursue the more direct approach of relating crop patterns to local ethanol processing

activity. As in related recent applications (Fatal and Thurman 2014; Motamed, McPhail, and Williams 2016; Li, Miao, and Khanna 2019; Y. Wang et al. 2020), we develop an index that captures the local demand pressure arising from the presence of ethanol plants.

To motivate the index we develop, consider first a single ethanol plant j with capacity K_j that enters the local market with local corn price p. This price, which represents the farmers' opportunity cost, may differ across regions because of pre-existing basis patterns (reflecting distance to terminal markets and/or the extent of local corn demand for other uses, such as livestock feed). The plant procures corn from the surrounding area, and there is a per-unit transportation cost $\tau > 0$ to move corn to the ethanol plant. Hence, the plant must offer a premium $\Delta_j > 0$ to attract delivery, such that the plant-gate price paid by the plant is $p + \Delta_j$. If farmer i, located a distance d_{ij} from the plant, sells their corn to the ethanol plant, they receive a net premium $\Delta_{ij} = \Delta_j - \tau d_{ij}$, and will find it profitable to deliver to the plant as long as the net premium is positive. Thus, for a given plant-gate corn price premium Δ_j, the plant will attract supplies from farmers located within the radius $d_j = \Delta_j / \tau$.

To operate at full capacity, the plant will offer a plant-gate premium Δ_j^0 that is just high enough for the implied procurement radius d_j^0 to encompass sufficient supplies. If corn is produced uniformly in the surrounding area, then the procurement area is proportional to the plant's capacity and the procurement radius is proportional to the square root of the capacity. The demand pressure that plant *j* exerts on farmer *i* located at distance d_{ij} from the plant is proportional to the net premium, which, based on the foregoing, can be rewritten as $\max\left\{\tau\left(d_j^0 - d_{ij}\right), 0\right\}$. Because the procurement radius depends on the plant's capacity, one can equivalently express the demand pressure by the index

$$Z_{ij} = \max\left\{\kappa\sqrt{K_j} - d_{ij}, 0\right\} \tag{1}$$

This formulation thus provides a simple framework for analyzing how an ethanol plant's capacity and location combine to impact farms in the surrounding area. Importantly, it also provides guidance when multiple ethanol plants operate in the same area.

Consider a second ethanol plant m, with capacity K_m and radius of procurement d_m^0, that enters the area described above. The two plants, j and m, are located a distance $\tilde{d}_{jm}$ from each other. If $\tilde{d}_{jm} \geq d_j^0 + d_m^0$, then there is no overlap in the plants' procurement areas, and they each operate exactly as described above. However, if $\tilde{d}_{jm} < d_j^0 + d_m^0$, then the plants' procurement areas do overlap. This overlap means that some farmers who had been selling their corn to plant j will now be able to receive a higher net premium from m. Similarly, plant m will fail to attract corn from some farmers within its stand-alone procurement radius because they receive a higher net premium from j.

There are two key points worth noting in the case of overlapping procurement areas. First, each plant will need to increase its plant gate premium in order to attract sufficient corn to meet capacity. This action effectively extends the procurement radius, with plant j's procurement radius increasing to $\overline{d}_j > d_j^0$, and similarly for plant m. Second, arbitrage considerations imply that farmers will always sell their corn to the plant providing them with the highest net premium. As shown in the foregoing for the single-plant case, the net premium is proportional to the farmer's distance to the boundary of the plant's procurement area. Thus, a farmer located inside multiple plants' procurement areas will sell their corn to the plant whose procurement boundary is farthest from the farmer's location. Putting these two points together, we can write the ethanol demand pressure on farmer i, for the general case of a set of plants $\mathcal{J}$ with overlapping procurement areas, as:

$$Z_i = \max\left\{\max_{j\in\mathcal{J}}\left(\overline{d}_j - d_{ij}\right), 0\right\} \tag{2}$$

The key insight here is that the number of plants or total ethanol capacity within an area does not enter directly into the farmer's decision-making. Previous studies use measures such as total

capacity within 100 kilometers of a grid-cell (Motamed, McPhail, and Williams 2016), total capacity weighted by distance from plant to county centroid (Fatal and Thurman 2014), or total capacity weighted by the share of a circle around the plant that falls within a county (Li, Miao, and Khanna 2019), directly in their estimating equation. Our index, on the other hand, recognizes the importance of the locations and capacities of nearby plants in determining $\bar{d}_j$, the procurement radius accounting for any overlap in individual procurement areas, and further maintains that these capacities influence a farmer's decisions only through their impact on $\bar{d}_j - d_{ij}$, the farmer's distance to the procurement area boundary.

The foregoing index addresses similar issue as Y. Wang et al. (2020), who note that the problem of exactly computing an adjusted procurement radius for each plant that properly accounts for interference from competing plants is intractable. They solve this problem by linearizing onto the two-dimensional space, such that each plant's procurement area overlaps with at most two other plants' areas. We provide an alternative, and somewhat more transparent approach, by approximating the true adjusted procurement radius via the procedure below. In essence, Y. Wang et al. (2020) compute an exact solution to the simplified problem of plants arranged linearly, whereas we compute an approximate solution using the true spatial distribution of plants on the two-dimensional plane. The algorithm for finding this approximate solution, which we implement in Matlab R2022a, is as follows:

(a) Compute each ethanol plant's independent procurement radius d_{jt}^0—the procurement radius before accounting for any interference from other plants. (Note that we now add a time subscript t to d_{jt}^0 and other variables from the conceptual model that may change year to year.)
(b) For each plant j, find all "nearby plants"—those whose procurement areas overlap j's own procurement area.
(c) Compute an adjusted capacity for each plant by adding the capacities of all nearby plants, weighted by the distance between the plants.

(d) Compute an adjusted procurement radius $\bar{d}_{jt}$ that accounts for interference from nearby plants, using the adjusted capacity computed in (c).

(e) Use the adjusted procurement radius of each plant to compute Z_{it}, the ethanol demand index for each field in each year.

More details on the computation of the ethanol plant index are provided in Appendix B**.** The end result is that we can compute the ethanol demand index Z_{it} for each field i and each year t, according to equation (2). This index of course reflects the observed history of ethanol plants' location and entry into production. Using the counterfactual history of ethanol plant capacities described in section 2, the algorithm can be used to compute the counterfactual values Z_{it}^{CF} that each plot would have experienced in each year had the RFS2 had not been enacted. Figure 3 shows the spatial distribution of the indices in selected years. Appendix Table A.1 provides summary statistics for the actual and counterfactual ethanol demand indices for the fields in our analysis.

4. Empirical Framework

To assess the impact of the RFS2 on cropping decisions, we estimate and compare the long-run areas of various crop patterns under both the actual and counterfactual ethanol histories. Following Hendricks, Smith, and Sumner (2014), and Pates and Hendricks (2021), we adopt a Markov chain approach that is consistent with farmers' use of rotation strategies. We adapt and extend the framework in two directions. First, as noted, our setup explicitly includes three planting options on any one plot: corn, soybeans, and other crops. Thus, we avoid pooling soybeans with everything else, which likely reduces pooling bias issues at the estimation stage and provides more nuanced and insightful inferences at the counterfactual-simulation stage. Second, we derived the Markov transition probabilities from a choice probability model rather than estimating transition probabilities directly. The latter approach, adopted by Pates and Hendricks (2021), requires different samples for different Markov transition probabilities. This makes it impossible to account for fixed effects that cut across planting histories, which are an integral part of our identification strategy.

4.1. Choice probabilities and Markov transition probabilities

At any point in time, a field of cropland can be planted with either corn (C), soybeans (S), or any other crop (O). Let $y_{it} \in \{C,S,O\}$ denote the choice pertaining to field i at time t. The Markov *transition probability* $P_{it}^{\ell k}$ for field i is the probability of observing crop choice k at time t conditional on having observed crop choice ℓ at time $t-1$: $P_{it}^{\ell k} \equiv \Pr[y_{it} = k | y_{it-1} = \ell]$, for $k, \ell \in \{C,S,O\}$. Thus, given the set of three possible "states" at time $t-1$, there are nine such transition probabilities, which can be arranged into a 3x3 Markov transition matrix:

$$\mathbf{M}_{it} = \begin{pmatrix} P_{it}^{cc} & P_{it}^{cs} & P_{it}^{co} \\ P_{it}^{sc} & P_{it}^{ss} & P_{it}^{so} \\ P_{it}^{oc} & P_{it}^{os} & P_{it}^{oo} \end{pmatrix} \tag{3}$$

Note that there are only six independent transition probabilities because each row of $\mathbf{M}_{it}$ must sum to 1, that is $P_{it}^{\ell c} + P_{it}^{\ell s} + P_{it}^{\ell o} = 1$, $\ell \in \{C,S,O\}$.

Transition probabilities can be recovered from a standard choice model where the objects of interest are choice probabilities. The multinomial logit model is a natural framework for choice probabilities. Let the payoff (expected profit) associated with choosing crop k on field i at choice occasion (time) t be

$$\Pi_{kit} = V_{kit} + \varepsilon_{kit} \quad , \; k \in \{C,S,O\} \tag{4}$$

where V_{kit} is the expected per-unit payoff and ε_{kit} is an i.i.d. random term. With this setup, choice probabilities are defined as $\Pr[y_{it} = k] = \Pr[\varepsilon_{\ell it} - \varepsilon_{kit} \le V_{kit} - V_{\ell it}, \forall \ell \ne k]$, which makes it clear that only differences in expected payoff matter. Thus, without loss of generality, we normalize the payoff of the "other" option, that is $V_{oit} \equiv 0$, $\forall i,t$. Further assuming that the

random terms ε_{kit} follow a Type 1 Extreme Value (T1EV) distribution, then the choice probabilities of interest can be written as (see, e.g., Train 2009):[5]

$$\Pr\left[y_{it}=k\right]=\frac{e^{V_{kit}}}{1+\sum_{\ell\in\{c,s\}}e^{V_{\ell it}}} \quad , \quad k=C,S,O\,. \tag{5}$$

To estimate choice probabilities, the expected plot-specific payoffs are conditioned on a set of relevant variables. Crucial to our focus on rotational effects to be captured by transition probabilities, the conditioning variables include state variables identifying the crop planted on plot i in the previous period. This is accomplished by defining two binary dummy variables C_{it} and S_{it}, which take a value of 1 if, respectively, the plot was planted to corn or soybeans at time t (the third option, that the field was planted with other crops, corresponds to $C_{it}=S_{it}=0$). For a first order Markov model with regional and time fixed effects, as well as the covariate related to ethanol expansion, we write:

$$V_{ikt}=\alpha_k+\beta_{ck}C_{it-1}+\beta_{sk}S_{it-1}+\lambda_k Z_{it}+\varphi_k R_t+\gamma_k^r+\omega_k^t \quad , \quad k\in\{c,s\} \tag{6}$$

Here Z_{it} is the local ethanol demand index described earlier, which provides a metric for the maximum net ethanol premium available to plot i given the spatial configuration and capacities of plants in year t. Beyond the local impact of the RFS via the index Z_{it}, the payoff of each crop in (6) also responds to the aggregate RFS mandate R_t, which, together with the foregoing local impact, permits the model to capture the overall impact of biofuel policies on crop choices. The binary dummy variables $\{C_{it-1},S_{it-1}\}$ capture the impacts of crop rotation practices on current crop payoffs (conceivably by affecting yields and/or costs). Their inclusion

[5] Train (2009, pp. 50–52) briefly discusses panel data settings, as well as dynamic contexts that involve state variables (such as inertia). He also notes what he calls the "variety seeking" situation, where the agent gets a higher utility by not choosing the same alternative as in the previous period, which applies directly to the rotation problem we are studying.

allows us to estimate the Markov transition probabilities and subsequently the long-run unconditional probabilities of interest.

The fixed effects in equation (6) are critical to our identification strategy. By assumption, the national ethanol mandate, denoted R_t, and the local ethanol demand index Z_{it} are driven by the RFS policy and are exogenous to farmers' choices. The crop-specific time fixed effects ω_k^t are meant to flexibly capture residual components of crop-specific payoffs that vary across time, such as commodity demand effects arising from non-biofuel sources and input prices. It is apparent, however, that identifying $\varphi_k R_t$ directly in (6), separately from the other time fixed effects ω_k^t, is not desirable.[6] Our strategy, therefore, is to carry out the analysis first by use of the composite time fixed effects $\delta_k^t \equiv \varphi_k R_t + \omega_k^t$. Having estimated these composite time fixed effects, we can subsequently identify the portion attributable to the systematic impact of the RFS2 expansion.

The region fixed effect γ_k^r flexibly controls for aspects of the payoff that vary across space, such as the pre-existing local basis, and climate and soil conditions that affect production costs. We note at this juncture that the model in equations (5)–(6) is applied separately for each of the seven LRRs discussed earlier (Table 2). Thus, the region identifier r pertains to sub-regions of each LRR. One possibility is to identify these regions by the MLRAs of each LRR—for the seven LRRs, this would give us a total of 79 subregions. As noted by a reviewer, however, a more refined delineation may be desirable to ensure homogeneity in pre-existing crop patterns, which is necessary for the fixed effects to control for potential endogeneity in the location choices of ethanol plants. To achieve a narrower spatial delineation, therefore, the region superscript r here identifies the product of MLRAs and Crop Reporting District (CRD). This procedure results in a total of 436 distinct regions across the seven LRRs of the study area.

[6] Direct estimation of the φ_k parameters requires dropping one of the time fixed effects ω_k^t, and the resulting estimate is conditional on which of these fixed effects is dropped.

Markov transition probabilities are readily identified from the estimated choice probabilities. Recalling that $P_{it}^{\ell k} \equiv \Pr\left[y_{it} = k \middle| y_{it-1} = \ell \right]$, then from equation (5) the Markov transition probabilities satisfy:

$$P_{it}^{ck} = \frac{e^{\alpha_k + \beta_{ck} + \lambda_k Z_{it} + \gamma_k^r + \delta_k^t}}{1 + \sum_{\ell=c,s}^{2} e^{\alpha_\ell + \beta_{c\ell} + \lambda_\ell Z_{it} + \gamma_\ell^r + \delta_\ell^t}} \quad , k = C, S \quad \text{and} \quad P_{co} = 1 - P_{cc} - P_{cs} \tag{7}$$

$$P_{it}^{sk} = \frac{e^{\alpha_k + \beta_{sk} + \lambda_k Z_{it} + \gamma_k^r + \delta_k^t}}{1 + \sum_{\ell=c,s} e^{\alpha_\ell + \beta_{s\ell} + \lambda_\ell Z_{it} + \gamma_\ell^r + \delta_\ell^t}} \quad , k = C, S \quad \text{and} \quad P_{so} = 1 - P_{sc} - P_{ss} \tag{8}$$

$$P_{it}^{ok} = \frac{e^{\alpha_k + \lambda_k Z_{it} + \gamma_k^r + \delta_k^t}}{1 + \sum_{\ell=c,s} e^{\alpha_\ell + \lambda_\ell Z_{it} + \gamma_\ell^r + \delta_\ell^t}} \quad , k = C, S \quad \text{and} \quad P_{oo} = 1 - P_{oc} - P_{os} \tag{9}$$

4.2. Related Literature

Some comments on related literature may be helpful at this juncture. Our setup here is similar to that of Quoss et al. (2011), who provide a concise summary of earlier applications of Markov chains to crop cover, and discuss the problem of estimating transition probabilities. They note the distinction between "unconditional measures"—i.e., based on observed frequencies of state transition, as in earlier statistical characterizations of Markov chains (e.g., Billingsley 1961)—and estimates that account for time-varying and geographically fixed determinants (as reflected, in our setting, in the conditioning variables in (6)). Kim et al. (2020) also appear to derive transition probabilities from an estimated multinomial logit model of choice probabilities. Pates and Hendricks (2021), and Paton et al. (2014), on the other hand, estimate the transition probabilities directly. In the foregoing parameterization, this procedure would entail the estimation of equations (7), (8), and (9) directly. Such an approach is unappealing, however, because to estimate (7) one would need to condition on fields for which corn was grown at time $t-1$, whereas to estimate (8) one would need to condition on fields for which soybeans were grown at time $t-1$. Separate estimation based on disjointed data samples makes it problematic to account for fixed effects that cut across histories (such as the time fixed effects δ_k^t and the regional fixed effects γ_k^r), which is straightforward in our model. Furthermore, because Markov

transition probabilities are nested in the estimated choice probability model, the framework we have outlined also permits tests of hypotheses that cut across transition probabilities.

4.3. Stationary probabilities and crop patterns

By affecting choice probabilities, and thus Markov transition probabilities, exogenous factors such as the RFS2-induced corn demand expansion impact farmers' crop choices. To characterize the long-run consequences in terms of crop patterns, we focus on two main aspects. First, we assess *crop totals*—that is, the long-run total area allocated to a particular crop in a given year (such as the total corn area). Second, we consider the long-run frequencies of two-year *crop sequences*—for example, corn followed by corn (corn-corn) or soybeans followed by corn (soybeans-corn).

Because of the Markov chain nature of the model, predicted field crop allocations are inherently probabilistic. To characterize the change in crop patterns due to the RFS2, therefore, we rely on the long-run probabilities of observing those crop patterns, unconditional on the crops previously planted. The long-run unconditional probability of observing each crop is given by the stationary row vector of the Markov chain (e.g., Norris 1998):

$$\boldsymbol{\pi}_{it}\mathbf{M}_{it} = \boldsymbol{\pi}_{it} \tag{10}$$

where $\boldsymbol{\pi}_{it} = \begin{bmatrix} \pi_{it}^{c} & \pi_{it}^{s} & \pi_{it}^{o} \end{bmatrix}$ is the row vector of unconditional probabilities of corn, soybeans, and other, respectively; $\mathbf{M}_{it}$ is the Markov transition matrix defined in equation (3); and, the elements of the stationary vector also satisfy the normalizing condition $\pi_{it}^{o} = 1 - \pi_{it}^{c} - \pi_{it}^{s}$. The subscript "$it$" emphasizes that the corresponding vector and matrix are specific to field i and as determined by the vector of conditioning variables pertaining to year t. Under some regularity conditions (the Markov process is irreducible and recurrent), the stationary probabilities are unique and, for our small state-space case, can be readily solved analytically in terms of the Markov transition probabilities. Suppressing subscripts for notational clarity, the unconditional probabilities of corn and soybeans are, respectively:

$$\pi^{c} = \frac{P^{os}(P^{sc} - P^{oc}) + P^{oc}(1 - P^{ss} + P^{os})}{(1 - P^{ss} + P^{os})(1 - P^{cc} + P^{oc}) - (P^{sc} - P^{oc})(P^{cs} - P^{os})} \tag{11}$$

$$\pi^{s} = \frac{P^{oc}(P^{cs} - P^{os}) + P^{os}(1 - P^{cc} + P^{oc})}{(1 - P^{ss} + P^{os})(1 - P^{cc} + P^{oc}) - (P^{sc} - P^{oc})(P^{cs} - P^{os})} \tag{12}$$

The stationary probabilities in equations (11)–(12) can be used to characterize how a shock to the Markov transition probabilities (as arising from the ethanol demand expansion) translates into changes in the long-run crop acreage distribution. A complementary characterization of the impacts of such a shock relies on the probability of observing a given two-year sequence of crop choices—say, x_{it-1} followed by choice y_{it}, where $x, y \in \{c, s, o\}$. The long-run probability of this choice sequence, denoted p_{it}^{xy}, can then be computed as:

$$p_{it}^{xy} = \pi_{it}^{x} P_{it}^{xy} \tag{13}$$

We estimate the model in equations (5) and (6) separately for each LRR to allow for heterogeneity in the response of crop choices to nearby ethanol plants across the study area. Having estimated the choice probability model, we derive the Markov transition probabilities as per equations (7), (8), and (9). We can then compute the unconditional crop total and sequence probabilities with (11)–(12) and (13). We compute the (expected) areas of crop totals and two-year crop sequences from these unconditional probabilities.

5. Results

As noted earlier, in the main body of the paper we report results from using "large" fields only (i.e., fields with and area greater than 10 acres). Results for "small" fields, and for all fields, are reported in the Supplementary Appendix.

5.1. Baseline model

The estimated parameters of the multinomial logit choice model defined by equations (5) and (6) are reported in Table 3, separately for each LRR and for both the corn and soybean choice

probabilities.[7] Standard errors are adjusted for clustering at the region (MLRA×CRD) level to account for possible spatial effects.[8] These coefficients are somewhat difficult to interpret on their own. Nevertheless, interesting patterns are visible in Table 3. The sign on the ethanol demand index parameter is generally positive for both the corn and soybeans equations, indicating that ethanol demand increases the value of planting both corn and soybeans relative to any other crop. The value of corn-soybean crop rotation is apparent in the estimated parameters of the state variables (previous crop) in both the corn and soybeans equations—for all LRRs, the soybean dummy typically has a higher value than the corn dummy in the corn equation, while the opposite is true in the soybean equation.

The econometric relevance of the terms included in the model is explored in Table 4, which reports the results of the Wald test concerning four hypotheses.[9] The first hypothesis is that there are no regional fixed effects, that is $H_0: \gamma_k^r = 0,\ \forall\, r,\ k = C, S$. This hypothesis, therefore, concerns whether the various MLRAs in the same LRR can be treated as homogeneous for the purpose of the analysis. The second hypothesis is that there are no time fixed effects, that is $H_0: \delta_k^t = 0,\ \forall t, k = C, S$. Failing to reject this hypothesis would imply that crop choices were not affected by other systematic, time-varying conditions beyond those included in the model. The third hypothesis is that the state variables (capturing the identity of the previous year's crop choice) matter, that is $H_0: \beta_{\ell k} = 0,\ \ell, k = C, S$. Failing to reject this hypothesis would imply that history does not matter for the specific crop choice in a given field (i.e., there are no rotation effects). Finally, the fourth hypothesis concerns the impact of the ethanol demand index (i.e.,

[7] Estimated with the `mlogit` command in Stata 17.

[8] Abadie et al. (2023) note that clustered standard errors can be severely inflated, especially in applications such as this one where all potential clusters of the population are observed. We report clustered standard errors, but note that they are likely overly conservative estimates of the true standard errors of our estimates.

[9] Because the Wald test relies on the estimated covariance matrix of the unrestricted model only, it can readily accommodate the clustering of standard errors we have implemented.

$H_0: \lambda_k = 0,\ \ k = C, S$). Failing to reject this hypothesis would suggest that the diffusion of ethanol plants induced by the RFS2 did not affect crop choices.

The Wald statistics of Table 4 were computed separately for each of the seven LRRs. From Table 4 it is apparent that the hypotheses of interest were decisively rejected at the 1% significance level in almost all cases. The two exceptions concern the test for the fourth hypothesis (about the ethanol demand index) for regions F and K. This is not unreasonable, given that these LRRs are at the northern periphery of the area of study (see Figure 1). Overall, the results in Table 4 provide support for the empirical relevance of the estimated model.

5.2. Local impact of the RFS2

The estimated coefficients associated with the ethanol index variable Z_{it} are informative of the "local" impact of the RFS2—that is, the impact on a given field from its proximity to ethanol processing plants, conditional on the non-local effects that, at this stage, are subsumed in the estimated time fixed effects. To characterize the extent of these local effects, we compare the predicted areas of the estimated baseline model with those of the model with the counterfactual ethanol plant capacities that would have prevailed absent the RFS2. Specifically, counterfactual choice probabilities and counterfactual transition probabilities are obtained by replacing Z_{it} with Z_{it}^{CF} (the counterfactual ethanol demand index values) in computing the Markov transition probabilities from equations (7), (8), and (9). As described above, we then compute unconditional crop total and sequence probabilities and areas for the counterfactual ethanol history. Comparing the estimated areas of crop patterns obtained under the actual and counterfactual ethanol capacity histories gives an estimate of the impact of the RFS2-induced expansion in Midwest ethanol production capacity on surrounding crop patterns.

Results for the foregoing exercise are reported in Table 5. We find that over the 2008–2018 period, when the RFS2 was in effect, the local impacts increased corn area by an average 1.01

million acres per year, an increase of 1.6% from the counterfactual area.[10] Overall, this expansion of corn area was mostly at the expense of other crops (the area for which decreased by 1.4%). Looking at predicted crop sequence probabilities, the local impacts of the RFS2 entail a higher frequency of farmers planting corn after corn (which increases by 4.3%), and also some increase in corn/soybean rotation sequences (corn following soybeans increases by 0.7%, and soybeans following corn increases by 0.6%).

Figure 4 gives a more detailed picture of the local impacts of the RFS2-induced ethanol expansion on crop patterns. The focus of this figure, as well as later similar figures, is the main crop patterns of interest: crop totals for corn and soybeans, and rotations involving exclusively corn and soybeans. For the latter, note that because the order of crops matters for the two-year sequences, the sequence probability for corn-soybeans in general differs from that of soybeans-corn. In practice, however, these two sequences move in tandem, and we combine them into a single measure of corn-soybean rotation.

Increases in corn area were highest in a belt from the eastern Dakotas, through Iowa, southern Minnesota, and parts of Wisconsin and Illinois, to Indiana and western Ohio. The local impacts of ethanol capacity on soybean area was more varied, as Iowa, southern Minnesota, and northern Illinois switched from corn-soybean rotation to corn monoculture. Meanwhile, areas of the Dakotas, Nebraska, and Kansas switched from cultivating other crops to corn-soybean rotation, increasing their total soybean areas.

5.3. Overall impacts of the RFS2

The logic of the model, so far, has been to identify field-level choice probabilities, and Markov transition probabilities, conditional on an index tracking the local corn demand impact of the RFS2-induced proliferation of ethanol plants, and conditional on time fixed effects. The latter, as

[10] To compute standard errors in Table 5 (and later in Table 7), we draw a random vector of parameter values using the estimated parameter values and covariance matrix. We then use this random parameter vector to compute the actual and counterfactual crop pattern areas as described in the text. We repeat this procedure 1000 times and report the standard deviation of the resulting predicted areas.

noted earlier, include the market-wide impact of the RFS mandates. Specifically, we have postulated that $\delta_k^t \equiv \varphi_k R_t + \omega_k^t$, where R_t is the metric for the national RFS mandate, and ω_k^t is the pure time fixed effect meant to control for elements of crop-specific payoffs that vary across time for reasons unrelated to biofuel policies (such as changes in commodity demand from other sources, and/or input prices).

That the RFS may be a major explanation for the estimated composite time fixed effects can be evinced from Figure 5, where we plot these time fixed effects for LRR region M (which includes the central Corn Belt and accounts for almost half of the cultivated acres in the study area). It is apparent that these estimated fixed effects have been consistently positive since 2007 for corn, and since 2009 for soybeans, clearly indicating a pattern consistent with the time path of the RFS mandate. Corresponding diagrams for all other LRRs in this study, reported in Appendix D, show a broadly similar pattern.

As in some other contexts, it is useful to relate estimated fixed effects to observables (e.g., Nevo 2001; Arcidiacono et al. 2012). Given the consistent estimates of the composite fixed effects $\hat{\delta}_k^t$, we can write $\hat{\delta}_k^t \equiv \delta_k^t + e_k^t$ and thus consider the regression

$$\hat{\delta}_k^t = \mu_k + \varphi_k R_t + u_k^t \quad , k = C, S \tag{14}$$

where $\mu_k \equiv E\left[\, \omega_k^t \,\right]$ is the unconditional expectation of the pure time fixed effects, and $u_k^t \equiv e_k^t + (\omega_k^t - \mu_k)$. Under the assumption that the exogenous RFS mandate R_t is orthogonal to the pure time fixed effects ω_k^t, ordinary least squares on equation (14) provide an unbiased estimate of the portion $\hat{\varphi}_k R_t$ of the estimated composite time fixed effects attributable to the RFS.

Results of the fixed effects regressions are reported in Table 6. It is apparent that the RFS mandate explains a sizeable portion of the estimated composite time fixed effects—the weighted average of the R^2 across the seven regions is about 0.75 for corn and 0.56 for soybeans. Having estimated the parameters in Table 6, we can combine these results with those

in Table 3 and perform a counterfactual experiment that assesses the overall additionality impact of the RFS2.

The counterfactual experiment is analogous to that described in the previous subsection to estimate the local impact of RFS2-induced ethanol capacity expansion on crop patterns. In addition, we construct a counterfactual composite time fixed effect for each year as:

$$\tilde{\delta}_k^t = \hat{\delta}_k^t + \hat{\varphi}_k(R_t^{CF} - R_t) \quad , k = C, S \tag{15}$$

Here, the counterfactual variable R_t^{CF} is the level of the RFS mandate that would have prevailed without the RFS2, which we assume to be the observed levels of the RFS1 in Table 1, with the mandate continuing at its 2012 level for 2013–2018.

We use the counterfactual time fixed effects in (15), along with the counterfactual values of the ethanol index Z_{it}^{CF} discussed earlier, to compute a set of counterfactual Markov transition probabilities. These transition probabilities are used, in turn, to compute the predicted stationary probabilities and expected crop areas of the counterfactual scenario. Comparing the outcomes of this counterfactual with the predicted outcomes of the baseline model provides an estimate of the full impact of the RFS2.

Results for the full impact of the RFS2 are presented in Table 7. We find that the 2007 expansion of the RFS increased corn area in the study region by 9.7% and soybean area by 6.7%. These substantial increases came as the area of other crops decreased by 13.9%. Turning to the two-year sequences, the area of corn following corn showed the largest relative increase, at 16.8%, while the corn-soybean and soybean-corn sequences each increased by about 10%.

There is substantial geographic heterogeneity in the RFS2's impact on crop patterns. Figures 6 and 7 show the changes in the areas of five key crop patterns by state and county, respectively. South Dakota saw the largest increase in corn area due to the RFS2, with a concurrent large increase in soybean area as farmers switched from cultivating other crops and adopted corn-soybean rotation. Iowa saw the next largest increase in corn area due to the RFS2, with

negligible increase in soybean area, as many farmers shifted from corn-soybean rotation to corn monoculture.

Figure 6 reveals these patterns at an even finer geographic resolution, showing the county-level change in crop pattern areas. We see that corn-corn area increased throughout much of the study area, with increases especially concentrated in Iowa, southern Minnesota, and northern Illinois. The corn-soybean crop rotation pattern increased drastically in North Dakota, South Dakota, Nebraska, Indiana, and Ohio. Together, these changes led to an increase in corn area throughout the study area, and a shift at the margins of soybean area from Iowa, southern Minnesota, Illinois, and Indiana toward the western edge of the Corn Belt.

6. Discussion and Conclusion

This study adds to ongoing efforts to analyze the impacts of U.S. biofuel policy on observed crop choices. The extant research in this area has employed a wide variety of methodologies and counterfactuals in determining these impacts (Austin, Jones, and Clark 2022 provide a comprehensive review). A contribution of this study is to derive a modeling framework whereby the exogenous corn demand expansion engineered by the RFS directly impacts field-level crop choices, and to characterize the spatial heterogeneity of this impact mediated by the proliferations of RFS-induced ethanol plants.

Similar to several previous related studies, we develop an index of ethanol demand pressure to estimate the local impacts of ethanol expansion on nearby crop production. We make two innovations to the development and use of such an index. First, like Y. Wang et al. (2020), we overcome the complexities of grounding the index in an explicit spatial equilibrium model. Whereas Y. Wang et al. (2020) pursue an exact solution to a simplified spatial problem, we implement an approximate solution that uses the actual two-dimensional spatial configuration of plants. Second, we perform our analysis at the level of an individual field rather than a county or grid-cell. This approach differs from existing ethanol index-based studies, but is similar to previous work estimating crop price elasticities (Hendricks, Smith, and Sumner 2014; Pates and Hendricks 2021) and the impact of the RFS2 (Lark et al. 2022). Unlike these previous studies, however, we embed Markov transition probabilities within a multinomial logit model

of choice probabilities. This allows us to control for time and regional fixed effects that cut across crop histories. A further innovation of our study is in analyzing the rich spatial data based on satellite imagery using the newly released and publicly available CSB dataset.

Due to the differing units of observation, comparing our local-impact results to those of grid-based studies that utilize an ethanol index is difficult. Two points are worth mentioning, however. First, our study allows for spatial heterogeneity in the impact of ethanol expansion on crop patterns and finds it to be economically significant. We show that the RFS2-induced expansion in ethanol capacity increased corn area broadly, but especially along a belt from the eastern Dakotas, through southern Minnesota, Iowa, and parts of Wisconsin and Illinois, and into Indiana and western Ohio. Meanwhile, the local impact of ethanol expansion on soybean area was even more spatially heterogeneous, with decreases in soybean area in Iowa and parts of Minnesota and Illinois as fields shifted from corn-soybean rotation to more continuous corn planting, and increases in soybean area around the periphery of the Corn Belt as fields transitioned from cultivating other crops to corn-soybean rotation.

The central research question of the paper is determining the impact of the RFS2 on crop patterns in the United States. Li, Miao, and Khanna (2019) estimate that the expansion of ethanol capacity from 2008 to 2014 increased corn area by 3.1%. Lark et al. (2022) find that the RFS2 increased corn area by 8.7% during the 2008 to 2016 timeframe. Our estimate, that the RFS2 increased corn area in the U.S. Corn Belt by an average of 9.7% over the 2008 to 2018 period, aligns well with the results of Lark et al. (2022). The geographic pattern of our results is broadly similar to those found by both Li, Miao, and Khanna (2019) and Lark et al. (2022), with the largest increases in corn area coming in South Dakota, Iowa, and Minnesota.

A novel contribution of the Markov chain parameterization that we have developed is to measure the RFS2's impact on soybean cultivation as well, separately from the impact on corn and other crops. We find that the RFS2 significantly increased soybean area. In some parts of the western Corn Belt, this increase was partly through the local impact of ethanol expansion, as increased ethanol capacity in the area led more fields to switch from other crops to corn-soybean rotation. Most of the increased soybean area, however, was due to the RFS2's estimated

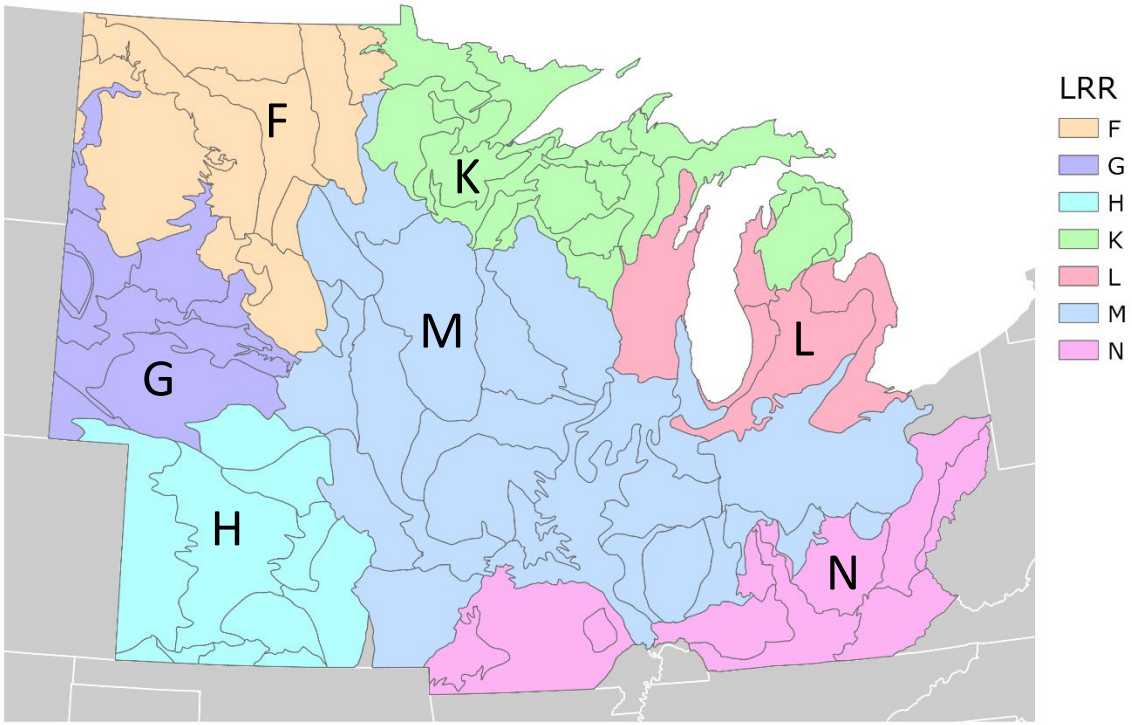
impact on national demand levels, as measured by our regression of the estimated year fixed effects on the level of the RFS mandate. This increase in soybean area is consistent with the implication of most equilibrium models used to analyze biofuel policies, whereby the increase in corn demand directly attributable to the RFS spills over to related commodity markets, as substitution effects in both demand and supply take place. Consistent with these results, Lark et al. (2022) estimate that the RFS2 increased soybean and wheat prices by about 20%, in addition to increasing corn prices by about 30%, whereas Roberts and Schlenker (2013) estimate that the RFS2 increased prices for corn, rice, wheat, and soybeans by 20%.

Is the analysis of this paper attributing too much to the RFS2? To put the results in a broader context, consider the pre-RFS years of the study period. Over the 5 years 2001–2005, USDA statistics for the 13 states in the study area show that corn cultivation averaged 67.2 million acres and soybeans averaged 62.6 million acres. During the last 5 years of the study period (2014–2018) for these 13 states, by contrast, corn area averaged 76.4 million acres, a 9.2 million acre increase, while soybean area averaged 71.2 million acres, a 8.6 million acre increase. Over the study timeframe, therefore, our estimates attribute about 62% of the observed change in corn area, and about 40% of the observed change in soybean area, to the implementation of the RFS2.

Farmers' response to the large increase in corn demand created by this biofuel policy has been mediated by changes in the widespread system of corn-soybean rotation, a practice that provides many environmental benefits, including breaking pest cycles and replenishing nitrogen in the soil. As with some previous studies, rotation choices are a central element of this paper's modeling approach. By focusing on the spatial relationships between ethanol plants and crop patterns, we provide direct evidence that the RFS2 has led to an increase in corn area and corn-corn rotation practices throughout the U.S. Corn Belt and a shift in the geography of soybean cultivation and corn-soybean rotation practices. How these shifts in rotation practices ultimately translate to measurable environmental outcomes is beyond the scope of the current paper. Lark et al. (2022) have made a substantial effort in that direction, and the complementary analysis we have developed in this paper might provide opportunity for future work.

Figures and Tables

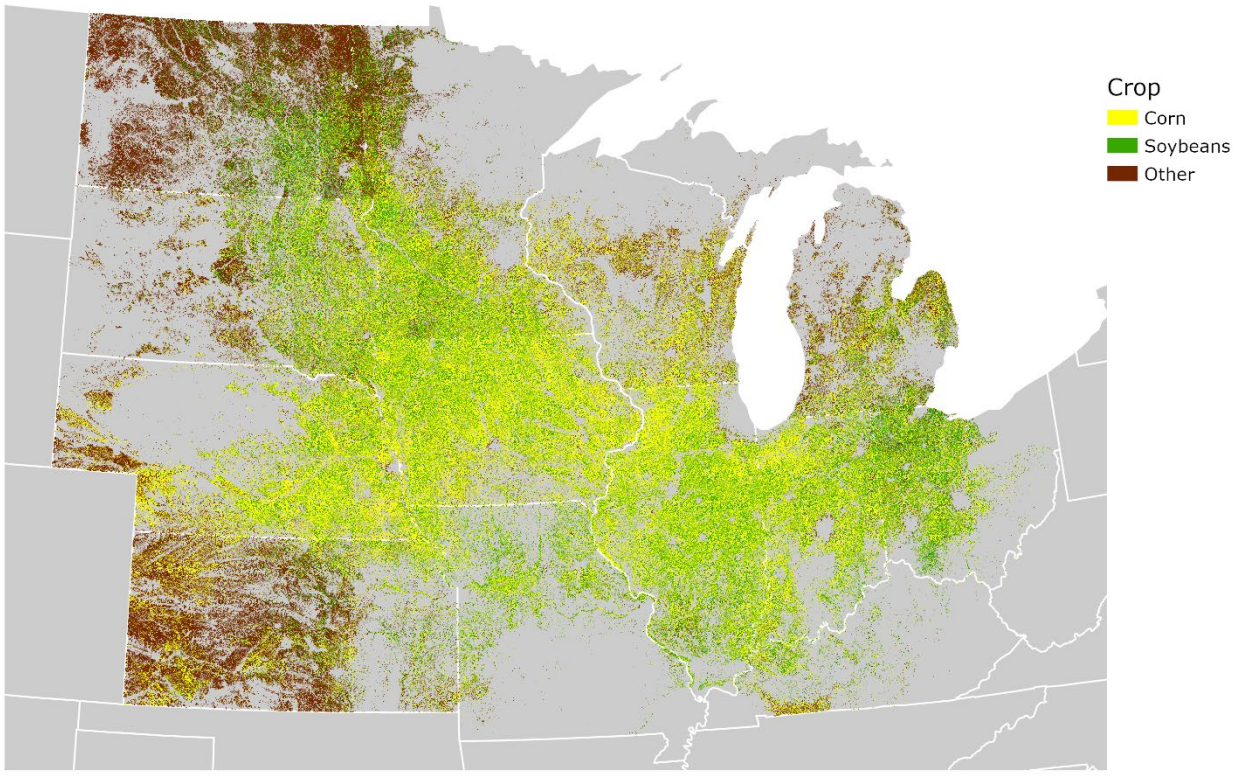
Figure 1. Study area



Note: This figure depicts the area covered by this study. We define the study area as the intersection of the 13 Midwest states covered by the Corn-Soy Data Layers with seven Land Resource Regions (LRRs) that cover the majority of the area in these states. Our analysis is performed separately for each of the seven LRRs to allow for differing responses to ethanol across the study area.

Each LRR is composed of several Major Land Resource Areas (MLRAs), whose boundaries are shown in the figure.

Figure 2. Field-level data

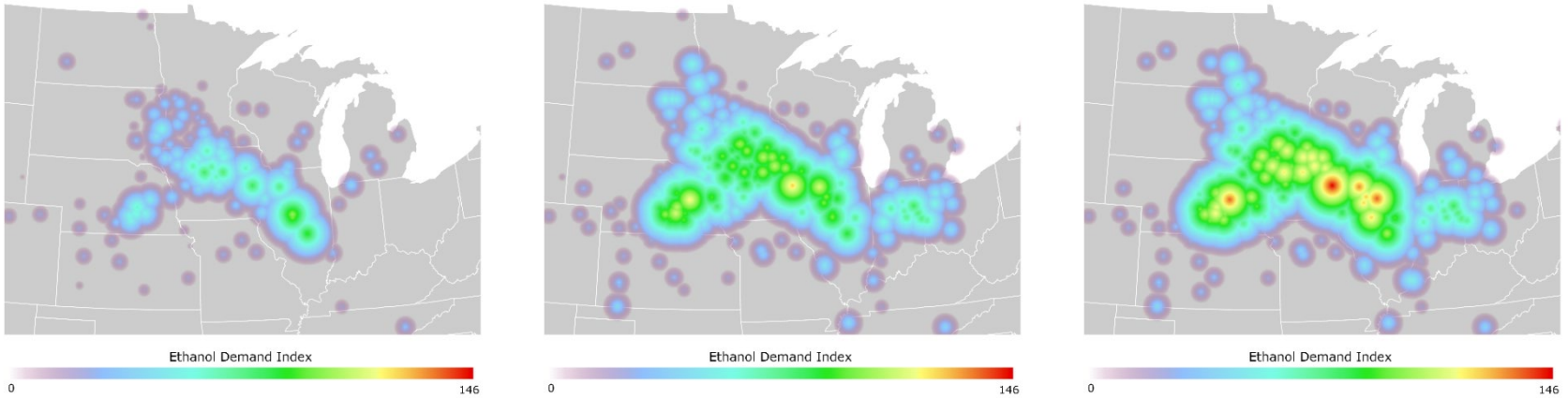


Note: This figure shows the universe of fields in our study area, colored by 2018 crop. Fields are defined by the 2018 USDA Crop Sequence Boundary dataset.

Figure 3. Ethanol demand index in selected years

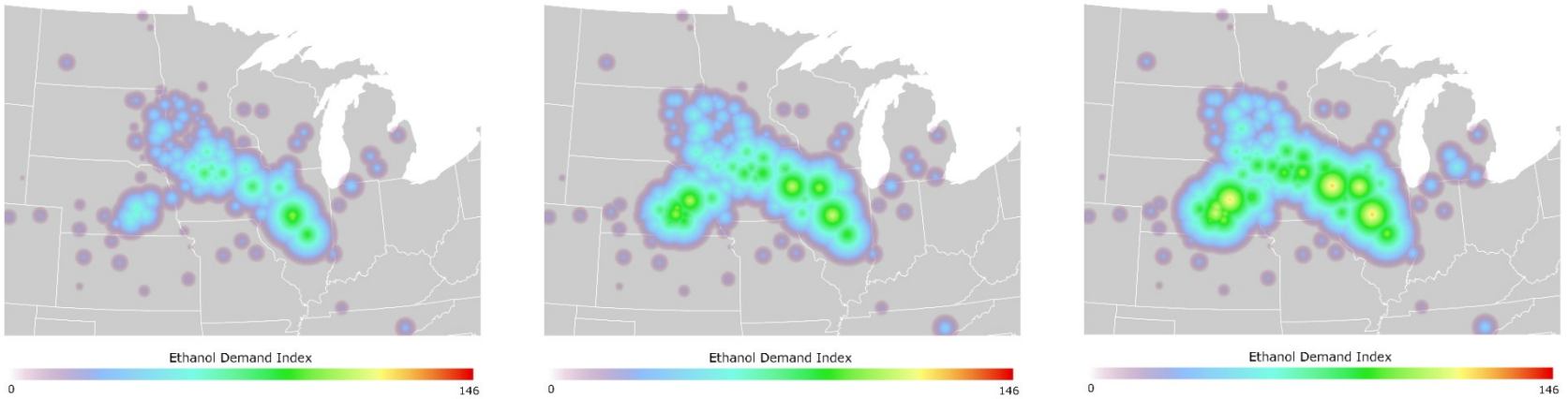
Actual ethanol history

2007 2012 2018



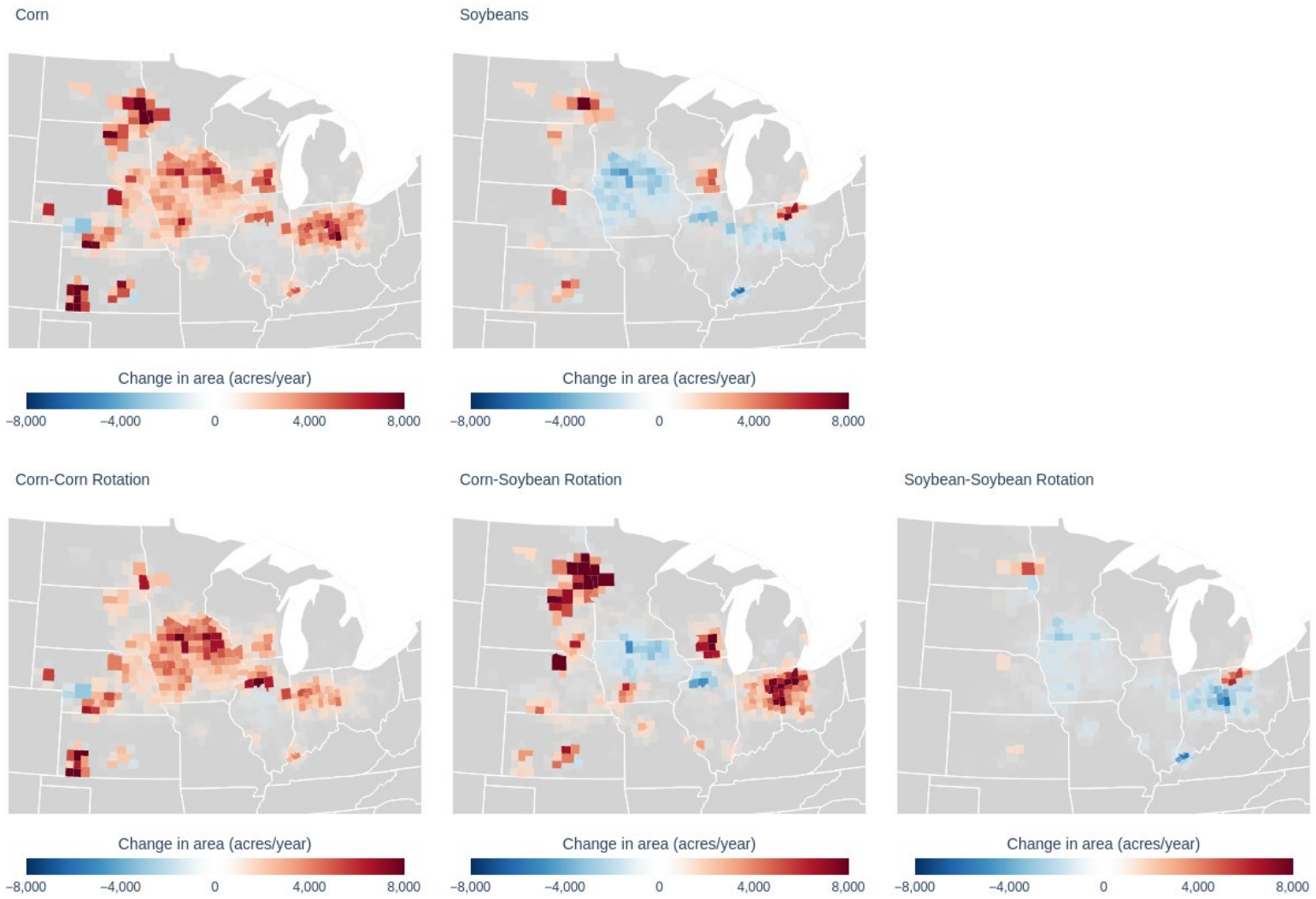
Counterfactual ethanol history

2007 2012 2018



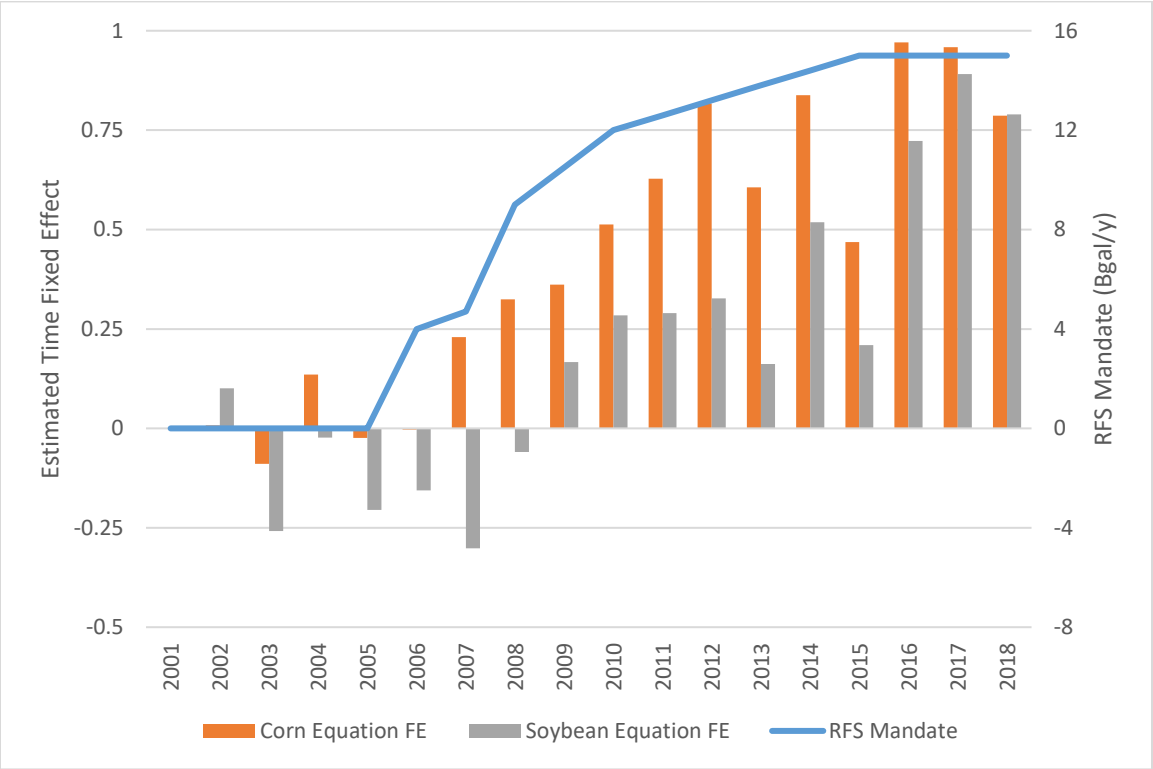
Note: This figure illustrates the geographic distribution of the ethanol demand index, for both the actual and counterfactual ethanol histories, in selected years.

Figure 4. County-level changes in crop pattern areas due to local ethanol capacity expansion



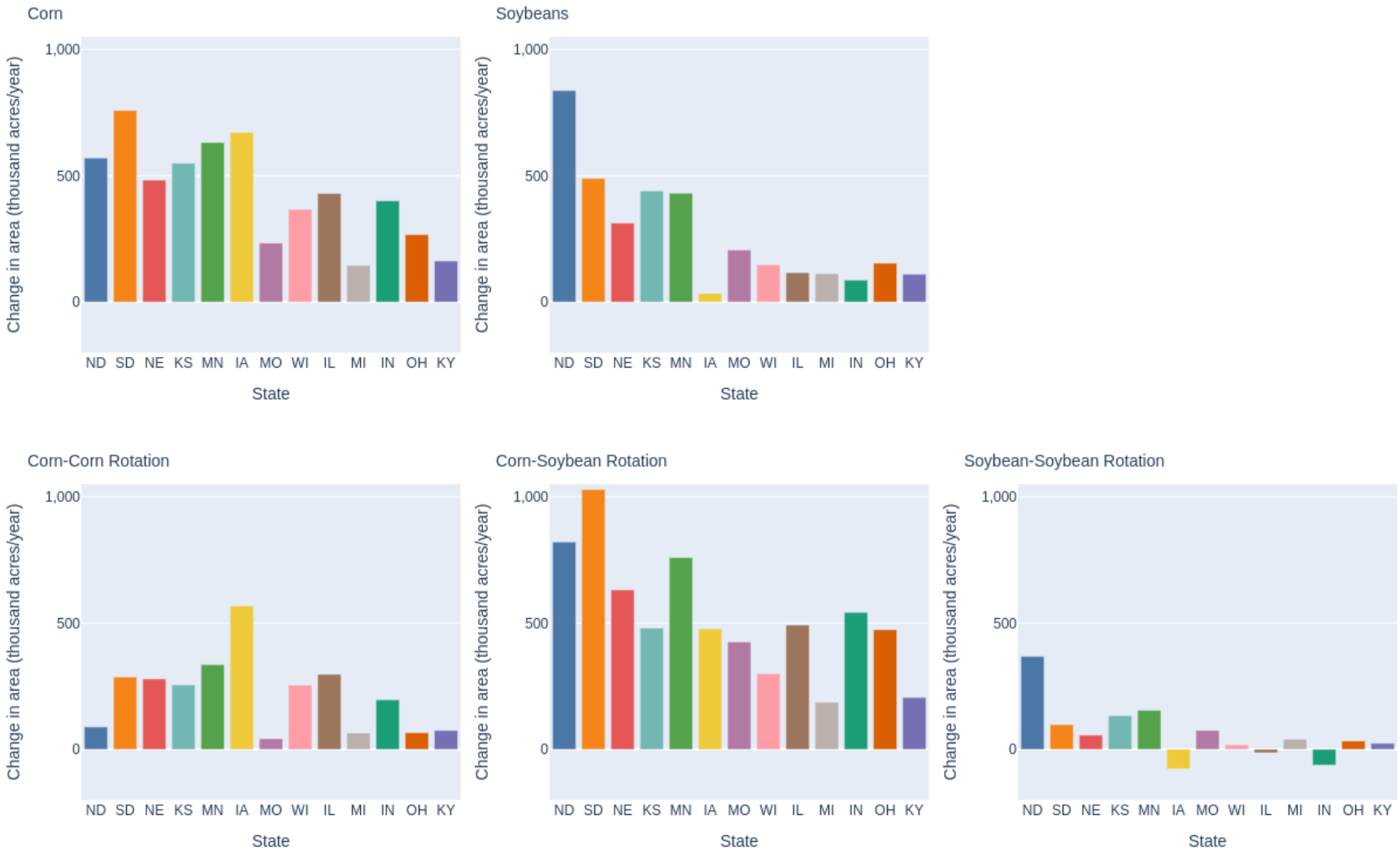
Note: This figure illustrates the impact of the RFS2-induced expansion of ethanol capacity in the Midwest on the areas of key crop patterns.

Figure 5. Estimated composite time fixed effects (region M) and the RFS mandate



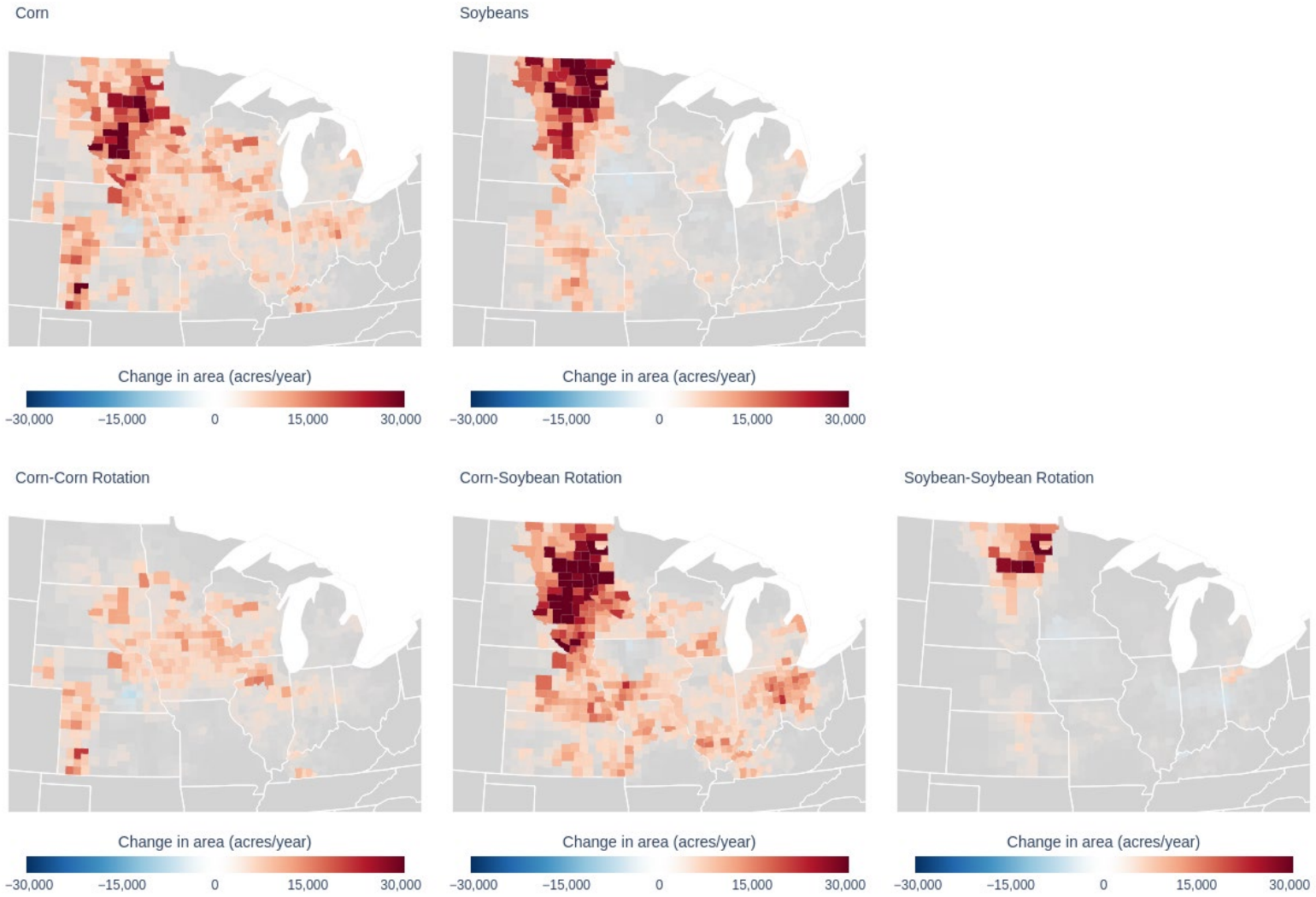
Note: This figure illustrates the estimated time fixed effects for LRR region M, for both corn and soybean equations (left axis), and the RFS mandate (right axis). The mandate corresponds to the RFS1 for years 2006 and 2007, and to the RFS2 for 2008 and later years. See Table 1 for more details.

Figure 6. State-level changes in crop-pattern areas due to RFS2



Note: This figure illustrates the RFS2's impact on the area of key crop patterns by state.

Figure 7. County-level changes in crop pattern areas due to RFS2



Note: This figure illustrates the RFS2's impact on the area of key crop patterns by county.

Table 1. Actual and counterfactual ethanol plants and capacity

	Actual		*Counterfactual*				
Year	N	Total Capacity	N	Total Capacity	RFS1 Mandate	RFS2 Mandate	USDA Projection[a]
2001	44	1,862	44	1,862			
2002	49	2,313	49	2,313			
2003	56	2,673	56	2,673			
2004	62	3,073	62	3,073			
2005	72	3,623	72	3,623			
2006	85	4,314	85	4,314	4,000		4,700
2007	102	5,509	101	5,474	4,700		5,600
2008	132	7,870	109	6,223	5,400	9,000	6,200
2009	154	10,182	113	6,813	6,100	10,500	6,700
2010	170	11,749	113	6,850	6,800	12,000	7,100
2011	180	12,977	113	7,367	7,400	12,600	7,400
2012	186	13,834	113	7,662	7,500	13,200	7,600
2013	170	12,512	113	7,745		13,800	7,700
2014	179	13,268	113	7,749		14,400	7,900
2015	185	14,076	113	7,771		15,000	8,100
2016	186	14,548	113	7,980		15,000	
2017	188	15,282	113	8,363		15,000	
2018	187	15,817	113	8,731		15,000	

Note: All capacity and volume data are in millions of gallons per year (Mgal/y). The values in the USDA Projection column are annual ethanol volumes projected in 2006, before the introduction of the 2007 Renewable Fuel Standard (RFS2). Source: Carter, Rausser, and Smith (2017) and https://usda.library.cornell.edu/concern/publications/qn59q396v). The close match between our counterfactual ethanol history and the USDA Projection gives us confidence that the counterfactual is a reasonable estimate of what would have happened without the RFS2.

Table 2. Field summary statistics

	Land Resource Region (LRR)							
	F	G	H	K	L	M	N	Total
Number of fields								
All	2,056,159	419,223	1,292,529	566,121	888,106	3,953,600	536,652	9,712,338
Large	848,018	155,938	594,691	199,174	407,994	1,961,520	156,045	4,323,380
Small	1,208,141	263,233	697,838	366,947	480,112	1,992,080	380,607	5,388,958
Average crop area, 2000-2018 (thousand acres)								
				All fields				
Corn	5,479	1,093	7,915	1,615	5,887	46,225	1,487	69,702
Soybeans	8,600	275	3,365	877	4,587	41,489	1,406	60,600
Other	25,855	4,824	18,807	3,864	4,419	14,911	3,212	75,892
Total	39,935	6,192	30,088	6,357	14,893	102,624	6,105	206,194
				Large fields				
Corn	4,940	933	7,274	1,242	5,062	42,977	1,261	63,688
Soybeans	7,762	240	3,019	690	3,996	38,322	1,147	55,177
Other	20,505	3,550	15,912	2,402	3,124	10,407	1,685	57,585
Total	33,207	4,722	26,206	4,334	12,182	91,705	4,093	176,450
				Small fields				
Corn	539	160	641	373	825	3,248	227	6,014
Soybeans	838	35	346	187	591	3,168	259	5,423
Other	5,350	1,274	2,896	1,462	1,295	4,504	1,526	18,307
Total	6,728	1,468	3,882	2,023	2,712	10,919	2,011	29,744

Note: This table presents summary statistics for the observed fields in the study. Summary statistics are broken down by Land Resource Region (LRR) and by field size. "Large" fields are those greater than or equal to 10 acres in size.

Table 3: Estimated parameters of the multinomial logit choice model

	F	G	H	K	L	M	N
Corn							
Ethanol demand index	0.00495	0.0104	0.0126	0.00175	0.00586	0.00445	0.00352
	(0.00278)	(0.00361)	(0.00278)	(0.00281)	(0.00236)	(0.00076)	(0.00272)
Previous crop							
Corn	0.535	1.165	1.381	1.942	1.558	1.495	0.786
	(0.104)	(0.204)	(0.137)	(0.097)	(0.203)	(0.141)	(0.486)
Soybeans	1.11	1.567	1.471	2.171	1.454	2.34	2.394
	(0.167)	(0.292)	(0.129)	(0.123)	(0.237)	(0.136)	(0.266)
Constant	-4.653	-1.459	-0.66	-3.163	-0.113	-2.613	-2.376
	(0.156)	(0.159)	(0.0674)	(0.146)	(0.188)	(0.119)	(0.172)
Soybeans							
Ethanol demand index	0.00263	0.0142	0.00822	0.00623	0.00688	0.000759	-0.00871
	(0.00256)	(0.00417)	(0.00233)	(0.00509)	(0.00272)	(0.00108)	(0.00396)
Previous crop							
Corn	1.774	2.226	1.947	2.682	2.678	3.131	2.537
	(0.181)	(0.203)	(0.15)	(0.127)	(0.211)	(0.122)	(0.27)
Soybeans	0.114	1.412	0.51	2.055	1.097	1.196	2.382
	(0.133)	(0.17)	(0.118)	(0.163)	(0.201)	(0.103)	(0.236)
Constant	-4.037	-3.663	-2.872	-2.069	-1.51	-1.011	-2.021
	(0.167)	(0.188)	(0.0867)	(0.124)	(0.109)	(0.0772)	(0.141)
Time fixed effects	34	34	34	34	34	34	34
Region fixed effects	88	74	88	110	68	314	116
Pseudo-R^2	0.21	0.18	0.23	0.19	0.15	0.25	0.21

Note: Results are estimated separately by Land Resource Region (LRR) and field size. This table reports the estimated parameters for fields greater than 15 acres in size. Standard errors are adjusted for clustering at the region (CRD × MLRA) level. The number of time fixed effects and region fixed effects estimated in the model and the total number of observations used for estimation are also presented.

Table 4. Hypotheses tests

	Land Resource Region (LRR)						
	F	G	H	K	L	M	N
Regional fixed effects	4.64E11	1.40E12	6.59E11	4.07E12	5.39E10	2.20E11	2.61E09
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Time fixed effects	79,239	502,112	179,066	140,466	3.25E08	8,488	60,155
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
State variables	328	346	180	1,215	487	2,327	551
(previous crop)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ethanol demand	3.36	13.65	22.85	1.50	9.65	48.73	16.38
index	(0.186)	(0.001)	(0.000)	(0.472)	(0.008)	(0.000)	(0.000)

Note: For each of the four sets of tests, the null hypothesis of the Wald test is that the affected parameters are zero. For example, for the regional fixed effects, $H_0: \gamma_k^r = 0, \forall r, k = C, S$. For each test, the top number reports the Wald test statistic, and the p-value of the test statistic is shown in parentheses below.

Table 5: Estimated change in crop areas from local ethanol plant impacts, 2008-2018.

	Baseline area	Counterfactual baseline area	Change in area	Percent change
Crop totals				
Corn	64.23	63.22	1.01	1.6%
	(0.157)	(0.215)	(0.129)	
Soybeans	55.63	55.82	-0.19	-0.3%
	(0.153)	(0.246)	(0.158)	
Other	56.60	57.42	-0.82	-1.4%
	(0.229)	(0.322)	(0.153)	
Crop sequences				
Corn-Corn	19.57	18.77	0.80	4.3%
	(0.235)	(0.257)	(0.125)	
Soy-Corn	36.51	36.24	0.27	0.7%
	(0.289)	(0.322)	(0.088)	
Other-Corn	8.14	8.20	-0.06	-0.7%
	(0.656)	(0.636)	(0.076)	
Corn-Soy	38.20	37.96	0.24	0.6%
	(0.250)	(0.278)	(0.081)	
Soy-Soy	11.05	11.31	-0.26	-2.3%
	(0.259)	(0.296)	(0.101)	
Other-Soy	6.38	6.55	-0.17	-2.6%
	(0.340)	(0.371)	(0.141)	
Corn-Other	6.46	6.49	-0.02	-0.4%
	(0.158)	(0.161)	(0.030)	
Soy-Other	8.06	8.27	-0.20	-2.5%
	(0.190)	(0.203)	(0.036)	
Other-Other	42.07	42.66	-0.59	-1.4%
	(0.717)	(0.704)	(0.072)	

Note: This table presents the average change in yearly crop totals and crop sequence areas over 2008-2018. All areas are in units of million acres. Two-year crop sequences are written as "previous crop-current crop". The baseline area is the average area under the actual history of ethanol expansion. The counterfactual baseline area shows what would have happened if ethanol expansion had followed our counterfactual trajectory, assuming that national price effects of the RFS2 still occurred as they did in reality.

Table 6: Regressions of time fixed effects on RFS mandate trajectory

	Land Resource Region (LRR)						
	F	G	H	K	L	M	N
Corn							
Constant	0.205	-0.605	-0.285	0.273	-0.082	-0.037	-0.265
	(0.075)	(0.150)	(0.079)	(0.124)	(0.050)	(0.046)	(0.142)
RFS Mandate	0.072	0.037	0.026	0.043	0.024	0.053	0.081
	(0.008)	(0.012)	(0.008)	(0.010)	(0.004)	(0.006)	(0.012)
R^2	0.78	0.49	0.45	0.63	0.69	0.85	0.81
Soybeans							
Constant	0.134	-0.933	0.059	-0.362	0.123	-0.182	-0.073
	(0.132)	(0.226)	(0.148)	(0.154)	(0.092)	(0.088)	(0.116)
RFS Mandate	0.064	0.034	0.050	0.028	0.022	0.044	0.069
	(0.013)	(0.019)	(0.013)	(0.013)	(0.008)	(0.009)	(0.013)
R^2	0.61	0.20	0.56	0.30	0.38	0.60	0.62

Note: This table presents the results of regressing the time fixed effects estimated in the multinomial logit choice model on the annual level of the RFS mandate.

Table 7: Estimated total changes in crop areas due to the RFS2, 2008-2018.

	Baseline area	Counterfactual baseline area	Change in area	Percent change
Crop totals				
Corn	64.23	58.55	5.68	9.7%
	(0.157)	(0.090)	(0.212)	
Soybeans	55.63	52.15	3.48	6.7%
	(0.153)	(0.077)	(0.209)	
Other	56.60	65.75	-9.15	-13.9%
	(0.229)	(0.103)	(0.301)	
Crop sequences				
Corn-Corn	19.57	16.76	2.81	16.8%
	(0.235)	(0.208)	(0.148)	
Soy-Corn	36.51	33.01	3.51	10.6%
	(0.289)	(0.285)	(0.111)	
Other-Corn	8.14	8.78	-0.64	-7.3%
	(0.656)	(0.624)	(0.109)	
Corn-Soy	38.20	34.87	3.32	9.5%
	(0.250)	(0.255)	(0.110)	
Soy-Soy	11.05	10.20	0.84	8.3%
	(0.259)	(0.216)	(0.119)	
Other-Soy	6.38	7.07	-0.69	-9.7%
	(0.340)	(0.288)	(0.158)	
Corn-Other	6.46	6.92	-0.45	-6.6%
	(0.158)	(0.183)	(0.055)	
Soy-Other	8.06	8.94	-0.88	-9.8%
	(0.190)	(0.214)	(0.062)	
Other-Other	42.07	49.89	-7.82	-15.7%
	(0.717)	(0.716)	(0.106)	

Note: This table presents the average change in yearly crop totals and crop sequence areas over 2008-2018. All areas are in units of million acres. Two-year crop sequences are written as "previous crop-current crop". The baseline area is the area under the actual history of ethanol expansion. The counterfactual baseline area shows what would have happened without the local and national impacts of the RFS2.

References

Abadie, A., Athey, S., Imbens, G.W. and Wooldridge, J.M., 2023. When should you adjust standard errors for clustering?. ***The Quarterly Journal of Economics***, 138(1): 1-35.

Arcidiacono, P., Foster, G., Goodpaster, N. and Kinsler, J., 2012. Estimating spillovers using panel data, with an application to the classroom. ***Quantitative Economics***, 3(3): 421-470.

Austin, K.G., J.P.H. Jones, and C.M. Clark. 2022. "A review of domestic land use change attributable to U.S. biofuel policy." ***Renewable and Sustainable Energy Reviews***, 159: 112181.

Billingsley, P., 1961. Statistical methods in Markov chains. ***The Annals of Mathematical Statistics***, pp. 12-40.

Brown, T.R., 2020. ***Biomass-Based Diesel: A Market and Performance Analysis***. Fuels Institute. https://www.fuelsinstitute.org/Research/Reports/Biomass-Based-Diesel-A-Market-and-Performance-Anal

Carter, C.A., G.C. Rausser and A. Smith. 2017. "Commodity Storage and the Market Effects of Biofuel Policies." ***American Journal of Agricultural Economics***, 99(4): 1027-1055.

Cui, J., H. Lapan, G. Moschini, and J. Cooper. 2011. "Welfare Impacts of Alternative Biofuel and Energy Policies." ***American Journal of Agricultural Economics***, 93(5): 1235-1256.

de Gorter, H., and D.R. Just. 2010. "The Social Costs and Benefits of Biofuels: The Intersection of Environmental, Energy and Agricultural Policy." ***Applied Economic Perspectives and Policy*** 32(1): 4-32.

de Gorter, H., D. Drabik, and D.R. Just. 2015. ***The Economics of Biofuel Policies: Impacts on Price Volatility in Grain and Oilseed Market***. NY: Palgrave-Macmillan.

Falconi, T.M.A., Kazemiparkouhi, F., Schwartz, B. and MacIntosh, D.L., 2022. Inconsistencies in domestic land use change study. ***Proceedings of the National Academy of Sciences***, 119(51), p. e2213961119.

Farrell, A.E., Plevin, R.J., Turner, B.T., Jones, A.D., O'hare, M. and Kammen, D.M., 2006. Ethanol can contribute to energy and environmental goals. ***Science***, 311(5760): 506-508.

Fatal, Y.S. and Thurman, W.N., 2014. The response of corn acreage to ethanol plant siting. ***Journal of Agricultural and Applied Economics***, 46(2): 157-171.

Gerveni, M., T. Hubbs and S. Irwin, 2023. "Overview of the Production Capacity of U.S. Biodiesel Plants." ***farmdoc daily*** (13):32, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, February 22.

Hendricks, N.P., A. Smith, and D.A. Sumner. 2014. "Crop Supply Dynamics and the Illusion of Partial Adjustment." ***American Journal of Agricultural Economics***, 96(5): 1469-1491.

Hunt, K.A., Abernethy, J., Beeson, P., Bowman, M., Wallander, S., and Williams, R., 2023. "Crop Sequence Boundaries (CSB): Delineated Fields Using Remotely Sensed Crop Rotations." USDA-NASS, Washington, D.C., USA. https://www.nass.usda.gov/Research_and_Science/Crop-Sequence-Boundaries/index.php

Kim, G., Nemati, M., Buck, S., Pates, N. and Mark, T., 2020. Recovering Forecast Distributions of Crop Composition: Method and Application to Kentucky Agriculture. ***Sustainability***, 12(7).

Lark, T.J., Hendricks, N.P., Smith, A., Pates, N., Spawn-Lee, S.A., Bougie, M., Booth, E.G., Kucharik, C.J. and Gibbs, H.K., 2022. Environmental outcomes of the US renewable fuel standard. ***Proceedings of the National Academy of Sciences***, 119(9), p. e2101084119.

Li, Y., Miao, R. and Khanna, M., 2019. Effects of ethanol plant proximity and crop prices on land-use change in the United States. ***American Journal of Agricultural Economics***, 101(2): 467-491.

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.

Moschini, G., J. Cui and H. Lapan. 2012. "Economics of Biofuels: An Overview of Policies, Impacts and Prospects." ***Bio-based and Applied Economics*** 1(3): 269-296.

Moschini, G., Lapan, H. and Kim, H., 2017. The renewable fuel standard in competitive equilibrium: Market and welfare effects. ***American Journal of Agricultural Economics***, 99(5): 1117-1142.

Motamed, M., McPhail, L. and Williams, R., 2016. Corn area response to local ethanol markets in the United States: A grid cell level analysis. ***American Journal of Agricultural Economics***, 98(3): 726-743.

Nevo, A., 2001. Measuring market power in the ready-to-eat cereal industry. ***Econometrica***, 69(2): 307-342.

Norris, J.R., 1998. ***Markov chains***. Cambridge University Press.

Nebraska Department of Environment and Energy. 2018. Ethanol Facilities Capacity by State and Plant. Accessed April 29, 2022. https://neo.ne.gov/programs/stats/inf/122_archive.htm

Pates, N.J. and Hendricks, N.P., 2021. Fields from Afar: Evidence of Heterogeneity in United States Corn Rotational Response from Remote Sensing Data. ***American Journal of Agricultural Economics***, 103(5): 1759-1782.

Paton, L., Troffaes, M., Boatman, N., Hussein, M. and Hart, A., 2014, July. Multinomial logistic regression on Markov chains for crop rotation modelling. In ***International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*** (pp. 476-485). Springer, Cham.

Quoss, V., T.H. Savage, D.L. Sunding, J. Terhors, 2011. Forecasting land use from estimated Markov transitions. Available at SSRN: http://dx.doi.org/10.2139/ssrn.1866003

Renewable Fuels Association. 2022. ***Annual Industry Outlook***. Accessed April 29, 2022. https://ethanolrfa.org/resources/annual-industry-outlook

Roberts, Michael J. and W. Schlenker. 2013. Identifying supply and demand elasticities of agricultural commodities: Implications for the US ethanol mandate. American Economic Review, 103(6): 2265-2295.

Schnepf, R. and B.D. Yacobucci. 2013. ***Renewable Fuel Standard (RFS): Overview and Issues***. Congressional Research Services Report No. R40155, October.

Taheripour, F., Mueller, S., Kwon, H., Khanna, M., Emery, I., Copenhaver, K., and Wang, M. 2022. "Comments on 'Environmental Outcomes of the US Renewable Fuel Standard'." https://greet.es.anl.gov/publication-comment_environ_outcomes_us_rfs , March 21.

Train, K.E., 2009. ***Discrete choice methods with simulation***, 2nd edition. Cambridge University Press.

U.S. Environmental Protection Agency. 2021. ***Renewable Fuel Standard Program***. December 21. Accessed May 10, 2022. https://www.epa.gov/renewable-fuel-standard-program

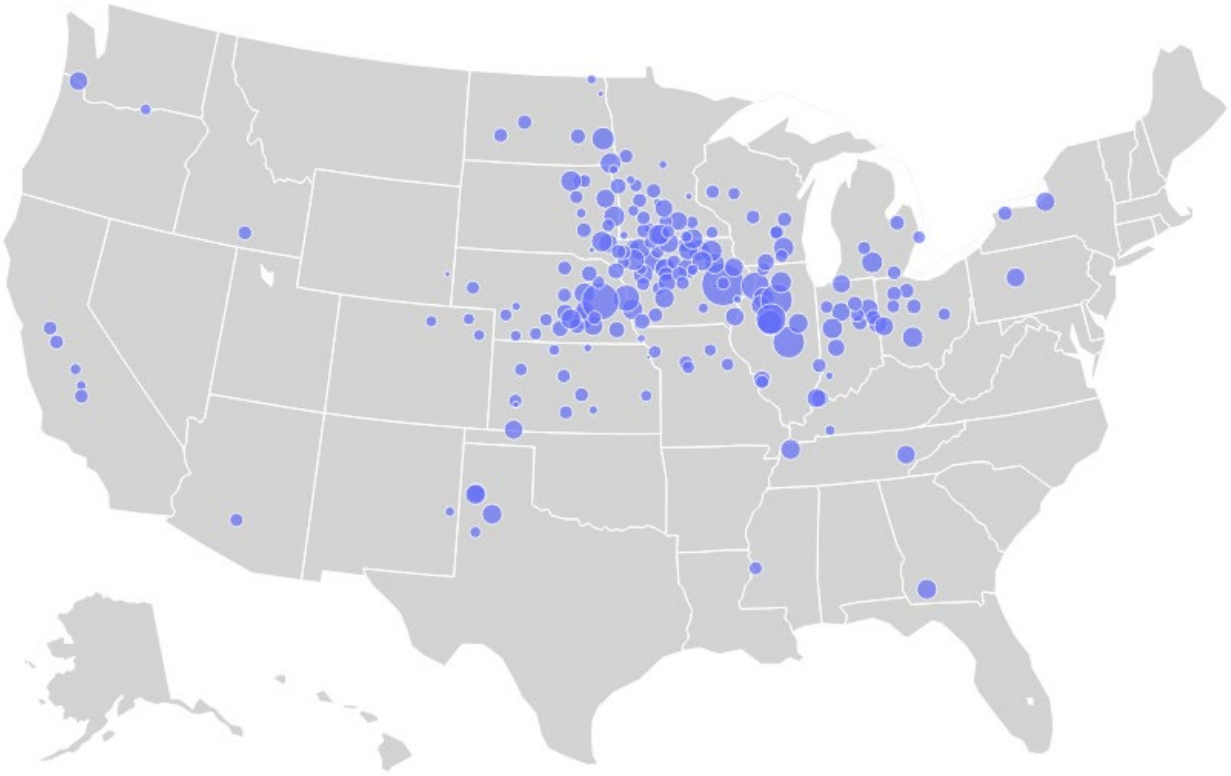
Wang, S., Di Tommaso, S., Deines, J.M. and Lobell, D.B., 2020. Mapping twenty years of corn and soybean across the US Midwest using the Landsat archive. ***Scientific Data***, 7(1): 307.

Wang, Y., Delgado, M.S., Sesmero, J. and Gramig, B.M., 2020. Market structure and the local effects of ethanol expansion on land allocation: a spatially explicit analysis. ***American Journal of Agricultural Economics***, 102(5): 1598-1622.

SUPPLEMENTARY APPENDIX

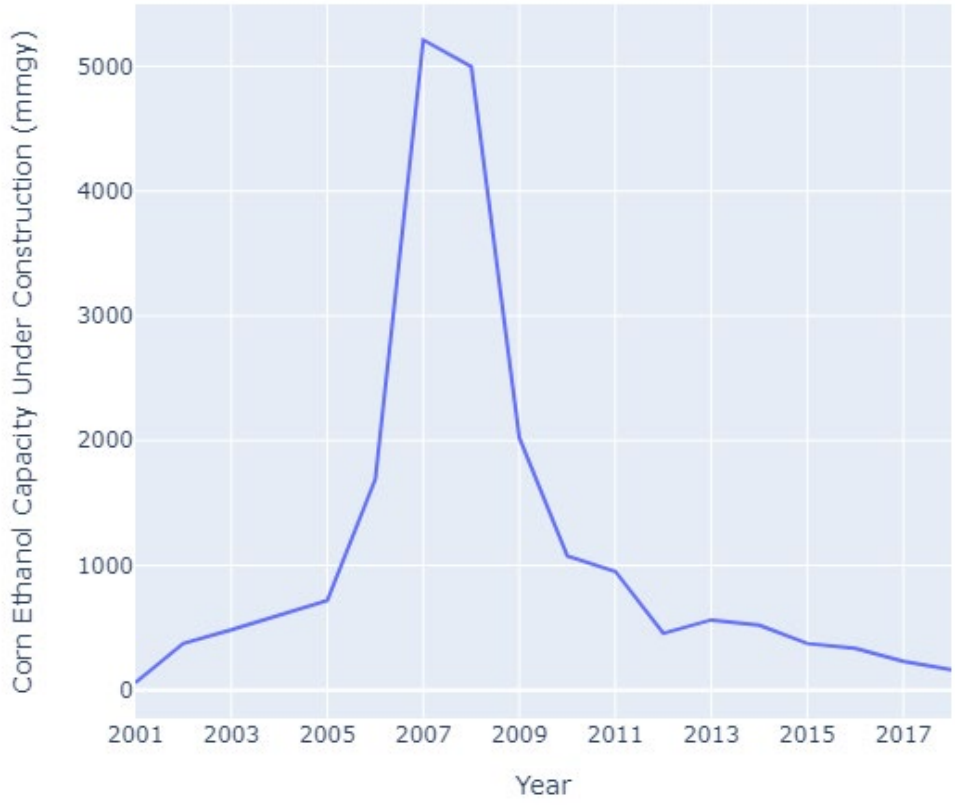
Appendix A. Ethanol Plants and the Renewable Fuel Standard

Figure A1. Ethanol plant locations



Note: This figure shows the locations of the 203 corn ethanol plants in our dataset. Circle sizes are proportional to the plant's highest annual capacity over the sample period.

Figure A2. Corn ethanol capacity under construction by year



Note: This figure illustrates the steep rise in ethanol capacity under construction in 2007 and 2008. We interpret this surge in construction as evidence that the response to the 2007 Renewable Fuel Standard began in 2007 in anticipation of its passage.

Table A1: Local ethanol demand index summary statistics

	Actual ethanol history				Counterfactual ethanol history			
Year	Mean	Std. Dev.	Max	Pct 0	Mean	Std. Dev.	Max	Pct 0
2001	3.1	9.8	84.2	85%				
2002	4.2	12.0	93.7	81%				
2003	4.6	12.2	93.7	78%				
2004	5.2	12.5	93.7	75%				
2005	6.0	13.2	93.9	71%				
2006	7.0	13.9	93.9	68%				
2007	9.1	16.5	98.5	63%				
2008	13.0	18.8	98.7	51%	10.5	17.7	98.7	59%
2009	17.1	22.6	114.0	44%	11.7	18.8	102.8	57%
2010	20.3	25.1	115.5	42%	11.8	18.9	103.0	57%
2011	22.7	27.4	123.1	39%	14.3	22.4	108.6	55%
2012	24.2	28.2	123.8	37%	15.1	23.3	108.6	54%
2013	22.7	27.2	123.7	39%	15.3	23.7	112.0	54%
2014	23.4	27.5	123.7	37%	15.4	23.7	112.0	54%
2015	25.0	28.9	124.4	36%	15.4	23.7	112.0	54%
2016	25.7	29.0	125.3	34%	15.9	24.0	113.1	53%
2017	28.4	32.3	140.6	34%	17.8	26.6	122.5	52%
2018	30.4	34.0	145.2	33%	19.0	27.6	124.0	51%

Note: This table presents summary statistics of the ethanol demand index, for both the factual and counterfactual ethanol histories, for the fields used in the analysis of the study. Pct 0 refers to the percent of fields whose ethanol index is 0 in each year. Counterfactual ethanol history begins in 2008 because that is when the two histories diverge.

Appendix B. Ethanol demand index procedure

To compute the independent procurement radius of each plant in isolation, we need to calibrate the value of κ in $Z_{ij} = \max\left\{\kappa\sqrt{K_j} - d_{ij}, 0\right\}$. If we assume that corn is produced uniformly in the surrounding area with density α bushels per unit area, then we can write $\kappa = (\alpha\rho\pi)^{1/2}$, where ρ is the conversion rate of corn into ethanol. We measure the value of α in bushels per square mile using county-level corn production data from the 2002 USDA Census of Agriculture and the land area of each county.[11] We use the industry standard value of $\rho = 2.7$ gallons per bushel as the conversion rate of corn into ethanol. Given the units of measurement for α and ρ, and measuring plants' capacities K_{jt} in gallons, the plant's procurement radius that we retrieve as $\kappa\sqrt{K_{jt}}$ is measured in miles.

The foregoing procedure yields a mean value of $\kappa = 0.0032$ for counties with significant corn production (at least 10,000 acres). For a typical 100 million gallon per year (Mgal/y) ethanol plant, this value of κ implies a procurement radius of 32 miles. Because this computation is derived from a highly stylized model, we supplement this procedure with information from Valero Renewable Fuels. Valero operates 12 corn ethanol plants with an average capacity of 132 Mgal/y. On the company's website, Valero reports the capacity and a rough procurement radius (e.g., "within 40-50 miles") for each plant.[12] The figures provided by Valero imply a value of around $\kappa = 0.0043$, or a 43 mile procurement radius for a 100 Mgal/y plant.

Using the combined evidence of our computation of county-level corn densities and Valero's rough procurement radii, we settle on a value of $\kappa = 0.004$ for our calibration, implying a 40-mile procurement radius for a 100 Mgal/y plant. We use this value to compute the independent procurement radius for each plant as $d_{jt}^0 = \kappa\sqrt{K_{jt}}$.

[11] We use the 2002 Census of Agriculture to capture the corn density before the impact of the ethanol boom.

[12] See https://www.valero.com/about/locations?location_type_id=2. Accessed 9 Jan 2023.

Next, we identify the set of plants with overlapping procurement areas for each plant: $\mathcal{J}^{\text{near}} = \left\{ k \in \mathcal{J} \mid \tilde{d}_{jk} < d_{jt}^0 + d_{kt}^0 \right\}$. Note that plant j itself is included in this set (that is, definitionally, each plant overlaps with itself). We can now compute each plant's procurement radius that accounts for any interference from nearby plants, $\bar{d}_{jt}$. We do this by first computing an "effective capacity" for each plant, defined as:

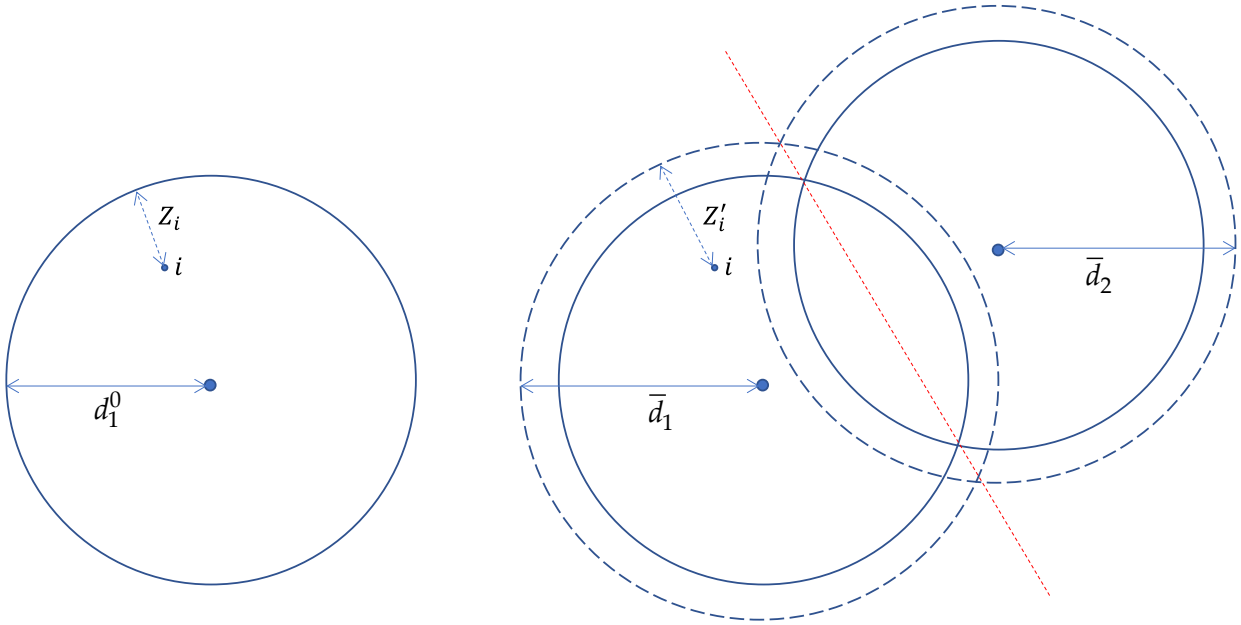
$$\bar{K}_{jt} = \sum_{k \in \mathcal{J}^{\text{near}}} \left(1 - \frac{\tilde{d}_{jk}}{d_{jt}^0 + d_{kt}^0} \right) K_{kt} \tag{16}$$

The effective capacity of plant j is essentially a sum of all plants with overlapping procurement radii (including plant j itself), weighted by the plant's distance from j. This definition of effective capacity has some intuitive properties. First, since $\tilde{d}_{jj} = 0$, j's own capacity K_{jt} enters the sum with a weight of 1. Second, as the distance between two plants j and k approaches $d_{jt}^0 + d_{kt}^0$, the weight of k's capacity on j's effective capacity (and vice versa) approaches 0, which aligns with the expectation that the two plants should not impact one another if their procurement areas do not overlap.

Finally, we use the effective capacities for each plant in each year to compute each plant's effective procurement radius accounting for interference from nearby plants, $\bar{d}_{jt} \equiv \kappa \bar{K}_{jt}$.

Figure B1 illustrates the (field-specific) ethanol demand index we have constructed. In Panel A there is a single ethanol plant, with a procurement radius d_1^0. The ethanol demand pressure Z_i experience by the parcel of land i is determined by its location's distance from the boundary of the procurement area. Panel B shows the implications of another plant's entry. The effective procurement radius of plant 1 expands to $\bar{d}_1$, and the ethanol demand pressure at location i increases to Z_i' (despite the fact that this location still supplies plant 1 and the distance between plant 1 and the parcel has not changed).

Figure B1. Ethanol demand pressure from an isolated plant and from competing plants



Appendix C. Supplementary results including small fields

This section presents results for the regressions, hypothesis tests, and predicted changes in area for fields less than 10 acres in size, which were excluded from the main analysis due to potential measurement error issues.

Table C1: Estimated parameters of the multinomial logit choice model, small fields.

	F	G	H	K	L	M	N
Corn							
Ethanol demand index	0.00098	0.00725	0.00837	0.00312	0.00388	0.00124	-0.00014
	(0.00264)	(0.00334)	(0.00262)	(0.00277)	(0.00197)	(0.0006)	(0.0022)
Previous crop							
Corn	0.845	1.178	1.385	1.963	1.848	1.805	1.489
	(0.121)	(0.183)	(0.109)	(0.0813)	(0.107)	(0.0631)	(0.27)
Soybeans	1.159	1.42	1.337	2.066	1.818	2.361	2.452
	(0.141)	(0.26)	(0.115)	(0.106)	(0.139)	(0.0729)	(0.167)
Constant	-4.693	-2.469	-1.127	-3.302	-1.103	-2.668	-2.984
	(0.143)	(0.127)	(0.0487)	(0.133)	(0.101)	(0.0661)	(0.113)
Soybeans							
Ethanol demand index	-0.00063	0.0112	0.0048	0.0096	0.00516	-0.00203	-0.0109
	(0.00218)	(0.00409)	(0.00223)	(0.00501)	(0.00245)	(0.00098)	(0.00318)
Previous crop							
Corn	1.632	1.953	1.686	2.429	2.493	2.793	2.535
	(0.15)	(0.178)	(0.133)	(0.11)	(0.123)	(0.0626)	(0.167)
Soybeans	0.468	1.506	0.835	2.139	1.658	1.653	2.41
	(0.124)	(0.153)	(0.113)	(0.138)	(0.141)	(0.0617)	(0.158)
Constant	-4.179	-4.598	-3.131	-2.31	-2.191	-1.426	-2.392
	(0.158)	(0.148)	(0.0714)	(0.117)	(0.097)	(0.0525)	(0.0835)
Time fixed effects	34	34	34	34	34	34	34
Region fixed effects	88	76	88	112	70	316	116
Pseudo-R^2	0.18	0.15	0.18	0.18	0.16	0.20	0.22

Note: Compare to Table 3 in the main text. Results are estimated separately by Land Resource Region (LRR) and field size. This table reports the estimated parameters for fields less than 10 acres in size. Standard errors are adjusted for clustering at the region (CRD × MLRA) level.

Table C2: Semi-elasticities of Markov transition probabilities

	Land Resource Region (LRR)						
Markov transition	F	G	H	K	L	M	N
Large fields							
Corn-Corn	0.0030	0.0019	0.0215	-0.0009	0.0008	0.0226	0.0043
Corn-Soy	0.0004	0.0031	-0.0048	0.0021	0.0058	-0.0189	-0.0046
Corn-Other	-0.0034	-0.0049	-0.0168	-0.0011	-0.0067	-0.0038	0.0003
Soy-Corn	0.0070	0.0030	0.0244	-0.0004	0.0069	0.0175	0.0049
Soy-Soy	-0.0005	0.0013	-0.0024	0.0013	0.0045	-0.0098	-0.0045
Soy-Other	-0.0065	-0.0043	-0.0220	-0.0009	-0.0114	-0.0077	-0.0004
Other-Corn	0.0061	0.0054	0.0296	0.0005	0.0114	0.0338	0.0044
Other-Soy	0.0013	0.0018	0.0019	0.0009	0.0078	-0.0090	-0.0029
Other-Other	-0.0074	-0.0072	-0.0316	-0.0015	-0.0193	-0.0248	-0.0015
Small fields							
Corn-Corn	0.0011	0.0015	0.0127	-0.0008	0.0006	0.0152	0.0021
Corn-Soy	-0.0013	0.0022	-0.0008	0.0025	0.0058	-0.0168	-0.0034
Corn-Other	0.0002	-0.0037	-0.0119	-0.0017	-0.0063	0.0016	0.0012
Soy-Corn	0.0013	0.0020	0.0147	-0.0005	0.0030	0.0128	0.0025
Soy-Soy	-0.0010	0.0015	-0.0005	0.0021	0.0046	-0.0114	-0.0031
Soy-Other	-0.0002	-0.0034	-0.0142	-0.0016	-0.0076	-0.0014	0.0006
Other-Corn	0.0009	0.0027	0.0129	0.0006	0.0059	0.0086	0.0006
Other-Soy	-0.0009	0.0012	0.0021	0.0010	0.0052	-0.0082	-0.0020
Other-Other	0.0000	-0.0039	-0.0150	-0.0016	-0.0111	-0.0004	0.0014

Note: This table shows the semi-elasticities of each Markov transition probability with respect to the ethanol demand index, i.e., $\partial P_{it}^{xy} \big/ \partial \ln Z_{it}$, where x denotes the previous crop and y denotes the current crop choice.

Table C3. Hypotheses tests, small fields.

	Land Resource Region (LRR)						
	F	G	H	K	L	M	N
Regional fixed effects	1.90E12	1.10E10	3.20E11	9.30E11	1.60E10	5.80E11	1.50E09
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Time fixed effects	35,163	1.00E06	106,969	234,062	6.80E07	8,142	42,295
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
State variables	297	446	220	1,272	483	2,980	595
(previous crop)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ethanol demand	0.70	8.69	10.43	3.80	7.00	22.17	16.11
index	(0.706)	(0.013)	(0.005)	(0.149)	(0.030)	(0.000)	(0.000)

Note: Compare to Table 4 in the main text. This table presents the results of hypothesis tests for fields less than 10 acres in size. For each of the four sets of tests, the null hypothesis of the Wald test is that the affected parameters are zero. For example, for the regional fixed effects, $H_0: \gamma_k^r = 0, \forall r, k = C, S$. For each test, the top number reports the Wald test statistic, and the p-value of the test statistic is shown in parentheses below.

Table C4: Estimated change in crop areas from local ethanol impacts, 2008-2018, small fields.

	Baseline area	Counterfactual baseline area	Change in area	Percent change
Crop totals				
Corn	6.92	6.83	0.08	1.2%
	(0.023)	(0.035)	(0.019)	
Soybeans	6.19	6.23	-0.04	-0.7%
	(0.021)	(0.033)	(0.021)	
Other	16.64	16.68	-0.04	-0.3%
	(0.031)	(0.043)	(0.024)	
Crop sequences				
Corn-Corn	2.37	2.30	0.07	3.0%
	(0.025)	(0.032)	(0.015)	
Soy-Corn	3.00	3.00	0.01	0.2%
	(0.028)	(0.03)	(0.009)	
Other-Corn	1.54	1.53	0.01	0.6%
	(0.025)	(0.025)	(0.004)	
Corn-Soy	3.19	3.18	0.01	0.2%
	(0.026)	(0.028)	(0.009)	
Soy-Soy	1.77	1.80	-0.04	-2.0%
	(0.025)	(0.03)	(0.014)	
Other-Soy	1.24	1.25	-0.01	-0.8%
	(0.021)	(0.023)	(0.004)	
Corn-Other	1.36	1.35	0.01	0.7%
	(0.021)	(0.021)	(0.004)	
Soy-Other	1.42	1.43	-0.01	-0.7%
	(0.025)	(0.026)	(0.004)	
Other-Other	13.86	13.90	-0.04	-0.3%
	(0.043)	(0.048)	(0.019)	

Note: Compare to Table 5 in the main text. This table presents the average change in yearly crop totals and crop sequence areas over 2008-2018 for fields less than 10 acres in size. All areas are in units of million acres. Two-year crop sequences are written as "previous crop-current crop". The baseline area is the average area under the actual history of ethanol expansion. The counterfactual baseline area shows what would have happened if ethanol expansion had followed our counterfactual trajectory, assuming that national price effects of the RFS2 still occurred as they did in reality.

Table C5: Estimated change in crop areas due to the RFS2, 2008-2018, small fields.

	Baseline area	Counterfactual baseline area	Change in area	Percent change
Crop totals				
Corns	6.92	5.78	1.14	19.6%
	(0.023)	(0.014)	(0.033)	
Soybeans	6.19	5.29	0.90	17.1%
	(0.021)	(0.012)	(0.03)	
Other	16.64	18.67	-2.04	-10.9%
	(0.031)	(0.019)	(0.043)	
Crop sequences				
Corn-Corn	2.37	1.83	0.54	29.4%
	(0.025)	(0.019)	(0.021)	
Soy-Corn	3.00	2.41	0.59	24.5%
	(0.028)	(0.026)	(0.012)	
Other-Corn	1.54	1.54	0.01	0.5%
	(0.025)	(0.023)	(0.01)	
Corn-Soy	3.19	2.59	0.59	22.9%
	(0.026)	(0.023)	(0.012)	
Soy-Soy	1.77	1.43	0.33	23.4%
	(0.025)	(0.017)	(0.018)	
Other-Soy	1.24	1.26	-0.03	-2.0%
	(0.021)	(0.02)	(0.008)	
Corn-Other	1.36	1.35	0.00	0.4%
	(0.021)	(0.02)	(0.009)	
Soy-Other	1.42	1.44	-0.02	-1.6%
	(0.025)	(0.024)	(0.008)	
Other-Other	13.86	15.88	-2.02	-12.7%
	(0.043)	(0.035)	(0.046)	

Note: Compare to Table 7 in the main text. This table presents the average change in yearly crop totals and crop sequence areas over 2008-2018 for fields less than 10 acres in size. All areas are in units of million acres. Two-year crop sequences are written as "previous crop-current crop". The baseline area is the area under the actual history of ethanol expansion. The counterfactual baseline area shows what would have happened without the local and national impacts of the RFS2.

The following tables present results for the predicted changes in area for small (<10 acres) and large (>=10 acres) fields combined.

Table C6: Estimated change in crop areas from local ethanol impacts, 2008-2018, all fields.

	Baseline area	Counterfactual baseline area	Change in area	Percent change
Crop totals				
Corn	71.14	70.05	1.09	1.6%
	(0.159)	(0.218)	(0.13)	
Soybeans	61.82	62.05	-0.23	-0.4%
	(0.154)	(0.248)	(0.159)	
Other	73.24	74.1	-0.86	-1.2%
	(0.231)	(0.325)	(0.155)	
Crop sequences				
Corn-Corn	21.94	21.07	0.87	4.1%
	(0.236)	(0.259)	(0.126)	
Soy-Corn	39.51	39.24	0.28	0.7%
	(0.29)	(0.323)	(0.088)	
Other-Corn	9.69	9.74	-0.05	-0.5%
	(0.656)	(0.637)	(0.076)	
Corn-Soy	41.38	41.14	0.24	0.6%
	(0.251)	(0.279)	(0.082)	
Soy-Soy	12.82	13.11	-0.3	-2.3%
	(0.26)	(0.298)	(0.102)	
Other-Soy	7.62	7.8	-0.18	-2.3%
	(0.341)	(0.372)	(0.141)	
Corn-Other	7.82	7.84	-0.01	-0.2%
	(0.159)	(0.162)	(0.03)	
Soy-Other	9.49	9.7	-0.21	-2.2%
	(0.192)	(0.205)	(0.036)	
Other-Other	55.93	56.56	-0.63	-1.1%
	(0.718)	(0.706)	(0.075)	

Note: Compare to Table 5 in the main text. This table presents the average change in yearly crop totals and crop sequence areas over 2008-2018 all fields in the study area. All areas are in units of million acres. Two-year crop sequences are written as "previous crop-current crop". The baseline area is the average area under the actual history of ethanol expansion. The counterfactual baseline area shows what would have happened if ethanol expansion had followed our counterfactual trajectory, assuming that national price effects of the RFS2 still occurred as they did in reality.

Table C7: Estimated total change in crop areas due to the RFS2, 2008-2018, all fields.

	Baseline area	Counterfactual baseline area	Change in area	Percent change
Crop totals				
Corns	71.14	64.33	6.81	10.6%
	(0.159)	(0.091)	(0.215)	
Soybeans	61.82	57.44	4.38	7.6%
	(0.154)	(0.078)	(0.211)	
Other	73.24	84.42	-11.19	-13.2%
	(0.231)	(0.105)	(0.304)	
Crop sequences				
Corn-Corn	21.94	18.59	3.35	18.0%
	(0.236)	(0.209)	(0.149)	
Soy-Corn	39.51	35.42	4.1	11.6%
	(0.29)	(0.286)	(0.112)	
Other-Corn	9.69	10.32	-0.63	-6.1%
	(0.656)	(0.624)	(0.109)	
Corn-Soy	41.38	37.47	3.91	10.4%
	(0.251)	(0.256)	(0.111)	
Soy-Soy	12.82	11.64	1.18	10.1%
	(0.26)	(0.217)	(0.12)	
Other-Soy	7.62	8.33	-0.71	-8.6%
	(0.341)	(0.289)	(0.158)	
Corn-Other	7.82	8.27	-0.45	-5.4%
	(0.159)	(0.184)	(0.056)	
Soy-Other	9.49	10.38	-0.9	-8.6%
	(0.192)	(0.215)	(0.063)	
Other-Other	55.93	65.77	-9.84	-15.0%
	(0.718)	(0.717)	(0.116)	

Note: Compare to Table 7 in the main text. This table presents the average change in yearly crop totals and crop sequence areas over 2008-2018 for all fields in the study area. All areas are in units of million acres. Two-year crop sequences are written as "previous crop-current crop". The baseline area is the area under the actual history of ethanol expansion. The counterfactual baseline area shows what would have happened without the local and national impacts of the RFS2.

Appendix D. Estimated time fixed effects vs. RFS mandate level

This section presents the counterparts of Figure 5 in the main text for all seven LRRs in the study area. The graph for LRR=M is included here, in addition to the main text, for ease of comparison.

Figure D1: Estimated time fixed effects vs. RFS Mandate, LRR=F

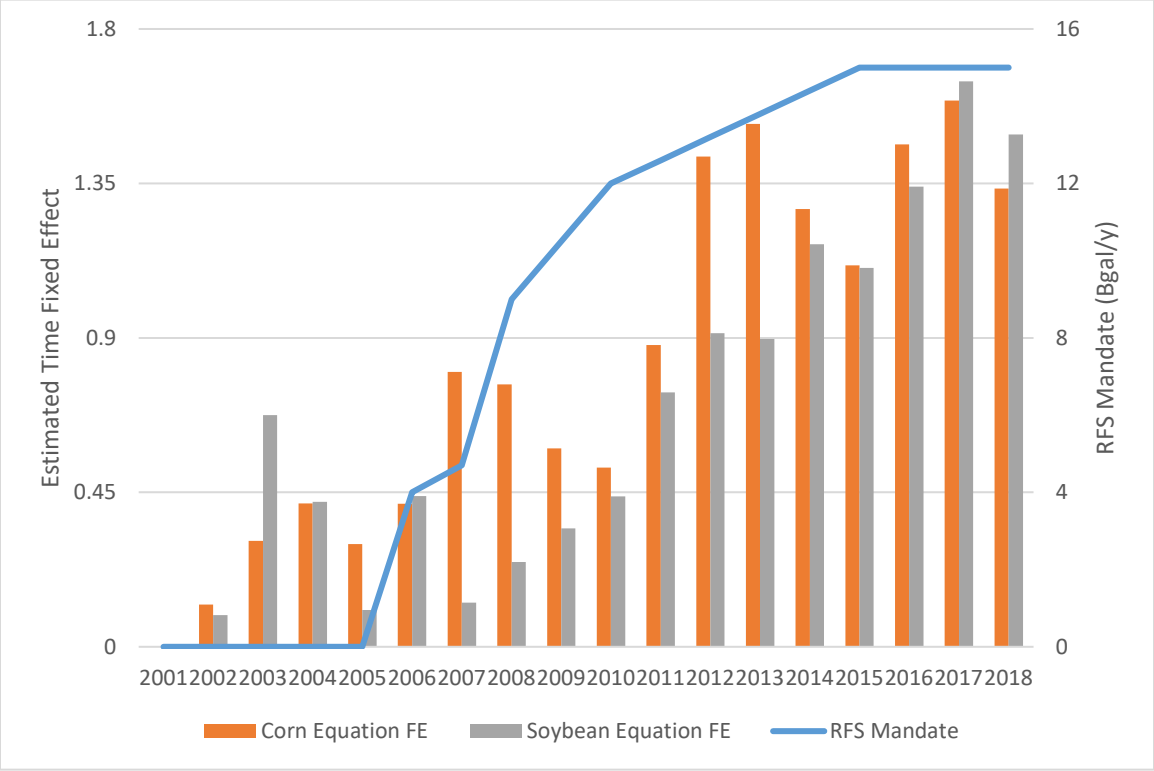


Figure D2: Estimated time fixed effects vs. RFS Mandate, LRR=G

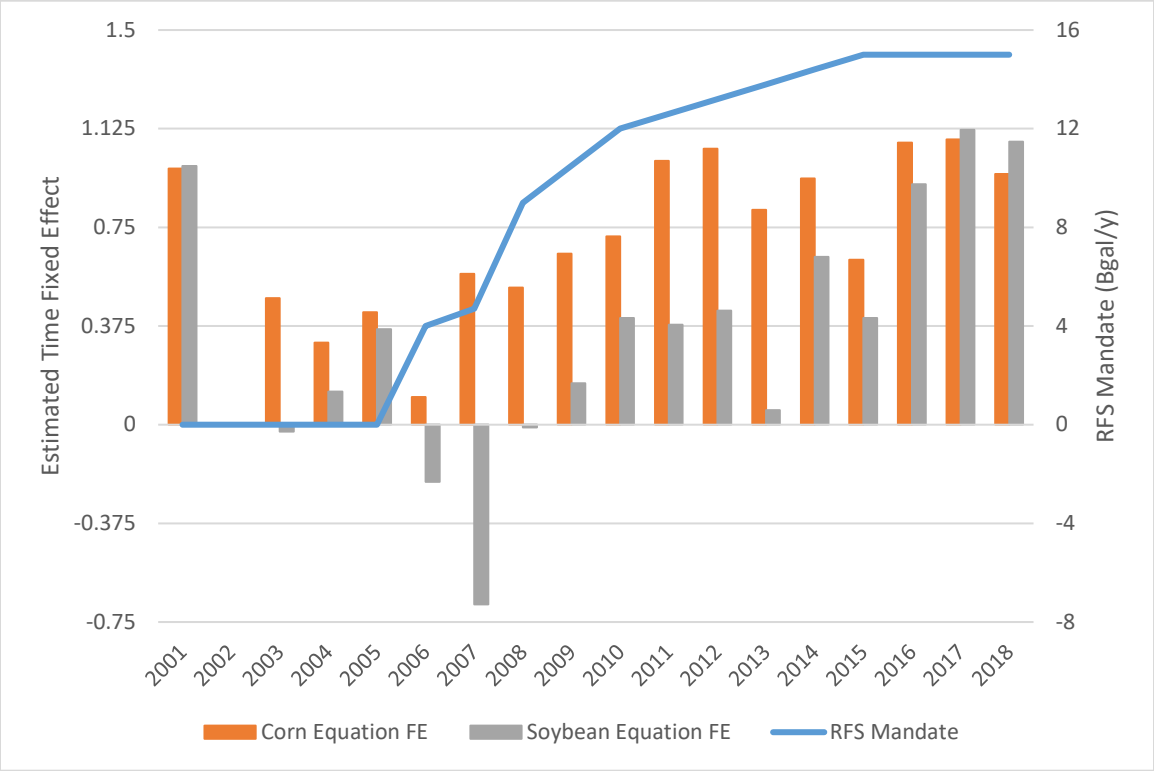


Figure D3: Estimated time fixed effects vs. RFS Mandate, LRR=H

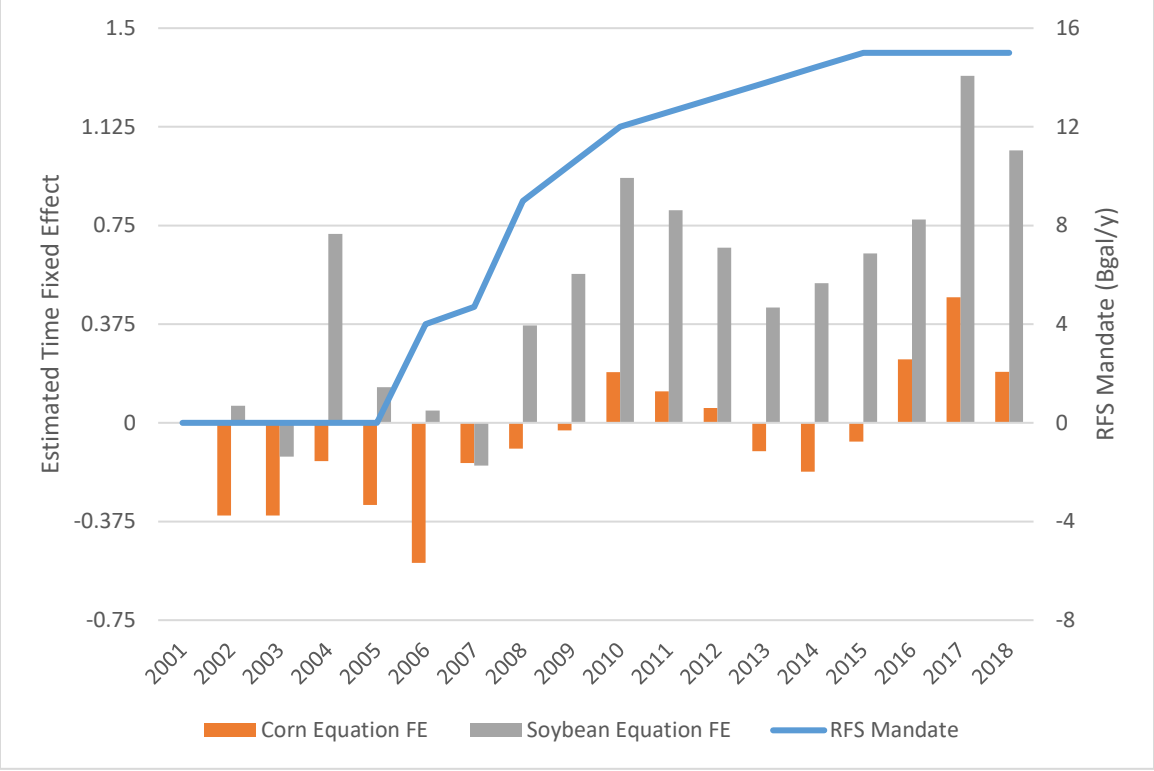


Figure D4: Estimated time fixed effects vs. RFS Mandate, LRR=K

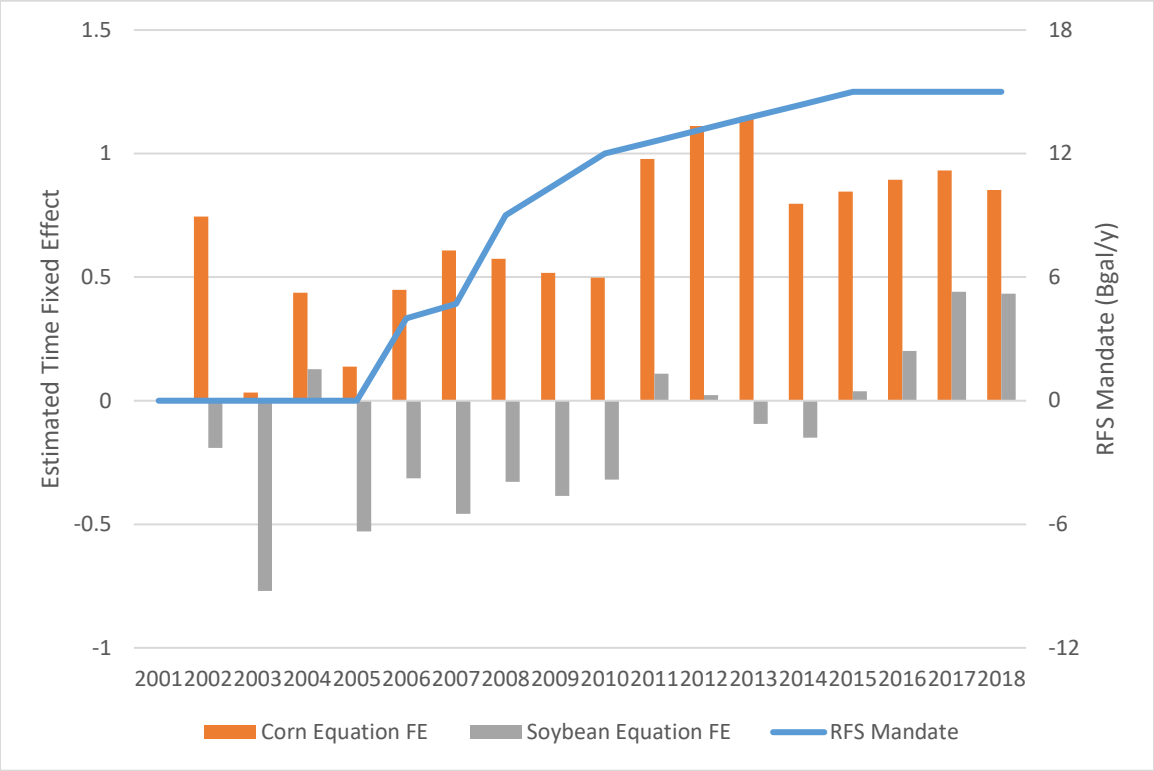


Figure D5: Estimated time fixed effects vs. RFS Mandate, LRR=L

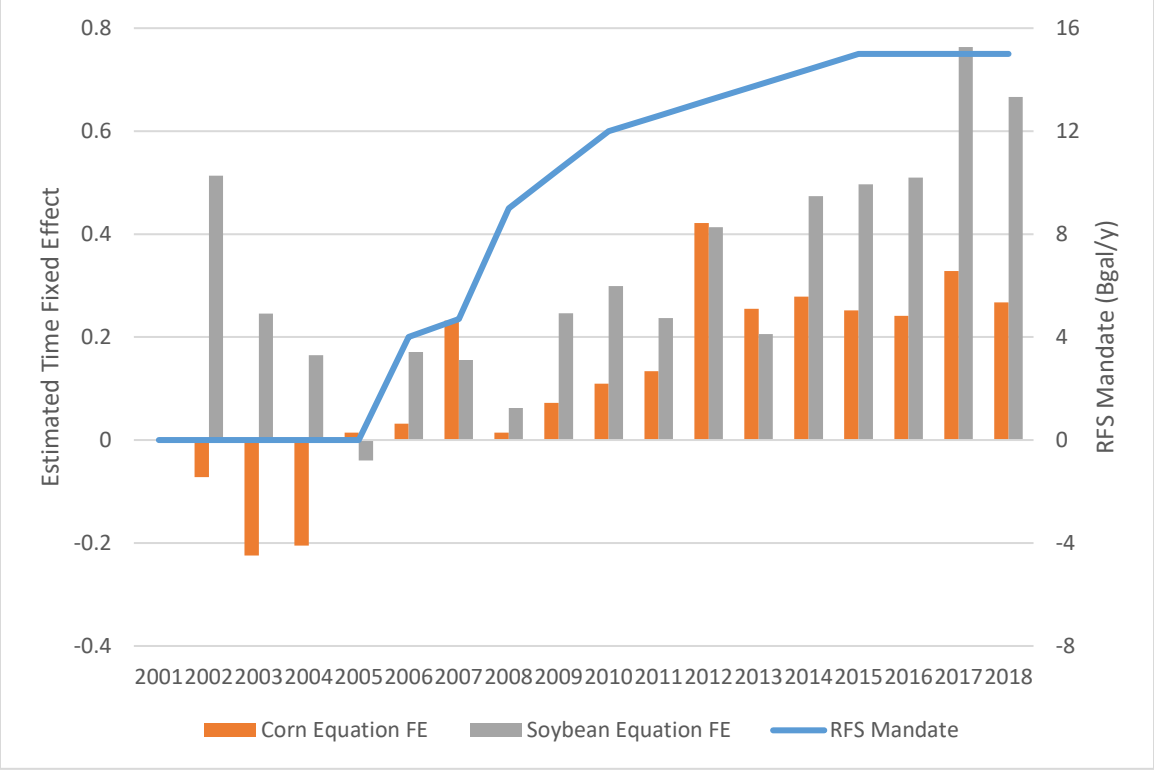


Figure D6: Estimated time fixed effects vs. RFS Mandate, LRR=M

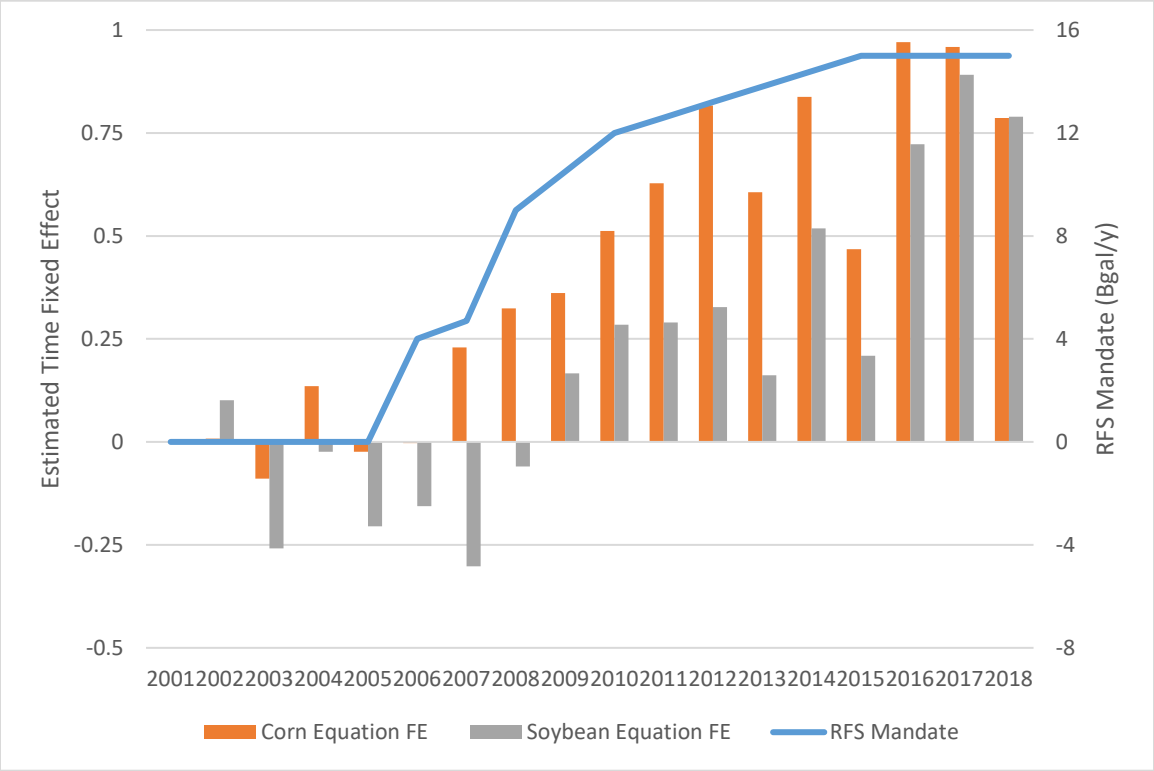


Figure D7: Estimated time fixed effects vs. RFS Mandate, LRR=N