Role of Ethanol Plants in Dakotas' Land Use Change: Analysis Using Remotely Sensed Data

Gaurav Arora
Dept. of Economics & Center for Agricultural and Rural Development, Iowa State University
Email: gaurav88@iastate.edu

Peter T. Wolter
Department of Natural Resource Ecology and Management
Email: ptwolter@iastate.edu

Hongli Feng
Department of Economics
Email: hfeng@iastate.edu

David A. Hennessy
Dept. of Economics & Center for Agricultural and Rural Development, Iowa State University
Email: hennessy@iastate.edu

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

This research is part of a collaborative project supported by the United States Department of Agriculture- National Institute of Food and Agriculture, award number 2014-67003-21772. The views in this paper are attributable solely to the authors.

Abstract:

North and South Dakota have experienced rapid land-use changes in the past decade. Recent studies have shown that these land-use changes are mainly characterized by conversions of grasslands to crop production, especially corn and soybeans. Approximately 271,000 hectares of grasslands were lost to corn and soy production in 2006-2011 period, almost seven times the losses in 1989-2003. The implications of these changing land-uses range from reduced biodiversity and loss of habitat for waterfowl species to low agricultural productivity on drought-sensitive marginal lands. While progress has been made in *characterizing* regional land-use changes, formal analyses establishing *causal relationships* at the local level are lacking. We construct a spatially delineated dataset for the Dakotas and utilize a Difference-in-Difference (DID) model in conjugation with Propensity Score Matching to estimate the impact of an ethanol plant on nearby corn-acres. We hold the advent of an ethanol plant to be a treatment that influences land-use on surrounding agricultural plots. In our preliminary work, based on the Parallel Paths assumption of the DID, we find that the effect of ethanol plants on corn production varies by plants and a single point estimate for all ethanol plants in a region, as usually provided in the literature, can be highly misleading. Surprisingly, we find both positive as well as negative effects of ethanol plants on corn-acres that may be statistically insignificant. Negative estimates are irreconcilable to the economic incentives due to these corn-based ethanol plants. We find intensified corn production and reduced soybeans due to the ethanol plants. Our analysis also reflects a difference in opportunity of converting from wheat to corn and from grass to corn. We use placebo tests and pre-treatment trends in corn acres to examine the Parallel Paths assumption that identifies the DID estimates. We find that this assumption fails and propose to carry out this analysis by incorporating differentiated trends into the DID framework through more flexible assumptions in future. An important contribution of this paper is that it presents a unique research design that uses quasi-experimental techniques to evaluate the impact of a change/policy upon availability of spatially delineated datasets. To this extent, our results are to be viewed as preliminary.

Background and Motivation

Characterizing the Dakotan Land Use Change

Recent research findings suggest rapid land use changes in North and South Dakota, where grasslands have been lost to corn and soybean cultivation. Wright and Wimberly (2013) characterize conversion rates from grass to corn and soybean in the U.S. Western Corn Belt (WCB) from 2006 to 2011. The authors attribute expanding biofuels production and increased crop prices as potential factors driving higher production of these crops and therefore, such land use changes. The WCB spans five states: North Dakota, South Dakota, Nebraska, Iowa and Minnesota. A total of 271,000 hectares of net grassland losses in the Dakots out of 528,000 hectares in all of the WCB's five states imply that conversions during this period were predominantly in the Dakotas. Spatial characterization of land use changes in these two states, using U.S. Department of Agriculture (USDA) Cropland Data Layer (CDL), finds westward expansion of the Corn Belt in regions east of Missouri River that intersect with the Prairie Pothole Region (PPR).

Johnston (2014) provides a longer-term perspective on cropland expansion in the Dakotas, utilizing USDA National Agricultural Statistical Service (NASS) state-level cropping acres from 1980 to 2011, along with USDA CDL spatial imagery from 2006 to 2012. She reports that land attributed to corn or soybean production almost tripled between 1980 and 2011, where in 1980 it accounted for only 5% of the total area in the two states. Author also characterizes land use transitions among various categories such that probability of corn/soy being re-planted to corn/soy increased from 68% in 2006-07 to 80% in 2011-12. On the other hand, such probability for grasslands decreased from 81% in 2006-07 to 74% in 2011-12. In addition, corn and soybeans replaced multiple land uses such as wheat and other small grains that were historically predominant in this region due to their climatic tolerance. Technological advancements yielding drought and cold resistant corn and soybean varieties are reported to be potentially driving such land use conversions.

One other study by Stephens (2008) estimates the probabilities of grassland conversion conditional on amounts of surrounding grasslands, slope and soil productivity. Their annual estimate of the probability of grassland conversion was 0.004 for the Dakotas from 1989 to 2003, amounting to 36,450 hectares of grassland conversion for the period of study. However, they find that probability of conversion is not uniform across all lands of high biological value. Thus, conservation policies for such lands should be prioritized based on the probabilities of conversion, conditional on their location and other land attributes. A 2015 study by Lark, Salmon and Gibbs evaluates the types, amounts and locations of converted lands for cultivation in the conterminous U.S from 2008 to 2012. North and South Dakota are found to have experienced greatest increase in new cultivated land around all U.S. states during this period, predominantly east of the Missouri river. However, northwestern and southeastern North Dakota experienced contraction of croplands in 2008-2012 period. To evaluate conversion rates on native prairies they utilize long-term trend analyses from U.S. Geological Survey spanning 1972-2002. For the Dakotas, they report 14-25 acres of previously native prairies converted per 10,000 acres of land on the east of Missouri river and 10-14 acres converted west of the river. Overall, the Dakotas stood out with highest conversion rates on lands previously attributed to native grasses. Soybeans were found to be the first crop planted upon conversion during 2008-2012 period on east of the Missouri river, whereas west of the river spring and winter wheat were the first crop planted upon conversion in North and South Dakota respectively.

Although Dakotas' native grasslands are a natural resource of national importance, most is under private ownership. Hence, the observed land use changes reported in the recent literature are an aggregate outcome of private decisions by individual landowners. These decisions could be a result of change in many factors including climatic conditions, technology, the local business environment, infrastructure, commodity prices, government payments towards conservation and crop insurance etc. For instance, Claassen *et al.* (2011) provide evidence that federal crop insurance subsidies have intensified cropping practices by reducing related risks. They conclude that the 2008 Sodsaver provision that restricts such subsidies could reduce grassland conversions by up to 9% in the PPR. These land use decisions have not only permanently or temporarily change the overall landscape of these states, but would also have long term impacts on the welfare of local farmers in the Dakotas.

Related Concerns and Policy Implications

Land use changes in the Dakotas raise many ecological, agronomic, environmental and economic concerns and related policy implications. The aforementioned study by Wright and Wimberly (2013) acknowledges the threat to existing wetlands and supported biodiversity from rapid agricultural conversions in the PPR, since wetlands are critical nesting and habitat sites for regional waterfowl species. Increased corn and soybean acres on originally native grassland imply loss of ecosystem services. Reduced populations of game species, when such conversions are in close proximity to the wetlands in the area, augment these losses (Wright and Wimberly, 2013; Johnston, 2014; Stephens *et al.* 2005). Another finding of Wright and Wimberly (2013) that raises concerns as well as interests to policymakers is that, in the Dakotas, corn/soybeans has replaced pasture and hay for livestock production on high quality lands (Land Capability Class II, explained hereafter in the *Data* section). First, higher production of corn and soybeans means fewer opportunities for livestock production. This may be due to an imbalance in incentives towards intensive cropping through reduced risks with insured crops and investments into developing tolerant genetically-engineered seed varieties. Second, rapid increase in corn and soybeans in the region would tailor the socio-economic structure of the region towards more crop-based infrastructure, thereby making crops even more attractive to farmers.

Agronomic issues arising from grassland conversions relate to reduced soil quality and increased soil erosion. Shifts from grass-based agriculture to crop-based agriculture reduce the water holding capacity of the soils, reduce soil ecosystem functions and decrease soil carbon thereby reducing soil productivity. Erosion due to intensified row cropping practices, especially corn, degrades soil quality and pollutes water streams in the region (Wright and Wimberly, 2013; Johnston, 2014). Degraded soils ultimately affect land productivity due to elevated vulnerability to drought due to less suitable climates of this region (Wright and Wimberly, 2013). Further intensification of agricultural activity and prolonged periods of extreme weather events like droughts in this region are considered serious threat to mostly ephemeral wetlands. Further, loss of stored carbon from uprooting the native grasses accounts towards environmental impacts of conversion (Johnston, 2014).

Among the policy suggestions, Johnston (2014) calls for policies that incentivize farmer behavior towards sustainable agricultural practices in light of detrimental environmental and soil-quality implications of intensive corn/soy production on these marginal lands. Further, whereas Stephens (2008) suggests conservation policies to prioritize land with higher chances of conversion based on their location and attributes, Wright and Wimberly (2014) suggest regulating location of biorefineries, deemed responsible for higher corn production in their study. Lark et al. (2015), while recognizing the broad economic and environmental impacts of land use conversion, point to the need for reformed policies aimed towards conserving natural ecosystems. Even though the new Renewable Fuel Standards program (RFS2) mandated procurement of grains for ethanol production only from lands under cultivation prior to December 2007, their study finds substantial increase in croplands in the United States. Further, the authors recognize the importance of the new Sodsaver provision in the 2014 U.S. Farm Bill. This provision, applicable in the PPR states including the Dakotas, dis-incentivizes conversion of native sod for agriculture after January 2014 through reduced crop insurance subsidies. Based on their analysis, the authors recommend a nationwide Sodsaver provision that covers forests and native ecosystems other than grasslands.

Our Contribution: Moving from Characterization towards Explaining Land Use Changes

The above studies characterize the rate and extent of land use conversions in the Dakotas at various spatial and temporal scales. They also speculate on potential factors that driver these land use changes in the region. However, detailed analyses to identify various phenomena that drive land use changes in Dakotas are lacking. We take a first step in understanding this phenomenon by evaluating the impact of ethanol plants on land use changes for these states. All Dakotas' ethanol plants are corn-based. Hence, we ask how the advent of an ethanol plant affects corn plantings in its proximity. There are 19 ethanol plants in Dakotas (four in ND and fifteen in SD) with a combined capacity of 1,386 million gallons per year (mgy, 363 mgy in ND and 1,023 mgy in SD). Together, the Dakotas provide for about 9% of the total U.S. ethanol production capacity, currently at 15,198 mgy. Fourteen (out of the nineteen plants in all) started operations in 2006-2008 period, i.e. after the first RFS program was launched under the Energy Policy Act of 2005 and when rapid land-use conversion rates are found by the pertinent literature, discussed above.

To motivate the economic incentives from ethanol plants, we compare trends in county-level corn basis, before 2006 and after 2008, for counties that house these 14 ethanol plants (see figure 1). An increase in corn basis implies an increase in local corn prices relative to the corn futures price. Such an increase in corn basis could be tied to the incentives from the ethanol plants to land owners with farms in the plants' proximity. It is possible for the ethanol plants to provide such incentives to the farmers who supply them corn from near-by areas, since it saves transportation costs for both supplier and the plant. Figure 1 shows a steeper basis trend for corn in post-2008 periods compared to the pre-2006 period. Therefore, we conjecture a positive and statistically significant impact of ethanol plants on local corn acreage. We also extend our models to analyze the effect of ethanol plants on corn-soybean rotations. We do this by separately analyzing evolution combined acreages of corn and soybeans in relation to the advent of an ethanol plant, and then compare these with that of corn acreage. If the effect of an ethanol plant on corn acreage is higher than on the combined acreage of corn and soybeans, then the implication is intensified corn cropping has occurred through reduced corn-soy rotations due to the ethanol plant.

This paper is subdivided into the various sections. First, a literature review section discusses the relevant findings of the impacts of ethanol plants from studies in the past. Second is a data section that discusses how we constructed a spatially delineated dataset for this analysis and provides a detailed explanation of the relevant variables. Third, the methodology section presents our research design and the Differences-in-Difference model in conjugation with Propensity Score Matching. Fourth is a section for estimation results for each ethanol plant. Lastly, we include discussions and conclusions in another section.

Literature Review

Earlier attempts in this direction involved evaluating indirect impact of ethanol plants on land use change by way of analyzing impacts on local corn prices and farmland values. In the more recent years studies have considered direct impact of ethanol plants on corn acres as measure of land use change. We provide a detailed review of the analyses of impacts on land acreage because these are of direct relevance to this article. We also provide a brief review of analyses involving grain prices and farmland values followed by direct impacts literature.

Direct Impacts: Corn Acreage

Miao (2013) has evaluated the proportion of corn acreage for the Iowa counties in response to location, capacity and ownership capacity of ethanol plants. He utilized a county-level panel data set from 1997 through 2009 and the Arellano-Bond generalized method-of-moments estimator to estimate the effect of ethanol plants on land use shares in the region. The specialized estimator attempts to controls for the endogeneity of ethanol plants and allows controlling for corn-soybean rotations by including lagged dependent variable (that is, proportion of corn acreage). He found a positive and significant impact of ethanol plants on intensity of corn production in Iowa. He further found that, all else equal, locally owned ethanol plants have twice as strong an effect on local corn acreage as their non-locally owned counterparts.

Motamed and McPhail (2011) used remotely sensed data to estimate a non-linear response of proximity to ethanol plants on corn acreage for 12 U.S. Midwestern states: ND, SD, NE, MN, WI, IA, KS, OK, MI, IL, IN, OH. They utilized a panel regression model with corn acreage on each of 10 km X 10 km land parcels from 2006 to 2010 as dependent variable. Their explanatory variables include capacity of the nearest ethanol plant, distance to the nearest ethanol plant and grain elevators, cash bids at the nearest grain elevator and a soil productivity index for these parcels. To incorporate non-linearity of response, their regression model includes logarithmic values of dependent and explanatory variables. They recognize that land parcels' corn acreage and their distance from the nearest ethanol plants are endogenous and use an instrumental variable approach as a corrective measure. They instrument each parcel's distance from the nearest ethanol plant on local transportation infrastructure, specifically distance from the nearest interstate ramp, primary/secondary roads and water ports. This analysis finds that upon moving one percent closer to an ethanol plant corn acreage increased by 0.64% within their region of study.

Turnquist *et al.* (2008) measure the impact of ethanol plants on farmland acreage for the state of Wisconsin between years 2000 and 2006. Although Wisconsin was reported to be losing its farmland to other uses during this period, fallow or undeveloped acres were found to increase. This indicated that factors other than development pressures were driving land use in Wisconsin. In addition, given that increases in fallow land are reversible to agricultural production, evaluating the impact of ethanol plants is interesting. The authors use land use data for municipalities in the state and define zones of 2, 10 and 50 miles around 4 operational ethanol plants during 2000-2006 period. The statistical differences between percentage change in agricultural acreage between 2000 and 2006, within- and outside each of these zones, evaluate the impact of ethanol plants in Wisconsin. Impact of ethanol plants on each of 3 zones' agricultural acreage is found to be statistically insignificant.

Mueller and Copenhaver (2009) analyzed the impact of two Illinois ethanol plants (Illinois River Energy Center (IRE) and Patriot Renewable Fuels (PRF)) on surrounding land use, as part of a larger study to deduce the impact of these plants on greenhouse gas emissions. They used satellite imagery and observe land use in corn supply regions for each plant in 2006, 2007, and 2008 to evaluate its impact. Defining these corn supply regions involved corn growers' surveys and inquiries from ethanol plants to adjudge the spatial extent of their corn suppliers. A 43-mile circle around IRE and 23-mile circle around PRF are respective corn supply regions. The study concluded a weak influence of ethanol plants on direct land use change in their vicinity, and inferred that increasing yields supported increasing exports as well as higher ethanol production.

Brown *et al.* (2014) utilized a spatial econometric regression framework to assess the land use decisions of farmers due to proximity to ethanol plants in Kansas. Using satellite imagery, they separately evaluate conversions from other cropland and non-cropland uses in 2007 to corn production in 2008 and 2009 on 5-acre parcels. The authors find that reducing parcel's distance to nearest refinery by 1% significantly increased non-cropland (other cropland) conversion to corn acres by 5% (4%) in a county 25 miles away from the refinery and by 15% (11%) in a county 75 miles from it. However, the authors recognize that their estimates may be biased due to endogenous ethanol plant locations.

Indirect Impacts: Local Corn Prices and Farmland Values

Miao (2013) also recognized that the literature lacks a consensus about impacts of ethanol plants on local grain prices and agricultural land values, which can be accounted as indirect effects of ethanol plants on land use change. Examples in the context of farmland values are Zhang *et al.* (2012), Henderson and Gloy (2008) and Du *et al.* (2007). Zhang *et al.* (2012) used disaggregated parcel-level data for Western Ohio to evaluate the impact of increased biofuels demand. They conducted difference-in-difference estimation on matched parcels to find increased farmland values in the vicinity of the ethanol plants, at a time that witnessed sharp dip in residential values. The study by Henderson and Gloy (2008) have used a hedonic framework to find a positive impact of ethanol plants on agricultural land values in 2007. Zhang *et al.* (2012) have, however, criticized the hedonic framework due to its inability to correct for selection bias of plant locations. Du *et al.* (2007), on the other hand, reject the hypothesis that ethanol plants significantly affect the cash rentals form farmlands in Iowa. In the context of local grain prices, Katchova (2009), O'Brien (2009), and McNew and Griffith (2005) found a positive significant impact of ethanol plants on local grain prices, whereas Lewis (2010) found that these positive impacts vary spatially. The author found a positive significant impact for MI and KS, and an insignificant impact for IA and IN.

The above review suggests disagreement on the direct and indirect impacts of ethanol plants on local land uses in the literature. Moreover, most studies utilize aggregated county-level datasets. An issue with such aggregated datasets for a location-based analysis is worth considering. Including an indicator (or dummy) variable for the existence of ethanol plants as a regressor assumes its location to be central to its home county when this variable equals 1. It, thereby, assumes that the corresponding ethanol plant will not impact the counties neighboring its home county. If the location ethanol plant is at the center of mass for each home county, we may treat the above as an assumption as reasonable. However, as in the Dakotas, an ethanol plant is often located near the shared boundaries of two or three counties. Consequently, it is appropriate to use spatially delineated data as some studies do. However, these studies ignore the issue of endogeneity that arises in these situations and provide biased estimates of the impacts of ethanol plants.

We make an extensive use of remote sensing tools that provide spatially delineated data with micro-resolutions of the researcher's choice. This article presents estimates of impact of ethanol plants using 500-acre plots as representative decision-making units.[1] This enables the evaluation of the effects of ethanol plants on a plant-by-plant basis, rather than by pooling county-level data for ethanol plants in an entire state or all of Midwestern United States. Adopting a methodology that allows for analyzing impacts of individual plants enables fine-detail scrutiny of local conversion effects. This provides an alternative approach to validate the estimates of the impacts of ethanol plants on corn acreage arrived at from more aggregate methods.

Data

We use remotely sensed data in the form of satellite imagery for the Dakotas from two main sources: land-use from the 'CropScape' portal of USDA-National Agricultural Statistical Service's Cropland Data Layer (CDL) Program and soil quality data from the Web Soil Systems portal of USDA-National Resource Conservation Service (NRCS).

USDA-Cropland Data Layer

CDL satellite imagery for South Dakota are available from 2006 to 2013 and for North Dakota from 1997 to 2013. CDL provides raster (pixel) data for all contiguous US states with different spatial resolutions, 56 m pixels for 2006-2009 and 30 m pixels for other years. To be able to compare land-use statistics across different years, we use remote sensing tools, namely ERDAS Imagine and ArcGIS, to bring each year's imagery to a uniform spatial resolution of 500 acres. To achieve this, each year's raster image was first converted to vector form (pixels to polygons), and then overlaid onto a grid-plot with 500 acre-polygons. Each polygon, which is our representative decision-making plot of land with a unique identifier is observed for every time

[1] We conducted our initial analyses at a much finer resolution (up to 160-acre plots). Aggregating the data up to 500 acres did not change our results significantly. However, higher aggregations suppress measurement errors from satellite imagery.

point thus, facilitating this analysis. Overall, we end up with approximately 104,000 land parcels for North Dakota and 99,000 parcels for South Dakota.

USDA National Resource Conservation Service - Web Soil Systems

We retrieve tabular data for Land Capability Classification (LCC) and representative slope data from the satellite imagery for both states using the Soil Data Viewer application developed by NRCS. Soil Data Viewer provides detailed definitions for both these variables. Briefly, LCC groups soils into eight broad classes each representing degree of limitations for cropping, with higher class codes assigned to greater limitations. LCC classes I and II are well-suited for cropping, whereas LCC classes III and IV require some special conservation practices for cropping, often restricting their use to pasture, rangeland or forests; and LCC V and above have severe limitations that make them impractical for crop cultivation. Representative slope simply measures the rise per unit run. The tabular data combines these soil attributes to geographically delineated and uniquely identified soil map units. Soil map units represent territories that require common management strategies for respective principal land-uses for the purpose of surveys (Soil Data Viewer 6.0 User Guide, 2011 pp. 11). Although map units are the finest spatial resolutions mapped by soil surveys, they may be composed of multiple map unit components that are typically horizontal strips of similar soil characteristics. Map unit size can range from 2 acres to 2,000 acres, depending upon the density of map components accommodated by each of them. Thus, both LCC and slope are aggregated up to map unit level facilitated by the Soil Data Viewer application. The aggregation criteria are contingent upon slope and LCC being continuous and categorical variable, respectively.

Representative slope was aggregated as average of slope values of all map components weighted by their respective share of area within the map unit. LCC, on the other hand, was aggregated by a ‘dominant condition’ criterion thereby assigning LCC category to the map unit represented by most acres among its components. In other words, the software first sums up the area of all map components that fall under the same LCC category and then assigns the LCC category with maximum acres to the corresponding map unit. Aggregation using the dominant condition criteria notes that the assigned LCC category t may represent as little as 25% of its area. In addition, where different LCC categories are represented by an equal area, ties are broken by choosing a higher LCC value. The former issue is related to the available aggregation process for this categorical variable. The latter issue is of minimal concern, arising in 4 out of 156 map units for North Dakota representing 0.7% area of all of state and for 2 out of 260 map units in South Dakota with only 0.6% area in this state. Moreover, the tie-breaker rule becomes even less irrelevant since we use percent land area (in our 500 acre land parcel) with LCC I, II as our independent variable. Choice of LCC I, II is based on above definitional and statistical reasons discussed in methodology section. Also, these soil quality variables remain constant temporally.

Ethanol Plants’ Spatial Coordinates

The spatial coordinates of ethanol plants, ultimately used to determine treatment and control groups, were acquired by using the Google Earth application in conjugation with online maps locating plants made available on *Ethanol Producer Magazine*’s website. Overall, there are 4 ethanol plants in North Dakota and 15 ethanol plants in South Dakota. We conduct our analysis using 8 ethanol plants (4 in each state), listed in table 1 with spatial locations in figure 2. Choice of ethanol plants is driven by our methodology and land-use data availability in South Dakota (2006-2013), discussed hereafter under ‘Estimation Results’.

Methodology

The objective is to quantify how the emergence of an ethanol plant affects local land-use change. The detailed micro-level panel dataset for the Dakotas allows us to implement a quasi-experimental method to evaluate the effects of ethanol plants on land-use patterns in their neighborhood. In this sense, we interpret the emergence of an ethanol plant as treatment where pre-and post-treatment year outcome levels are observed land-use patterns before and after its emergence, respectively. Specifically, we use the Difference-in-Difference (DID) estimation strategy in conjugation with propensity score matching (PSM) to evaluate the role of ethanol plants. Using the DID approach is reasonable since the location of an ethanol plant is endogenous to land-use trends in its locality. The issue of endogeneity arises because Dakotas' ethanol plants are corn-based facilities and thus their location decisions could place them in regions with high corn production in pre-plant years or with high potential for corn production in the post-plant years. DID controls for such endogeneity by estimating causal impacts as difference between average temporal trends of land-use acres across treated and untreated groups, assuming that in the absence of the ethanol plant land-use in both these groups would evolve equivalently. This assumption of parallel trends requires constituents (land plots, here) of treated and untreated groups to be alike, except for the land-use patterns potentially affected by the presence of these ethanol plants. That is, estimated treatment effects are unbiased if these land parcels are randomly assigned to the treatment group and we control for any other within-group or across-group dissimilarity among them. We utilize PSM to ensure random assignment of land parcels to treatment group by conditioning their treatment selection on the observed soil quality variables. The soil quality variables are central to land-use decisions, and thus potentially influence ethanol plants' location choice to regions with land attributes favoring corn production.

In particular the PSM strategy restricts the sample for estimating treatment effects to one where estimated conditional probability of treatment (or propensity score, PS) for each untreated parcel is close ‘enough’ to its treated counterpart. We implement a one-to-one nearest-neighbor propensity score matching algorithm and include only those treated parcels for which there exists an untreated parcel whose PS lies within a pre-assigned radius (absolute difference between PSs) of each corresponding treated parcel’s score. The choice of this radius involves a trade-off between bias and efficiency of treatment effects, as a smaller radius will yield more similar land parcels in both groups but at the same time a smaller sample for estimating treatment effects.[2] We report treatment effects calculated using samples from pre-assigned radius of 0.01.[3] In addition, post-matching heterogeneity in the distribution of soil quality variables among treated and untreated groups may also potentially bias our treatment effects’ estimates. Therefore, we conduct statistical checks on difference in mean of these observables across matched treated and untreated samples (also known as balancing). We find that reducing the pre-assigned radius yields higher balance across the two groups used for estimating treatment effects. It is noteworthy that while PSM controls for selection on observables, the DID estimation approach controls for selection on unobservables through individual and trend fixed-effects in the regression framework (List *et al.* 2003). In the DID regression framework using matched samples, we further control for pre-treatment land-use decisions as an opportunity to convert to

---

[2] We implement the PSM algorithm developed by Fraeman (2010), which optimizes the sample size in two steps. First, it searches for all possible matches to each treated sample within the pre-assigned radius and then while assigning matches to these treated parcels it prioritizes those with the least number of matches from the first step. We also use the SAS code that is published within Fraeman (2010) to implement this algorithm.

[3] Treatment effects calculated using a pre-assigned radii of 0.05, 0.10 and the unmatched cases are available upon request. The treatment effects’ estimates substantially differ with and without matching.

corn. Illustratively, a land plot almost entirely attributed to corn in pre-treatment years does not offer space for conversion irrespective of the advent of an ethanol plant. In addition, even if it was predominantly under wheat (or grass) in the pre-treatment year, the opportunity to convert comes with switching (or conversion costs), respectively. Further, in recognition of the fact that farmers usually grow corn and soybean in rotation, we evaluate treatment effects for corn as well as the combined acreage of corn and soy as our dependent variables.

Defining treatment and control (untreated) groups

Until now we have laid down our framework for estimating the treatment effects without providing a clear definition for treatment and control groups. The argument that the location of an ethanol plant is potentially influenced by opportunity as for corn production in its vicinity relates to a cost minimizing outcome. If an ethanol plant procures most of its annually required corn from near-by areas, it would save on transportation and related logistics costs, and so is willing to compensate local suppliers. Therefore, to define our treatment and control groups, we assume that these transportation costs are monotonic in the Euclidean distances of a land parcel from an ethanol plant and that the ethanol plant bears at least some these costs. In this scenario, a representative corn supplier nearer to the ethanol plant has higher incentive to grow corn on their field than one farther away, *all else equal*. Consequently, we choose to designate samples that lie closer to the ethanol plant as treatment samples and ones farther away as control (or untreated) samples.

How Significant are Transportation Costs? Empirical Evidence

To support our argument that transportation costs and distance are sensible treatment and control parameters, we present back of the envelope calculations. Consider transportation trucks with carrying capacity of 1 ton (=39.4 bushels[4]) corn and mileage of 134 ton-miles per gallon. According to the U.S. Energy Information Administration, the annual average diesel price in U.S. ranged from $2.4 - $4 post 2005. At such per gallon rates for diesel, the fuel cost of transporting 1 bushel of corn for 1 mile would range from 0.05 to 0.07 cents. O'Brien (2009) estimates the total transportation cost to be approximately 4 times the fuel cost. Therefore, the maximum willingness to pay for the owner of an ethanol plant to incentivize a farmer located 50 miles away from the plant to grow corn would range from 10 to 14 cents per bushel. On the other hand, cash rents for croplands ranged between $39-$46.5 in ND and $53-$71.5 in SD from 2006-10 (USDA NASS Land Values Summary, 2006-10). Given the corn yields of 111-132 bushels/acre in ND and 97-151 bushels/acre in SD (USDA NASS Quick Stats, 2012), the average cropland rents for the Dakotas were between 30-73 cents per bushel of corn. As the transportation costs are 14%-47% of the total cropland rental values there should be strong pressure for proximate landowners to engage in corn production.

The DID model in conjugation with PSM

This sub-section elaborates on the working of a standard DID model for the purpose of this article, in conjugation of the PSM strategy. We follow the standard DID model of Abadie (2005). Consider a representative land parcel i with $C_{i,t}$ and $CS_{i,t}$ as its corn acreage and combined corn and soy acreage respectively at time period t. We introduce binary variables d_i and d_t to designate treatment/control groups and pre-/post-treatment periods respectively. So, $d_i = 1$ for treated parcels and equals 0 otherwise, while $d_t = 1$ for time periods after the advent of an ethanol plant and equals 0 otherwise. Further, denote $t^+(t^-)$ as the set of post-treatment (pre-

[4] Bushel/Ton Converter. www.agriculture.alberta.ca

treatment) time periods with t_0 as the treatment year[5]. Intuitively, to evaluate a treatment effect for treated parcel i's corn acreage we would compare the outcome levels with and without ethanol plant in the post treatment era, that is $C_{i,i}$ [6] with $t \in t^+$. Consequently, the average treatment effect for the treated (ATT) equals $E[C_{i,t^+}^T - C_{i,t^+}^U \mid d_i = 1]$, where superscript T(U) denote presence (absence) of the plant.

The issue, though, is that we only observe the post-treatment acreage of corn with treatment while the without treatment outcome for these years is unobserved. The DID approach is primarily designed to overcome this issue. By assuming that treated and control parcels follow parallel land use changes if the ethanol plant had not emerged at t, we can evaluate the ATT using the pre- and post-treatment outcomes of both groups (Abadie, 2005). This assumption is key to identify the estimates of treatment effects because in the event that this assumption fails our estimates could not be trusted. Also, observing these land parcels individually lets us control for soil quality and land use shares at time $t_0 - 1$ as covariates. Hence, our ATT will be a conditioned on covariates other than the treatment dummy. The parallel land use changes assumption among both groups (discussed earlier) can be expressed mathematically as

(1) $$E[C_{i,t^+}^u - C_{i,t^-}^u \mid Z, d_i = 1] = E[C_{i,t^+}^u - C_{i,t^-}^u \mid Z, d_i = 0],$$

In equation (1) the superscript u signifies that we are considering the case of no treatment (both groups stay untreated) and Z is the set of covariates. If (1) holds true then the ATT is calculated as

[5] Example: For the Red Trail Energy ethanol plant that came up in 2007, $t^+ = \{1997, 1998, ..., 2006\}$ and $t^- = \{2008, 2009, ..., 2013\}$.

[6] We present the model for corn acreage. An extension for combined corn and soy acreage follows by changing the notation from $C_{i,t}$ to $CS_{i,t}$.

(2) $$ATT = E[C_{i,t^+} - C_{i,t^-} \mid Z_i, d_i = 1] - E[C_{i,t^+} - C_{i,t^-} \mid Z_i, d_i = 0]$$

ATT, in equation (2) can be estimated as β_3 from the regression framework in equation (3) below.

(3) $$C_{i,t} = \beta_0 + \beta_1 d_t + \beta_2 d_i + \beta_3 d_i d_t + \beta_{4,t} Z_i + \varepsilon_{i,t}$$

In equation (3) $\beta_0, \beta_1, \beta_2, \beta_3$ and $\beta_{4,t}$ are regression coefficients. Note that $\beta_{4,t}$ allows the effect of time-invariant covariates to vary across pre- and post-treatment years (Abadie, 2005).

To illustrate the extension of a standard DID model to incorporate PSM, consider the decomposition of the set of covariates $Z_i = \{X_i^a, X_i^b\}$. Here, the set X_i^a contains the soil quality variables LCC and slope and set X_i^b represents the initial land use conditions for parcel i. We match the parcels based on their soil quality parameters. The justification for matching on soil quality is that we seek to ensure random placement of these parcels into their respective groups relative to the location of ethanol plant. An ethanol plant's location decision must be based on the potential for corn production based on land quality. But to say that the plant chooses to locate on land use status in just the penultimate year of it starting operations is logistically infeasible. Miao (2013) acknowledges that the ethanol plant goes on-line as early as 3-years prior to starting operations. We use a logistic model with d_i as dependent variable and X_i^a as the set of regressors to estimate a propensity score (denoted by $P(X_i^a)$) for each parcel in the treatment and control groups. Specifically, we use the percentage land with LCC ≤ 2 (denoted, LCC2) and percentage land with slope ≤ 5 (denoted by Slope5) as regressors in the logit regression. We match the parcels using a nearest-neighbor matching algorithm discussed earlier. By matching, we seek to ensure that parcels' propensity to be treated is alike across groups, conditional of the time-invariant intrinsic property of land – soil quality. Post-matching, we use the DID regression framework as in equation (3) with covariates reduced to X_i^b. A conceptual expression for the ATT from our extended model, denote as ATT^m, can be written as

(4) $$ATT^m = E[C_{i,t^+} - C_{i,t^-} \mid P(X_i^a), X_i^b, d_i = 1] - E[C_{i,t^+} - C_{i,t^-} \mid P(X_i^a), X_i^b, d_i = 0]$$

The estimation of ATT^m follows from equation (3) with Z_i replaced by X_i^b and the sample data used for this post-matching estimation will be a subset of its counterpart in (3). Therefore, if β_3^m is the estimate of our new ATT, then it can be retrieved estimating the following regression equation

(5) $$C_{i,t} = \beta_0^m + \beta_1^m d_t + \beta_2^m d_i + \beta_3^m d_i d_t + \beta_{4,t}^m X_i^b + \varepsilon_{i,t}.$$

An aspect of our research design that differentiates it from most other quasi-experimental studies is a non-centrally administered or a non-exogenous (cite some examples) treatment. We designate the advent of an ethanol plants as treatment, which itself is a market outcome that must be bridging the supply-demand gap in commodities and biofuels markets in the Dakotas and even beyond. The implication of this non-exogenous intervention is that we do not have exogenous control groups. Rather, our treatment and control groups follow the ‘rule of thumb’ that treated parcels are located nearer to the ethanol plant than their untreated counterparts. This allows innumerable possibilities of treatment and control groups near each ethanol plant’s location and practically inexhaustible combinations that can be included in our article. However, to ensure robustness of our results we designate two treatment groups and two control groups for each ethanol plant. Based on our definition that parcels farther away from the ethanol plants are controls when compared with the treated, we conjecture that treatment effects using the nearest treatment and the farthest control groups will be larger in size and more significant than the other three combinations. We present the regression results for this particular combination and compare with others as a robustness strategy. In cases where we have sufficient pre-treatment and post-treatment years we also estimate treatment effects for multiple combinations of pre-or post-treatment years (advocated by Meyer, 1995). Bertrand *et al.* (2004) find serial correlation as a severe issue in studies implementing DID strategy on panel data with many pre-and post-treatment years, leading to over-rejection of null hypothesis (of no treatment effect). The authors point to high serial correlation of interaction term of individual- and trend-fixed effects as one of the causes of this problem, which we believe to affect our estimates. The remedy suggested by the authors to overcome this issue is aggregating through pre- and post-treatment years by using mean outcome levels rather than for individual years' is implemented here.

The Placebo Treatment Effects

Further, in recognition of the non-exogenous treatment we utilize placebo tests or falsified treatments to validate the robustness of our results. We conduct temporal placebos, meaning that we assume the advent of an ethanol plant in a year that predates the actual treatment. These temporal placebos are conducted for North Dakota plants since the data is available starting from 1997. This gives us a sufficient window of time periods to designate various falsified treatments and the pre- and post-treatment years for each of these. Placebo tests are important as they allow validating our identification strategy to estimate treatment effects.

The farthest treated and control parcels are located at a maximum distance of 100 km (62 miles) from each other in our empirical setup. We, therefore, anticipate that the physical characteristics of these parcels and their initial land use shares will play a major role in identifying treatment effects. Weather may be another variable of interest, which we assume to be uniform across our treated and control parcels. Since weather data points are collected at weather stations covering multiple counties and our analysis only spans 60 miles strips, we think that our assumption is reasonable. By definition, distance from ethanol plants are the sole differentiator of treated and untreated land parcels. However, these end up contained within multiple boundaries. Although the markets and incentive structure may vary substantially across counties, we do not expect these to affect how much corn farmers grow due to advent of an ethanol plant in their vicinity. Even if we were to consider county-fixed effects for each of the parcels, they would cancel out due to the first difference operator inherent to the DID estimator, on pre- and post-treatment outcome levels of each parcel. Despite of the fact that we have been careful in choosing the covariates for above regression framework, there might still be factors that we fail to control. An example would be matching the parcels based on soil moisture, not done here due to incomplete data. However, a good or bad rainfall year could influence the impact of advent of an ethanol plant in our treatment and control groups even if we assume uniform rainfall measured across all parcels. The right amount of precipitation leading to higher soil moisture on a LCC II, flat sloped land could influence farmers' decision to grow water-thirsty corn, with or without ethanol plant in vicinity. To address our inability to capture such effects that may confound the estimated treatment effects, we include temporal placebos. If we successfully control for all relevant covariates and our matching strategy is perfect, we should get a zero or statistically insignificant placebo treatment effect. However, a significant (positive or negative) placebo treatment effect would point towards ambiguity in our identification strategy and allow statistical correction of our estimates of the actual treatment.

Estimation Results

As mentioned earlier, there are 19 ethanol plants in North and South Dakota. We include all four North Dakota ethanol plants, but restrict our analysis for South Dakota to four out of 15 ethanol plant due to data availability. The CDL data for South Dakota only goes back until 2006.The four South Dakota ethanol plants, included here (see Table 1), started operations in 2008. This allows implementing the DID estimation strategy through pre- and post-treatment years. We analyze the effects of POET and NuGen ethanol plants together as cluster 1 and ABE and GLE as cluster 2 due to their spatial proximity. A vector description (dimensions and directions) for the treatment and control groups of each of these ethanol plants is provided in Table 2. These rectangular-shaped groups can be visualized in figure 3, as an example. Another factor that determined which land parcels entered treatment and control rectangles was existence of ‘other’ ethanol plants nearby. We follow the linear city model and consider all ethanol plants as market terminals with designated market capacity. So, while deciding which land parcels enter our rectangles we ensure that linear distance of a parcel is minimum to the ethanol plant under study. The linear distances are normalized by ethanol plants’ capacities. For instance, if two ethanol plants with annual capacities 20 and 80 million gallons are 100 km apart, then market designated for the larger (smaller) ethanol plant is 80 (20) km from its location. Such details for ethanol plants considered for our analysis are added in Table 2 (see the ‘Remarks’ column).

Treatment Effects’ Estimates

To estimate the treatment effects, we modify our regression framework (equation (5)) to include the first differences of pre- and post-treatment outcomes as dependent variables. Our regression estimates, therefore, are to be viewed as regression coefficients of equation (6) below.

(6) $$C_{i,t^+} - C_{i,t^-} = \beta_1^m + \beta_3^m d_i + \beta_4^m X_i^b + (\varepsilon_{i,t^+} - \varepsilon_{i,t^-})$$ [7]

[7]Equation (6) is retrieved by taking a difference on the pre- and post-treatment versions of equation (5). That is $\{C_{i,t^+} = \beta_0^m + \beta_1^m.1 + \beta_2^m d_i + \beta_3^m d_i.1 + \beta_{4,t^+}^m X_i^b + \varepsilon_{i,t^+}\}$ $-$ $\{C_{i,t^-} = \beta_0^m + \beta_1^m.0 + \beta_2^m d_i + \beta_3^m d_i.0 + \beta_{4,t^-}^m X_i^b + \varepsilon_{i,t^-}\}$. Again, similar results follow for the combined corn and soybeans case by changing the notation $C_{i,t}$ to $CS_{i,t}$.

Where, C_{i,t^+} is the average of corn acres in post-treatment years and C_{i,t^-} is the average if corn acres in pre-treatment years. In addition, β_1^m captures the trend-effects of moving between pre- and post-treatment periods, β_3^m is the estimate of ATT^m (defined earlier), and $\beta_4^m = (\beta_{4,t^+}^m - \beta_{4,t^-}^m)$ is the differentiated role of the set of controls X_i^b on change in corn acreage through time. We now present our estimation results for each ethanol plant included in Table 1. Our regression analysis also includes $\ln(C_{i,t^+}) - \ln(C_{i,t^-})$ as a dependent variable to compare rate of change in outcomes pre- and post-treatment. This is especially useful when, in the pre-treatment period, outcome levels (corn acres in this case) between treatment and control groups differ significantly. Illustratively, say the conditional mean of corn acreage for control groups is *a* acres and for treatment group is *2a* acres. If there were no treatment effect and both groups would grow by a factor 2, then post-treatment corn acres for control and treatment groups will be *2a* and *4a* respectively. Our definition of ATT will yield a positive treatment effect, even though it was zero. Using log-linear regressions will compare the rate of change and would help avoid such confounding results.

Red Trail Energy

For the Red Trail Energy ethanol plant (RTE) that started operations in year 2007, we have $t^- = \{1997, ..., 2006\}$ and $t^+ = \{2008, ..., 2013\}$. Consequently, $X_i^b = \{W_{i,2006}, G_{i,2006}\}$, where $W_{i,2006}$ is the 2006 wheat acreage on a representative parcel i and $G_{i,2006}$ is the 2006 grass cover on i. For RTE, the pre- and post-treatment summary statistics for both treatment and control groups are included in Table 3 and corresponding estimation results are included in Table 4.

Table 3 reveals that the unconditional change of mean corn acres is higher in the treatment group. However, the regressions revert that due to negative and significant impediments posed by grass acres are higher in case of control group rather than treatment group. A negative significant coefficient for grass acres in 2006 can be explained due to high switching costs. Such switching costs are further exaggerated due to pre-treatment low corn (and soy) acres treatment and control groups. Such low levels of corn and soy acres could mean lack of experience, information and technology to switch to corn in this region. This makes sense because RTE is located west of the Mississippi river, away from the western edge of the Corn Belt. The conditional rate of change of corn (and soy) are also negatively affected due to the ethanol plant, though the change is insignificantly different from zero at 95% confidence interval. It should be noted that the intercept, in absolute value, is large as compared to treatment and other controls. Since the intercept captures trend-effects (discussed earlier), large intercepts relative to treatment effects suggest that ethanol plants only provide for a small fraction of the overall land use change among the groups.

Blue Flint

For the Blue Flint ethanol plant (BF) that started operations in year 2007, we have $t^{-}=\{1997,...,2006\}$ and $t^{+}=\{2008,...,2013\}$. Consequently, $X_i^b=\{W_{i,2006},G_{i,2006}\}$, where $W_{i,2006}$ is the 2006 wheat acreage on a representative parcel i and $G_{i,2006}$ is the 2006 grass cover on i. For BF, the pre- and post-treatment summary statistics for both treatment and control groups are included in Table 5 and corresponding estimation results are included in Table 6.

Due to the ethanol plant, unconditional mean of corn acres among the two groups grew almost equivalently while the combined corn and soy acreage grew more for the treated. But again grass acres, that are a significant impeding factor for conversion to corn (or soy) here, are higher among the untreated parcel. This yields negative treatment effect, such that conditional corn acres increased more among untreated parcels. However, while comparing the rate of change among parcels treatment effect is positive for corn acres and negative for combined corn and soy, although insignificant. A positive growth rate of corn acres and negative rate for corn and soy combined may have implication for crop rotation. This suggests intensified corn cropping where -corn-soy-corn-soy- rotational structure shifting to -corn-corn-soy-corn-. Also, for the log linear regressions coefficients on initial wheat acres (in 2006) are positive and significant revealing opportunity to switch to corn. At the same time, negative (insignificant) coefficients on initial wheat acres in the linear regressions suggest costs of switching to corn production that are lower than conversion costs from grass acres. Again, large intercepts relative to the treatment effects suggest that ethanol plants are not a major determinant of the overall land use change among the groups.

Tharaldson Energy

For the Tharaldson Energy ethanol plant (TE) that started operations in year 2006, we have $t^{-} = \{1997,...,2005\}$ and $t^{+} = \{2007,...,2013\}$. Consequently, $X_i^b = \{W_{i,2005}, G_{i,2005}\}$, where $W_{i,2005}$ is the 2005 wheat acreage on a representative parcel i and $G_{i,2005}$ is the 2005 grass cover on i. For TE, the pre- and post-treatment summary statistics for both treatment and control groups are included in Table 7 and corresponding estimation results are included in Table 8.

A feature that distinctly distinguishes TE from RTE and BF is higher pre-treatment acres of corn and soybeans in among treated and untreated groups. Also, with treated groups having more than twice as many corn acres and, also that many combined corn and soy acres comparing rates of change is more reasonable than the absolute changes. Specifically, log-linear regressions will provide more reasonable inferences as compared to their linear counterparts. The treatment effect here is found to be negative, although more negative for combined acres of corn and soy. This suggests intensifying corn cropping and forgone corn-soy rotations in the process. Trend-effects again dominate the treatment effects in this case. However, grass acres may serve as opportunity to grow corn, despite higher conversion costs, due to lower grass cover prior to the ethanol plant.

Hankinson Renewable Energy

For the Hankinson Renewable Energy ethanol plant (HRE) that started operations in year 2008, we have $t^{-} = \{1997, ..., 2007\}$ and $t^{+} = \{2009, ..., 2013\}$. Consequently, $X_i^b = \{W_{i,2007}, G_{i,2007}\}$, where $W_{i,2007}$ is the 2007 wheat acreage on a representative parcel i and $G_{i,2007}$ is the 2007 grass cover on i. Since there are too many pre-treatment years compared to post-treatment years, we introduce $t_1^{-} = \{2003, ..., 2007\}$ as an alternative pre-treatment years to seek any difference in treatment estimates. For HRE, the pre- and post-treatment summary statistics for both treatment and control groups are included in Table 9 and corresponding estimation results are included in Table 10.

The regression results for HRE suggest that this ethanol plant has had a positive impact on corn acres, and the combined corn and soy acres. However, it is clear that impact on corn acreage has been greater than that on the combined corn and soy acreage. This may have implications for corn and soy rotation. Similar to our inferences above, corn acres seem to intensify, leading to lesser corn-soy rotations due to the advent of the ethanol plant. This inference on rotations is especially quite strong if we compare the historical pre-treatment years (starting 1997), rather than the recent ones (starting 2003). Once again, higher wheat acres in the year before the ethanol plant lead to positive significant increase in corn acres (and combined corn/soy acres as well). Also, unlike the previous three ethanol plants trend-effects are dominated by HRE's treatment effect for log-linear regressions while trend-effects dominate in the linear regressions case. Thus, HRE is an important determinant of the rate of change in corn production whereas it is not so important to absolute changes in corn acres.

Cluster 1: POET Bio refinery and NuGen Energy

Cluster 1, which is a conglomerate of POET Bio refinery and NuGen Energy, (PBNE) started operations in 2008. So, we have $t^{-}=\{2006,2007\}$, $t^{+}=\{2009,...,2013\}$ and $X_i^b=\{W_{i,2007},G_{i,2007}\}$, where $W_{i,2007}$ is the 2007 wheat acreage on a representative parcel i and $G_{i,2007}$ is the 2007 grass cover on i. We also include $t_1^{+}=\{2009,2010\}$ as an alternative post-treatment years' set to seek any difference in treatment estimates. For PBNE, the pre- and post-treatment summary statistics for both treatment and control groups are included in Table 11 and corresponding estimation results are included in Table 12.

The corn acres seem to be positively impacted by emergence of PBNE in later years (2011-2013), as the treatment effect is insignificant for the post-treatment years t_1^{+}. The rate of growth in corn acres, however, was not significant due the plants. Given that initial corn acreage between treated and control parcels is significantly different, the inference on rate of growth is more reliable than absolute acres. But, the effect of these ethanol plants on the combined acreage of corn and soybeans is unanimously positive and significant. This implies that, unlike in North Dakota, these South Dakota ethanol plants have de-intensified corn cropping and encouraged corn-soy rotations. Another finding that differs here from the analysis of North Dakota plants is negative trend effects. It seems as if the higher corn acres are driven due to the advent of these ethanol plants, since treatment effects and intercept are comparable in size. Further, higher initial (2007) wheat and grass acres have positive and significant impact on both, corn acreage and combined acreage of corn and soy.

Cluster 2: Aberdeen Bio energy and Glacial Lakes Energy

Cluster 2, which is a conglomerate of Aberdeen Bio energy and Glacial Lakes Energy, (ABGL) started operations in 2008. So, we have $t^{-} = \{2006, 2007\}$, $t^{+} = \{2009, ..., 2013\}$ and $X_i^b = \{W_{i,2007}, G_{i,2007}\}$, where $W_{i,2007}$ is the 2007 wheat acreage on a representative parcel i and $G_{i,2007}$ is the 2007 grass cover on i. We also include $t_1^{+} = \{2009, 2010\}$ as an alternative post-treatment years' set to seek any difference in treatment estimates. For ABGL, the pre- and post-treatment summary statistics for both treatment and control groups are included in Table 13 and corresponding estimation results are included in Table 14.

The initial average corn (combined corn and soy) acreage for treatment group is almost twice (thrice) when compared to the control group. Hence, we draw our inferences for this ethanol plant from rate of change equations. We find negative impacts of these ethanol plants on treated corn acreage, which is driven by the decreasing corn and combined corn and soy acreage for treatment groups coupled with corresponding increase for control group. A more negative treatment effect for combined corn and soy acreage points out to intensified corn cropping relative to corn-soy rotations. Also, as in the other cluster in South Dakota, trend-effects are negative and are dominated by the treatment effects here. Initial wheat and grass acres have positive significant effects on corn and soy production.

Summarizing the Estimation Results

The treatment effects are found to vary in size, sign and significance by individual ethanol plants. This finding disapproves estimation strategies used in the past that discover only one point estimate of the impact of ethanol plants for all of Iowa or, even, the U.S. Midwest. However, the negative significant treatment effects are both surprising and irreconcilable due to earlier argued higher relative incentives near the ethanol plants. This was because transportation costs (that are monotonic in distance) are quite significant compared to cropland rentals values in the Dakotas. To understand and validate these negative treatment effects, we examine impact of ethanol plants on county-level corn basis and evaluate placebo treatment effects. The placebos and robustness checks from multiple treatment and control groups are discussed in the next section.

We also find that intensity and type of impact of ethanol plants on local land use depends on its spatial location, rather than only its capacity as controlled for in previous literature. Specifically, for ethanol plants that lie on the Corn Belt (HRE, PBNE and ABGL) we find treatment effects to dominate or be at least comparable to the trend-effects. Whereas, for RTE, BF (located west of the edge of the Corn Belt) and TE (located north of the Corn Belt) the treatment effects are dominated by the trend-effects. So, ethanol plants could be a major factor in determining the overall evolution of corn and soybean acres in their proximity when they operate among areas densely planted in corn/soy. We also find that the advent of ethanol plants could impact corn-soy rotations in an area. In 5 out 6 cases considered in this analysis, we find corn intensification relative to combined corn and soy acreage. This points towards lesser corn-soy rotations in close proximity of these ethanol plants. We also find initial wheat and grass acres to significantly affect the evolution of corn and soybean acres in post ethanol plants years. Controlling for these variables reveals higher wheat to encourage corn relative to higher grass, which is a result of higher conversion costs (of sod-busting) than switching costs among crops.

Corn-Basis Analysis

We had conjectured earlier that proximity to ethanol plants could offer strong incentives to drive more corn production. This conjecture was primarily based on our back of the envelope calculations and also, partially, on the existing literature. Our findings, in contrast to the conjecture, of insignificant or negative treatment effects are indeed surprising. To better understand and reconcile these findings we analyze the effects of ethanol plants on corn basis. If the advent of an ethanol plant were to incentivize corn production in its proximity, these incentives should be observable in a market setting as increase in corn basis. So, the treated parcels should have a higher increase in basis post-treatment as compared to the untreated ones. This would ultimately feed into land-use decisions and lead to higher corn acres in close vicinity of the ethanol plant. Our back of the envelope calculations focused on the maximum willingness to pay for an ethanol plant to incentivize corn production for a supplier unit closer to its location. As opposed to the maximum willingness to pay, increase in basis will reveal an actual willingness to pay for the ethanol plants as observed in the market setting. In case that the actual willingness to pay for the ethanol plants does not increase as expected, we can, at least, justify the insignificant treatment effects estimates found earlier.

We retrieved a county-level dataset providing monthly corn basis from 2000 to 2013, for North and South Dakota[8]. We present comparative basis trend-plots from 2000 to 2013 for the counties that contained the treatment and control groups for 4 out of six ethanol plants (or clusters) included here (figures 4-7). In figures 4-7, the county that contains the ethanol plant (its home-county) is plotted as a solid series while others are plotted as hashed series. If the ethanol plant were to significantly increase the compensation to farmers for supplying corn in its close vicinity, we should be able to visualize it through its home-county's basis time-series plot. In an event of significant impact of the advent of an ethanol plant, we expect the basis series for it home-county to deviate upwards from its counterparts. Further, the home-counties for RTE and BF and their respective neighbors suffer with missing values and are inappropriate to deduce any impacts of these plants.

[8] Dataset Source: *Geo Grain*.

Figures 4-7 show increased relative basis for Richland county (home to HRE) and Turner county (Home to Cluster 1). This justifies the positive significant treatment effects for these two cases. However, the corn basis for Cass county (home to TE) seems to be stagnated post-treatment year. Also for cluster 2, stationed in two counties, corn basis for Brown had fallen relative to its neighbors, while there was a temporary rise in corn basis for Edmunds which was not sustained in the later years. This observation goes to provide some understanding as to why the ethanol plants yielded non-positive treatment effects for TE and Cluster 2. Note that our claims are not founded here on robust statistical tools (like regressions), but only on some summary statistics. Our purpose here is to only garner some understanding and support the quasi-experimental design of this study.

Discussion and Conclusions

Robustness Checks

➔ *Multiple Treatment and Control Groups*

As discussed earlier, the advent of an ethanol plant is not a centrally-administered change/policy but an outcome of perceived supply and demand gap in its input and output markets. Because the treatment for this study is non-exogenous, the control groups are also non-exogenous. As per their definitions, the only requirement for a control group is that it is more distant than a treatment group. Since the treatment groups are ad-hoc, so are the control groups. This calls for robustness checks on our treatment estimates. To incorporate those we include multiple treatment and control groups (see Table 2). We apply the steps of above stated methodology, including propensity score matches, for each combination of treatment and control groups for each ethanol plant. This means that we have estimated treatment effects estimates for each ethanol plant/cluster using four regressions, and hence twenty-four regressions in all. To save space we present the results of one of these combinations. Tables 15-16 summarize all twenty-four regressions by listing the treatment effects estimate for each one of them. We also include the control as per Tables 3-14 but mute their coefficient estimates to save space. Our robustness checks using multiple treatment and control groups reveal that the treatment estimates are generally stable across these combinations. The size and sign of these are especially similar by control group. That is, combinations 'T1 and C2' and 'T2 and C2' will generally yield similar estimates.

➔ *Placebo Tests*

Although we find our treatment estimates to be sufficiently stable, we still have not been able to reconcile negative treatment effects. In an attempt, to do so we present placebo treatment effects in Table 17 and Figure 8 shows the schematic that we follow to conduct these placebos. Ideally, we should have found the placebo treatment effects to be zero or, at least, statistically insignificant. Unfortunately, we do find significant placebo tests pointing out to the fact that either our matching strategy is not perfect or we are not able to control for all the factors that affect growth of corn acres in equation (6). To reconcile the failed placebo tests, we first consider the pre-treatment trends for treatment and control groups for the North Dakota ethanol plants to validate the Parallel Paths assumption of DID estimation strategy (see equation 1). Figure 9 shows that the Parallel Paths assumption has failed and that we need to incorporate differentiated trends between pre- and post-treatment periods and between treatment and control groups. We follow Ricardo and Mora (2012) to model trends into the DID model presented above (see appendix). This model will be further developed as part of future work.

Discussion

The literature on impact of ethanol plant is themed two-way: direct and indirect land use changes. The changes in agricultural prices and farmland values due to presence of ethanol plants are indirect land use impacts, while acreage changes of different land use types is the direct impact. In the U.S., a growing biofuels industry due to favorable policies and national security concern amidst increasing crude oil prices fueled the advent of a large number of ethanol plants. Fulfilling such increased demand of biofuels in the last decade or so required increased supply of corn to the corn-based ethanol industry. An increased demand of corn, in turn, required intensification of corn cropping coupled with higher per acre productivity from improved and more tolerant seed varieties. Since intensified corn cropping would require more land for its plantation, a debate on various impacts of ethanol plants on land use change is evident in the literature. This debate specifically focusses on whether new lands are brought into growing corn or existing croplands have been shifted towards its production by replacing other crops. This issue is especially relevant to the Dakotas which have primarily been perennial native grasslands supporting the livestock industry, but experiencing rapid shifts towards corn production along the western corn belt (Johnston, 2014; Wright and Wimberly, 2013). These shifts in production systems present policy-makers with at least three challenges. First, these lands are ecologically important. Continual loss of original mixed grass prairie surrounding pothole lakes poses serious threats to feeding and nesting habitat for migratory waterfowl, songbirds, many insects and their predators. Second, the Dakotas PPR offers marginal lands for row crop production with high drought/floods risks and low productivity. Historically, cropped land in the area has proven costly to agricultural support programs (Johnston, 2014; Wright & Wimberly, 2013). Third, higher conversion rates imply lesser scope for growing perennial bioenergy crops that would limit the scope of supporting the cellulosic ethanol industry in future (Wright & Wimberly 2013).

Apart from above policy related issues, intensified corn production can have serious concerns of environmental degradation and soil erosion. First, employing new lands for corn production by converting native grasslands releases carbon stored in them, thereby increasing greenhouse gas emissions. Second, more acres for corn production imply higher application of nutrients, nitrogen and phosphorous, which may run-off to the open water streams degrading surface water quality. Third, intensified corn production has implications for corn-soy rotations, shifting very alternate-year rotation strategy towards corn in more periods than soybeans. This could affect land quality by increased soil erosion, higher nutrient run-off affecting surface water quality and sustained aquatic life.

Conclusions

Our evaluation of the role of ethanol plants in Dakotas land use change is an attempt towards addressing these concerns. First, we find evidence that emergence of an ethanol plant disrupts the corn-soy rotations and shifts production systems towards more corn. This raises concerns about sustainable agricultural practices in these states, for farming on already marginal lands with plantation strategies that further degrade land quality can be detrimental to their productivity. Second, we find that increased corn production due to ethanol plants is a plant-specific issue. In some cases, especially outside the Corn Belt, we find the role of ethanol plants towards land use change rather limited. Here, grasslands are inhibiting factors towards corn production and high wheat acres encourage it. Whereas, for the South Dakota ethanol plants, that started operations within the Corn Belt, both grass and wheat acres encourage corn production. This suggests that a technological and informational divide among two regions such that within WCB, where there is a lot of corn to begin with, conversion costs are lesser than those outside of it. However, our results also suggest that switching from wheat to corn is less costly than converting from grass to corn. Such results imply that switching from other crops will precede conversion of grasslands towards corn production on these marginal lands, especially outside the Corn Belt. However, the policy concerns of loss of habitat, polluted surface water due to nutrient run-off and lower productivity due to soil erosion still hold valid for the Dakotas.

Ideas for future research in this area

This article provides a novel research design that incorporates remotely sensed data into applied economic analyses, especially those under the ambit of quasi-experimental studies. However, our design has its own shortcomings that provide opportunities for future research. First, we use Euclidean distances rather than 'actual' distances of land parcels from ethanol plants. These 'actual' distances using local road networks provided by the state Departments of Transportation can be incorporated using the 'Nearest Facility Analysis' tool on ArcGIS. Second, our research design uses ad-hoc treatment and control groups with our placebo tests suggesting that our matching strategies were not perfect. The placebo tests also suggest, in some cases, that treated parcels had lower rates of growth in corn acres than their untreated parcels. This observation leads to a question that why would the ethanol plants, in the first place, locate in regions where growth in corn production is initially lower. This is possible due to multiple reasons. The location of ethanol plants would definitely depend on ex-ante land use patterns in its proximity, but that is not the only factor that it considers. These plants would also want to consider areas with good public infrastructure, easy access to elevators and other market terminals while making location decision. In any case, such results can provide a springboard for researchers whose interest lies with understanding the effects of ethanol plants on the socio-economic environment in its proximity.

References

Abadie, A. 2005. “Semiparametric Difference-in-Difference Estimators,” *Review of Economic Studies* 72(1):1-19.

Bertrand. M., Duflo, E., Mullainathan, S. 2004. “How Much Should We Trust Difference-In-Difference Estimates,” *Quarterly Journal of Economics* 119(1):249-275.

Brown, J.C., Hanley, E., Bergtold, J.S., Caldas, M., Barve, V., Peterson, D, L. 2014. “Ethanol plant location and intensification vs. extensification of corn cropping in Kansas,” *Applied Geography* 52:141-148.

Claassen, Roger, Joseph C. Cooper, and Fernando Carriazo. 2011. “Crop Insurance, Disaster Payments, and Land Use Change: The Effect of Sodsaver on Incentives for Grassland Conversion,” *Journal of Agricultural and Applied Economics* 43 (2): 195–211.

Du, X., Hennessy, D.A., Edwards, W.M. 2007. “Determinants of Iowa Cropland Cash Rental Rates: Testing Ricardian Rent Theory,” *Working Paper 07-WP 454, Center for Agricultural and Rural Development, Iowa State University.*

Fraeman, K.H. 2010. “An Introduction to Implementing Propensity Score Matching With SAS®,” *Northeast SAS Users Group Conference Proceedings, Baltimore, M.D.*

Henderson, J., Gloy, B.A. 2009. “The Impact of Ethanol Plants on Cropland Values in the Great Plains,” *Agricultural Finance Review* 69(1): 36–48.

Johnston, C.A. 2014. “Agricultural Expansion: Land Use Shell Game in the U.S. Northern Plains,” *Landscape Ecology* 29(1):81-95.

Katchova, A.L. 2009. “The Spatial Effect of Ethanol Biorefinery Locations on Local Corn Prices,” Selected paper, *2009 Agricultural and Applied Economics Association Meeting, Milwaukee, WI*.

Lark, T.J., Salmon, J.M., Gibbs, H.K. 2015. “Cropland Expansion Outpaces Agricultural and Biofuel Policies in the United States,” *Environmental Research Letters* 10(4):044003.

Lewis, K.E. 2010. “The Impact of Ethanol on Corn Market Relationships and Corn Price Basis Levels,” *M.S. Thesis, Department of Agricultural, Food, and Resource Economics, Michigan State University*.

List, J.A., Millimet, D.L., Fredriksson, PG., McHone, W.W. 2003 “Effects of Environmental Regulations on Manufacturing Plant Births: Evidence from a Propensity Score Estimator,” *The Review of Economics and Statistics* 85(4): 944-952.

McNew, K., Griffith, D. 2005. “Measuring the Impact of Ethanol Plants on Local Grain Prices,” *Review of Agricultural Economics* 27(2): 164–180.

Meyer, B.D. 1995. “Natural and Quasi-Experiments in Economics,” *Journal of Business and Economic Statistics* 13(2): 151-161.

Miao, R. 2013. “Impact of Ethanol Plants on Local Land Use Change,” *Agricultural and Resource Economics Review 42(2):291-309.*

Mora, R., Reggio I. 2012. “Treatment Effect Identification Using Alternative Parallel Assumptions,” *UC3M Working Paper 12-33, Economic Series (48).*

Motamed, M.J., McPhail, L.L. 2011. “Spatial Dimensions of US Crop Selection: Recent Responses to Markets and Policy,” Selected Paper, *2011 AAEA & NAREA Joint Annual Meeting, Pittsburgh, PA.*

Mueller, S., Copenhaver, K. 2009. “Determining the Land Use Impact of Two Midwestern Corn Ethanol Plants,” *Energy Resources Center, University of Illinois at Chicago*. Retrieved from: http://www.erc.uic.edu/assets/pdf/twoplantlucstudyEPASubmissionv3.pdf

O’Brien, D. 2009. “The Effects of the Micro-Market Structure for Kansas Grain Elevators on Spatial Grain Price Differentials,” *NCCC-134 Applied Commodity Price Analysis, Forecasting and Market Risk Management Conference, St. Louis, Missouri*.

Reber, S.J. 2005. “Court-Ordered Desegregation Success and Failures Integrating American Schools since Brown versus Board of Education,” *Journal of Human Resources 40(3): 559-590.*

Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey. Available online at http://websoilsurvey.nrcs.usda.gov/. Accessed 01/15/2015.

Stephens, S.E., Rotella, J.J., Lindberg, S., Taper, M.L., Ringelman, J.K. 2005. “Duck Nest Survival in the Missouri Coteau of North Dakota: Landscape effects at multiple spatial scales,” *Ecological Applications* 15: 2137-2149.

Stephens, S.E., Walker, J.A., Blunck, D.R., Jayaraman, A., Naugle, D.E., Ringelman, J.K., Smith A.J. 2008. “Predicting Risk of Habitat Conversion in Native Temperate Grasslands,” *Conservation Biology* 22(5):1320-1330.

Turnquist, A., Fortenbery, R.T., Foltz, J. 2008. “Progress or Devastation? The Effects of Ethanol Plant Location on Local Land Use,” Selected paper, *2008 American Agricultural Economics Association Annual Meeting, Orlando, FL*.

USDA National Agricultural Statistics Service Cropland Data Layer. 1997-2013. Published crop-specific data layer [Online]. Available at http://nassgeodata.gmu.edu/CropScape/ (accessed 01/15/2015; verified 01/15/2015). USDA-NASS, Washington, DC

USDA National Agricultural Statistics Service Economics, Statistics and Market Information System. 2006-2010. "Agricultural Land Values". (http://usda.mannlib.cornell.edu/MannUsda/viewDocumentInfo.do?documentID=1446).

USDA National Agricultural Statistics Service Quick Stats. 2012. http://quickstats.nass.usda.gov/).

United States Department of Agriculture – Natural Resources Conservation Service, "Soil Data Viewer 6.0 User Guide", Document Version: 2, 2011.

Wright, C.K., Wimberly, M.C. 2013. "Recent land use change in the Western Corn Belt threatens grasslands and wetlands," *Proceedings of the National Academy of Sciences USA*, 110:4134-4139.

Zhang, W., Irwin, E.G., Nickerson, C. 2012. "The Expanding Ethanol Market and Farmland Values: Identifying the changing influence of proximity to agricultural delivery points," Selected Paper, *2012 American Agricultural Economics Association Annual Meeting, Seattle, WA.*

TABLES

Table 1: List of Ethanol Plants in North Dakota and South Dakota for our analysis

S. No.	Ethanol Plant	Year Established	Capacity (Million gallons per year)	Location
North Dakota				
1	Red Trail Energy	2007	50	Richardton, Stark County
2	Blue Flint Ethanol	2007	65	Underwood, McLean County
3	Tharaldson Ethanol LLC	2006	153	Casselton, Cass County
4	Hankinson Renewable Energy	2008	145	Hankinson, Richland County
South Dakota				
1	POET Bio refinery (POET)	2008	110	Chancellor, Turner County
2	NuGen Energy (NuGen)	2008	100	Marion, Turner County
3	Advanced Bio Energy (ABE)	2008	53	Aberdeen, Brown County
4	Glacial Lakes Energy (GLE)	2008	100	Mina, Edmunds County

Table 2: Schematics of the treatment and control groups of ethanol plants analyzed in this article.

Ethanol Plant	**T1**	**T2**	**C1**	**C2**
RTE	5km-35km South	10km-40km South	50km-80km South	70km-100km South
BF	5km-35km South	10km-40km South	50km-80km South	70km-100km South
TE	5km-35km West	10km-40km West	50km-80km West	70km-100km West
HRE	5km-35km West	10km-40km West	50km-80km West	70km-100km West
POET & NuGen	5km-35km West of POET*	25km-55km West of POET*	70km-100km West of POET*	90km-120km West of POET*
ABE & GLE	5km-35km West of ABE*	25km-55km West of ABE*	70km-100km West of ABE*	90km-120km West of ABE*

* GLE lies ~30 km west of ABE – the location of T & C groups can be visualized accordingly.

<u>Notes on Planar Dimensions of our Treatment and Control Rectangles (Part of Table 2):</u>

- <u>Red Trail Energy & Blue Flint Ethanol</u>: **30 km N-S X 50 km E-W**.
- <u>Tharaldson Ethanol</u>: **30 km E-W X 50 km N-S**.
- <u>Hankinson Renewable Energy</u>: **30 km E-W X 40 km N-S**. North Dakota State Boundary is located 15 km south of this ethanol plant. So the N-S dimensions are chosen to be: 30 km N to 10 km S of the ethanol plant, resulting in length of one side of the rectangles be 40 km (N-S).
- <u>Cluster (POET and NuGen)</u>: 30 km E-W X 40 km N-S RECTANGLES (25 km N + 15 km S). The rectangles included exclude a circle of radius 2.5 km from NuGen, to avoid permanent development in land use characterization.
- Cluster (ABE and GLE): 30 km E-W X 50 km N-S RECTANGLES. The rectangles here exclude a circle of radius 7 km from GLE to avoid a big water pond in land use characterization.

Table 3: Summary Statistics for Red Trail Energy, C2 and T1 combination. Caliper = 0.01.

	C2			T1		
	Mean	Std. dev.	N	mean	Std. dev.	N
C_{i,t^-}	2.04	3.01	492	1.47	2.39	492
C_{i,t^+}	9.90	16.88	492	10.84	18.03	492
CS_{i,t^-}	2.23	3.10	492	1.62	2.48	492
CS_{i,t^+}	10.00	17.00	492	10.88	18.04	492
$W_{i,2006}$	121.13	106.28	492	186.96	123.47	492
$G_{i,2006}$	327.84	120.18	492	246.85	127.59	492
LCC2	58.16	41.95	492	56.54	41.15	492
Slope5	16.44	31.33	492	15.57	31.25	492

Table 4: Treatment Effects' Estimates with Heteroskedasticity corrected t-stats in parentheses

	$C_{i,t^+} - C_{i,t^-}$	$CS_{i,t^+} - CS_{i,t^-}$	$\ln(C_{i,t^+}) - \ln(C_{i,t^-})$	$\ln(CS_{i,t^+}) - \ln(CS_{i,t^-})$
Treatment	-2.92	-2.93	-0.64	-0.61
	(2.78)***	(2.77)***	(1.91)*	(1.83)*
$W_{i,2006}$	-0.02	-0.02	-0.00	-0.00
	(1.06)	(0.98)	(0.28)	(0.37)
$G_{i,2006}$	-0.07	-0.07	-0.01	-0.01
	(4.35)***	(4.27)***	(4.20)***	(4.48)***
Constant	32.92	32.17	3.41	3.50
	(4.43)***	(4.34)***	(2.48)**	(2.56)**
R^2	0.16	0.16	0.08	0.09

* p<0.1; ** p<0.05; *** p<0.01

Table 5: Summary statistics for Blue Flint Ethanol: C2 and T1 combination. Caliper = 0.01

	C2			T1		
	mean	Std. dev.	N	mean	Std. dev.	N
C_{i,t^-}	4.81	5.70	465	3.02	7.82	497
C_{i,t^+}	21.47	25.00	465	20.25	31.05	497
CS_{i,t^-}	5.68	6.53	465	5.18	9.33	497
CS_{i,t^+}	22.64	26.00	465	24.79	36.46	497
$W_{i,2006}$	132.77	87.92	465	100.71	82.33	497
$G_{i,2006}$	291.92	107.26	465	235.13	124.25	497
LCC2	51.15	41.97	465	53.19	41.27	497
Slope5	21.87	37.53	465	24.16	37.97	497

Table 6: Treatment Effects' Estimates with Heteroskedasticity corrected t-stats in parentheses

	$C_{i,t^+} - C_{i,t^-}$	$CS_{i,t^+} - CS_{i,t^-}$	$\ln(C_{i,t^+}) - \ln(C_{i,t^-})$	$\ln(CS_{i,t^+}) - \ln(CS_{i,t^-})$
Treatment	-4.41	-3.53	0.50	-0.22
	(2.69)***	(1.95)*	(1.62)	(0.78)
$W_{i,2006}$	-0.02	-0.03	0.01	0.01
	(1.16)	(1.34)	(3.00)***	(3.26)***
$G_{i,2006}$	-0.12	-0.14	-0.01	-0.01
	(6.98)***	(6.49)***	(4.45)***	(5.56)***
Constant	55.82	62.28	1.70	1.87
	(7.02)***	(6.44)***	(2.47)**	(3.07)***
R^2	0.25	0.24	0.08	0.10

* p<0.1; ** p<0.05; *** p<0.01

Table 7: Summary statistics for Tharaldson Ethanol: C2 and T1 combination. Caliper = 0.01

	C2			T1		
	mean	Std. dev.	N	mean	Std. dev.	N
C_{i,t^-}	15.04	12.55	698	35.83	30.87	698
C_{i,t^+}	71.00	46.26	698	119.11	56.44	698
CS_{i,t^-}	97.97	43.65	698	206.04	66.16	698
CS_{i,t^+}	253.06	89.42	698	334.52	88.67	698
$W_{i,2005}$	129.09	101.59	698	112.36	86.51	698
$G_{i,2005}$	96.98	77.90	698	57.33	55.20	698
LCC2	93.08	20.69	698	92.69	21.00	698
Slope5	96.65	13.52	698	99.93	0.00	698

Table 8: Treatment Effects' Estimates with Heteroskedasticity corrected t-stats in parentheses

	$C_{i,t^+} - C_{i,t^-}$	$CS_{i,t^+} - CS_{i,t^-}$	$\ln(C_{i,t^+}) - \ln(C_{i,t^-})$	$\ln(CS_{i,t^+}) - \ln(CS_{i,t^-})$
Treatment	16.25	-38.97	-0.21	-0.44
	(7.10)***	(12.92)***	(4.17)***	(26.65)***
$W_{i,2005}$	-0.02	0.11	0.00	0.00
	(1.90)*	(6.68)***	(5.43)***	(9.36)***
$G_{i,2005}$	-0.27	-0.36	-0.00	0.00
	(16.36)***	(13.35)***	(3.11)***	(2.83)***
Constant	85.23	175.80	1.55	0.80
	(27.56)***	(40.82)***	(21.41)***	(28.91)***
R^2	0.23	0.26	0.06	0.37

* p<0.1; ** p<0.05; *** p<0.01

Table 9: Summary statistics for Hankinson Renewable Energy: C2 and T1 combination. Caliper = 0.01.

	C2			T1		
	mean	Std. dev.	N	mean	Std. dev.	N
C_{i,t^-}	89.48	56.75	483	95.84	64.90	483
C_{i,t_1^-}	101.41	61.15	483	112.65	70.02	483
C_{i,t^+}	123.98	69.46	483	153.49	76.88	483
CS_{i,t^-}	177.33	86.60	483	244.64	102.90	483
CS_{i,t_1^-}	210.96	98.06	483	265.25	113.34	483
CS_{i,t^+}	237.04	116.15	483	316.26	125.03	483
$W_{i,2007}$	19.66	42.19	483	30.27	48.69	483
$G_{i,2007}$	99.17	102.15	483	60.84	93.74	483
LCC2	63.66	43.97	483	63.63	44.43	483
Slope5	98.58	9.88	483	98.29	10.85	483

Table 10: Treatment Effects' Estimates with Heteroskedasticity corrected t-stats in parentheses

	$C_{i,t^+} - C_{i,t^-}$	$C_{i,t^+} - C_{i,t_1^-}$	$\ln(C_{i,t^+}) - \ln(C_{i,t^-})$	$\ln(C_{i,t^+}) - \ln(C_{i,t_1^-})$
Treatment	18.29	14.48	0.28	0.22
	(7.16)***	(5.56)***	(6.66)***	(4.99)***
$W_{i,2007}$	0.18	0.22	0.00	0.00
	(6.46)***	(8.15)***	(12.11)***	(11.12)***
$G_{i,2007}$	-0.08	-0.04	0.00	0.00
	(6.56)***	(3.01)***	(0.30)	(0.50)
Constant	38.86	21.89	0.17	0.03
	(16.77)***	(8.78)***	(3.67)***	(0.48)
R^2	0.16	0.12	0.13	0.09
	$CS_{i,t^+} - CS_{i,t^-}$	$CS_{i,t^+} - CS_{i,t_1^-}$	$\ln(CS_{i,t^+}) - \ln(CS_{i,t^-})$	$\ln(CS_{i,t^+}) - \ln(CS_{i,t_1^-})$
Treatment	4.82	20.27	0.06	0.19
	(1.39)	(6.12)***	(1.85)*	(5.43)***
$W_{i,2007}$	0.26	0.35	0.00	0.00
	(7.55)***	(10.00)***	(6.92)***	(7.72)***
$G_{i,2007}$	-0.11	-0.03	-0.00	-0.00
	(6.74)***	(1.57)	(0.88)	(0.79)
Constant	65.81	21.82	0.21	0.00
	(20.42)***	(6.84)***	(4.81)***	(0.05)
R^2	0.12	0.15	0.02	0.04

* p<0.1; ** p<0.05; *** p<0.01

Table 11: Summary statistics for PBNE, C2 and T1 combination. Caliper = 0.01

	C2			T1		
	mean	Std. dev.	N	mean	Std. dev.	N
C_{i,t^-}	123.93	49.73	521	193.99	68.39	521
C_{i,t_1^+}	127.76	49.46	521	193.99	61.92	521
C_{i,t^+}	145.81	46.54	521	210.76	60.85	521
CS_{i,t^-}	231.92	90.81	521	335.44	90.90	521
CS_{i,t_1^+}	264.96	88.24	521	357.35	86.70	521
CS_{i,t^+}	287.18	84.16	521	374.15	83.13	521
$W_{i,2007}$	58.85	66.17	521	21.01	35.61	521
$G_{i,2007}$	148.67	86.19	521	85.23	69.36	521
LCC2	98.96	6.51	521	98.89	6.73	521
Slope5	94.94	15.63	521	94.77	15.54	521

Table 12: Treatment Effects' Estimates with Heteroskedasticity corrected t-stats in parentheses

	$C_{i,t_1^+} - C_{i,t^-}$	$C_{i,t^+} - C_{i,t^-}$	$\ln(C_{i,t_1^+}) - \ln(C_{i,t^-})$	$\ln(C_{i,t^+}) - \ln(C_{i,t^-})$
Treatment	12.25	11.65	0.12	0.09
	(4.06)***	(4.30)***	(4.85)***	(4.06)***
$W_{i,2007}$	0.31	0.35	0.003	0.003
	(10.68)***	(14.66)***	(7.83)***	(11.25)***
$G_{i,2007}$	0.07	0.05	0.001	0.001
	(4.28)***	(3.65)***	(2.89)***	(5.73)***
Constant	-24.51	-6.84	-0.21	-0.13
	(6.62)***	(2.09)**	(6.23)***	(4.40)***
R^2	0.14	0.19	0.36	0.27

	$CS_{i,t_1^+} - CS_{i,t^-}$	$CS_{i,t^+} - CS_{i,t^-}$	$\ln(CS_{i,t_1^+}) - \ln(CS_{i,t^-})$	$\ln(CS_{i,t^+}) - \ln(CS_{i,t^-})$
Treatment	13.81	10.05	0.05	0.04
	(4.42)***	(3.29)***	(3.69)***	(2.73)***
$W_{i,2007}$	0.54	0.56	0.003	0.003
	(14.88)***	(15.83)***	(10.02)***	(10.23)***
$G_{i,2007}$	0.07	0.09	0.001	0.001
	(4.16)***	(5.17)***	(6.50)***	(9.60)***
Constant	-9.33	9.60	-0.10	-0.06
	(2.43)**	(2.59)***	(5.27)***	(3.45)***
R^2	0.32	0.36	0.31	0.40

* p<0.1; ** p<0.05; *** p<0.01

Table 13: Summary statistics for ABGL, C2 and T1 combination. Caliper = 0.01.

	C2			T1		
	mean	Std. dev.	N	mean	Std. dev.	N
C_{i,t^-}	51.34	50.71	425	116.48	80.99	425
C_{i,t_1^+}	50.29	50.44	425	58.69	46.15	425
C_{i,t^+}	63.10	52.66	425	108.67	51.91	425
CS_{i,t^-}	75.02	71.86	425	216.20	124.78	425
CS_{i,t_1^+}	89.65	83.76	425	128.65	86.06	425
CS_{i,t^+}	106.85	89.68	425	215.78	97.66	425
$W_{i,2007}$	67.18	79.79	425	47.50	75.42	425
$G_{i,2007}$	315.95	134.59	425	187.07	137.86	425
LCC2	66.56	43.85	425	66.50	44.03	425
Slope5	77.69	39.68	425	99.93	0.00	425

Table 14: Treatment Effects' Estimates with Heteroskedasticity corrected t-stats in parentheses

	$C_{i,t_1^+} - C_{i,t^-}$	$C_{i,t^+} - C_{i,t^-}$	$\ln(C_{i,t_1^+}) - \ln(C_{i,t^-})$	$\ln(C_{i,t^+}) - \ln(C_{i,t^-})$
Treatment	-23.38	7.57	-0.58	0.37
	(6.27)***	(2.26)**	(3.15)***	(2.26)**
$W_{i,2007}$	0.26	0.32	0.00	0.01
	(9.82)***	(12.08)***	(1.87)*	(4.40)***
$G_{i,2007}$	0.22	0.16	0.00	0.01
	(13.01)***	(11.17)***	(1.87)*	(6.92)***
Constant	-87.83	-60.81	-0.67	-1.35
	(11.64)***	(9.42)***	(2.51)**	(6.48)***
R^2	0.39	0.27	0.03	0.07
	$CS_{i,t_1^+} - CS_{i,t^-}$	$CS_{i,t^+} - CS_{i,t^-}$	$\ln(CS_{i,t_1^+}) - \ln(CS_{i,t^-})$	$\ln(CS_{i,t^+}) - \ln(CS_{i,t^-})$
Treatment	-58.15	-2.80	-0.79	-0.06
	(12.37)***	(0.62)	(5.84)***	(0.44)
$W_{i,2007}$	0.40	0.38	0.00	0.00
	(12.01)***	(13.24)***	(2.17)**	(2.55)**
$G_{i,2007}$	0.28	0.17	0.00	0.00
	(13.91)***	(10.00)***	(1.87)*	(6.47)***
Constant	-100.86	-47.54	-0.16	-0.72
	(11.28)***	(6.37)***	(0.93)	(4.05)***
R^2	0.52	0.27	0.06	0.07

* p<0.1; ** p<0.05; *** p<0.01

Table 15: Robustness Checks for treatment effects on Corn Acres. All combinations of multiple treatment and control groups.

Ethanol Plant	Combinations	$\ln(C_{i,t^+})-\ln(C_{i,t^-})$	$\ln(C_{i,t^+})-\ln(C_{i,t_1^-})$	$\ln(C_{i,t_1^+})-\ln(C_{i,t^-})$
Red Trail Energy	T1 and C2	**-0.64***	n/a	n/a
	T2 and C2	**-0.87*****	n/a	n/a
	T1 and C1	**-0.80*****	n/a	n/a
	T2 and C1	**-1.11*****	n/a	n/a
Blue Flint	T1 and C2	**0.50**	n/a	n/a
	T2 and C2	**0.05**	n/a	n/a
	T1 and C1	**0.47****	n/a	n/a
	T2 and C1	**0.33**	n/a	n/a
Tharaldson Ethanol	T1 and C2	**-0.21****	n/a	n/a
	T2 and C2	**-0.18*****	n/a	n/a
	T1 and C1	**-0.18*****	n/a	n/a
	T2 and C1	**-0.12****	n/a	n/a
Hankinson Renewable Energy	T1 and C2	**0.28****	**0.22****	n/a
	T2 and C2	**0.34*****	**0.30*****	n/a
	T1 and C1	**0.17*****	**0.12*****	n/a
	T2 and C1	**0.18*****	**0.15*****	n/a
Cluster 1: POET Bio Refinery and NuGen Energy	T1 and C2	**0.09****	n/a	**0.12****
	T2 and C2	**0.09*****	n/a	**0.12*****
	T1 and C1	**0.02**	n/a	**0.02**
	T2 and C1	**0.03***	n/a	**0.03**
Cluster 2: Advanced Bio Energy and Glacial Lakes Energy	T1 and C2	**0.37****	n/a	**-0.58****
	T2 and C2	**-0.25***	n/a	**-0.74*****
	T1 and C1	**0.42****	n/a	**-0.38*****
	T2 and C1	**-0.07**	n/a	**-0.64*****

* p<0.1; ** p<0.05; *** p<0.01; **N/A** means 'not applicable' for the case.

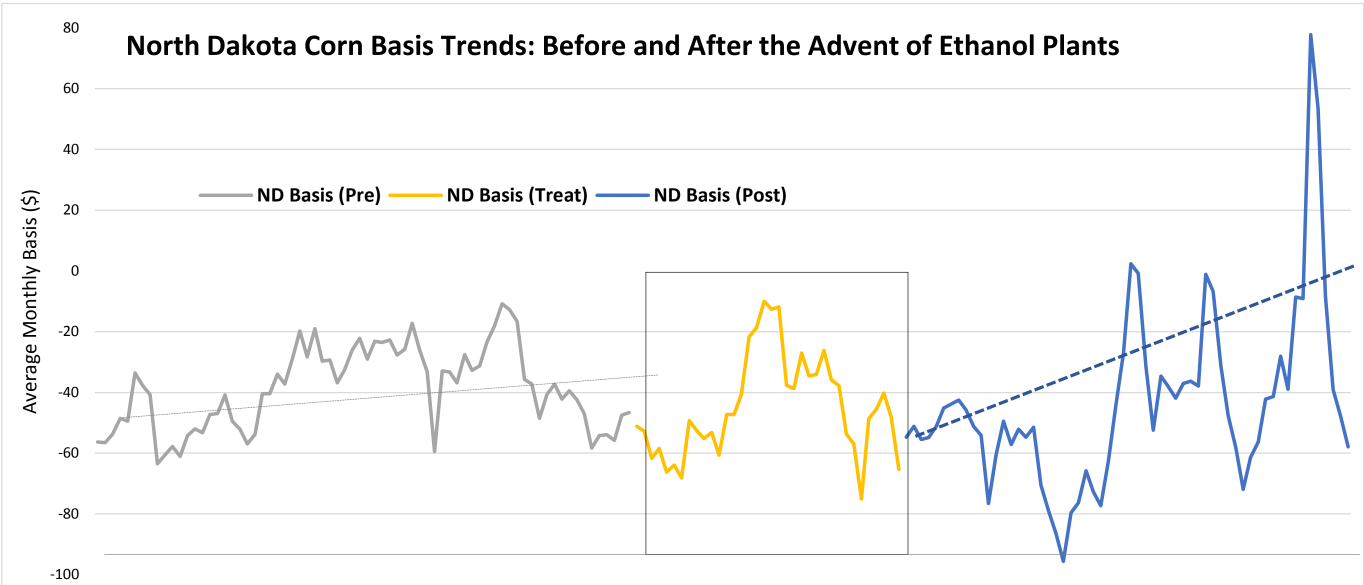
Table 16: Robustness Checks for treatment effects on Corn Acres. All combinations of multiple treatment and control groups.

Ethanol Plant	Combinations	$\ln(CS_{i,t^+}) - \ln(CS_{i,t^-})$	$\ln(CS_{i,t^+}) - \ln(CS_{i,t_1^-})$	$\ln(CS_{i,t_1^+}) - \ln(CS_{i,t^-})$
Red Trail Energy	T1 and C2	**-0.61***	n/a	n/a
	T2 and C2	**-0.88*****	n/a	n/a
	T1 and C1	**-0.83*****	n/a	n/a
	T2 and C1	**-1.11*****	n/a	n/a
Blue Flint	T1 and C2	**-0.22**	n/a	n/a
	T2 and C2	**-0.65****	n/a	n/a
	T1 and C1	**-0.12**	n/a	n/a
	T2 and C1	**-0.22**	n/a	n/a
Tharaldson Ethanol	T1 and C2	**-0.44*****	n/a	n/a
	T2 and C2	**-0.43*****	n/a	n/a
	T1 and C1	**-0.28*****	n/a	n/a
	T2 and C1	**-0.26*****	n/a	n/a
Hankinson Renewable Energy	T1 and C2	**0.06***	**0.19*****	n/a
	T2 and C2	**0.04**	**0.16*****	n/a
	T1 and C1	**0.02**	**0.09*****	n/a
	T2 and C1	**0.01**	**0.09*****	n/a
Cluster 1: POET Bio Refinery and NuGen Energy	T1 and C2	**0.04*****	n/a	**0.05*****
	T2 and C2	**0.04*****	n/a	**0.07*****
	T1 and C1	**0.04*****	n/a	**0.05*****
	T2 and C1	**0.04*****	n/a	**0.06*****
Cluster 2: Advanced Bio Energy and Glacial Lakes Energy	T1 and C2	**-0.005**	n/a	**-0.72*****
	T2 and C2	**-0.31*****	n/a	**-0.69*****
	T1 and C1	**0.08**	n/a	**-0.46*****
	T2 and C1	**-0.25*****	n/a	**-0.63*****

* p<0.1; ** p<0.05; *** p<0.01; **N/A** means 'not applicable' for the case.

FIGURES

Figure1: Comparative Corn Basis Trends for Counties that House Dakotas' Ethanol Plants that started operations in the 2006-2008 period. The acronym 'treat' denotes the period when these ethanol plants started operations, 'pre' ('post') means years prior (after) to the 2006-2008 period.



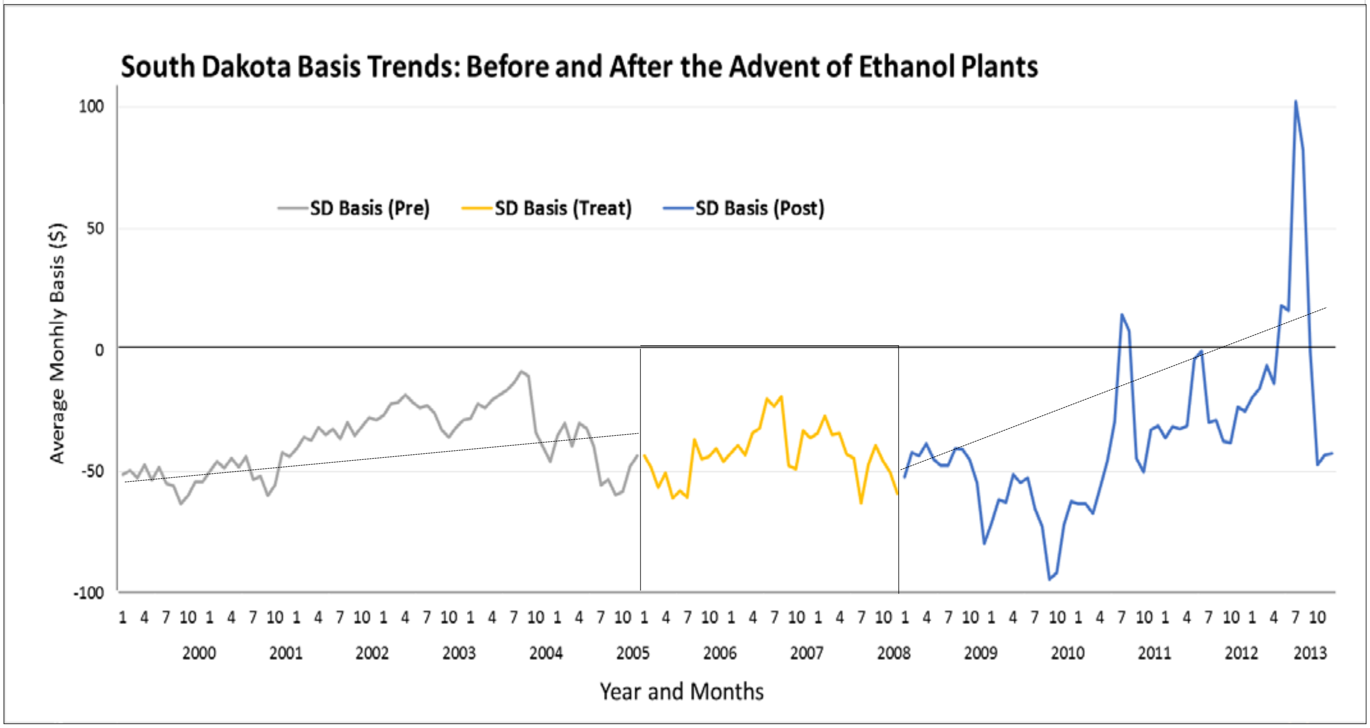
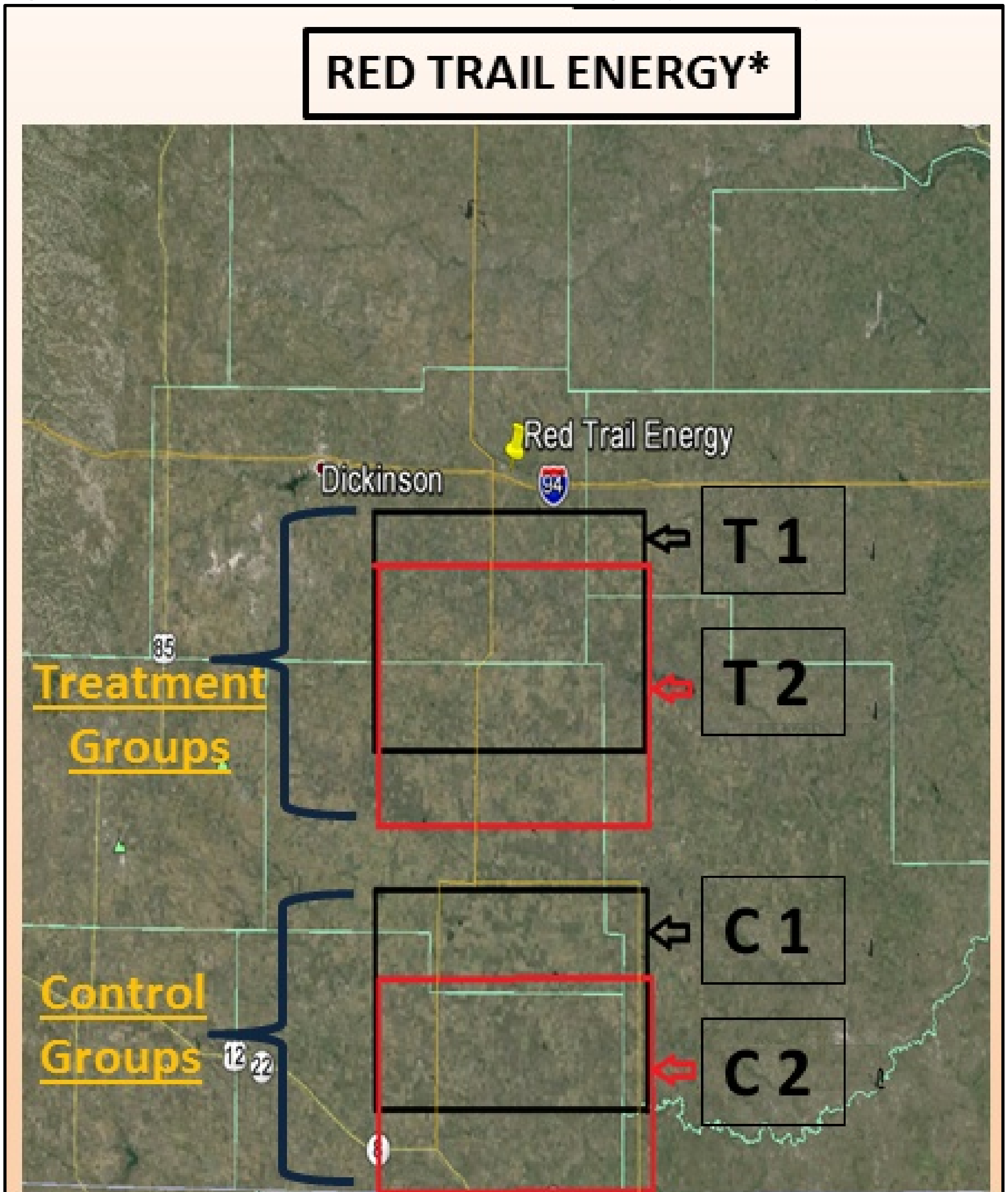


Figure 2: Spatial locations of the 8 ethanol plants included in this analysis

Source: "North and South Dakota." 5122554.70 m N and 393724.99 m E. **Google Earth.** April 9, 2013. April 21, 2015.

Figure 3: Schematics of treatment and control group: An Example



Source: "North Dakota" 337504.21 m E, 5249274.59 m N. **Google Earth.** April 9, 2014. October 20, 2014. * NOT TO SCALE

Figure 4: ***Spatially Comparative* Corn-Basis Trends for *Cass* vs. Neighboring Counties**
Cass is home-county to Tharaldson Ethanol. Established in 2006, Capacity = 153 mgy.

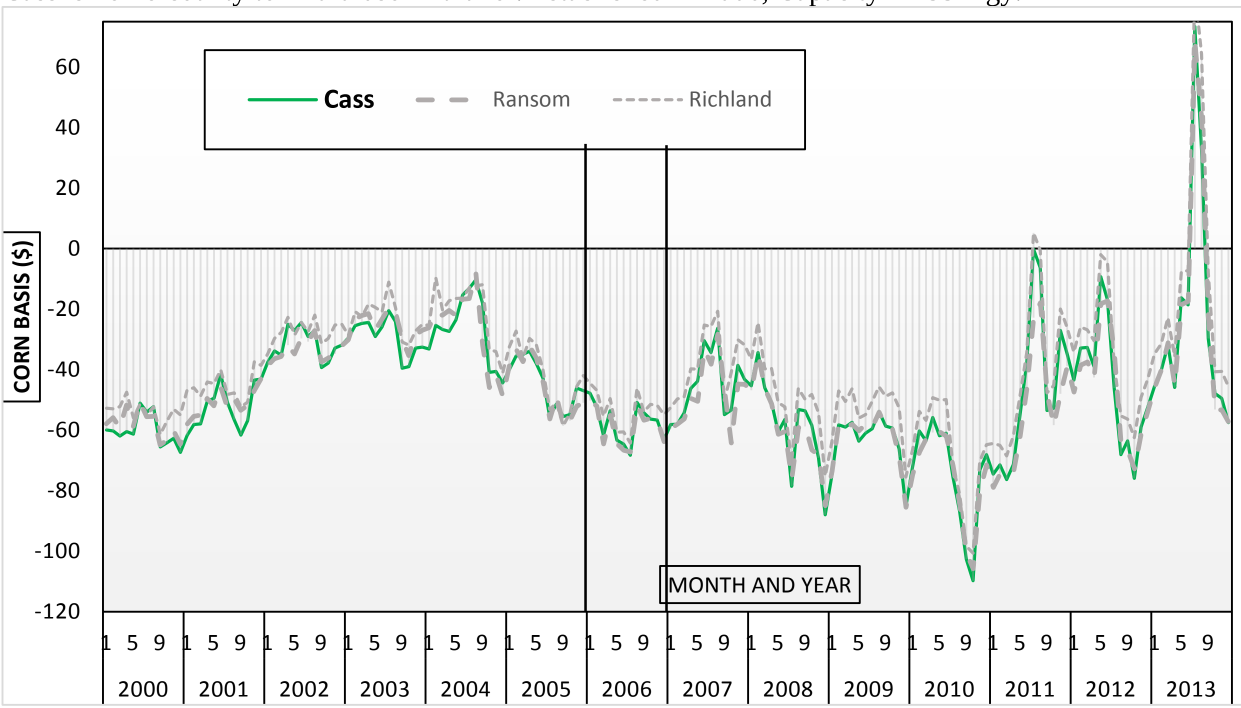


Figure 5: ***Spatially Comparative* Corn-Basis Trends for *Richland* vs. Neighboring Counties**
Richland is home-county to Hankinson Renewable Energy. Established in 2008, Capacity = 145 mgy.

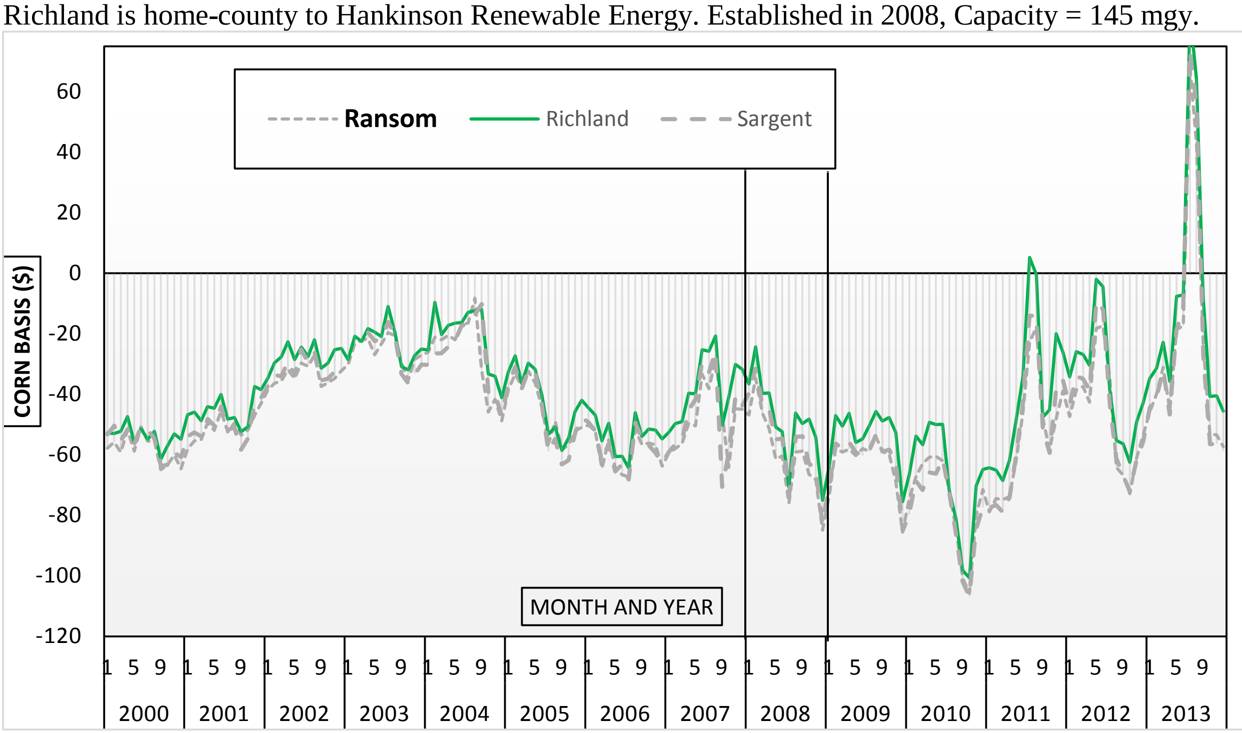


Figure 6: Spatially Comparative **Corn-Basis Trends for *Turner* vs. Neighboring Counties**

Turner is home-county to POET Bio refinery (110 mgy) and NuGen Energy (115 mgy). Established in 2008

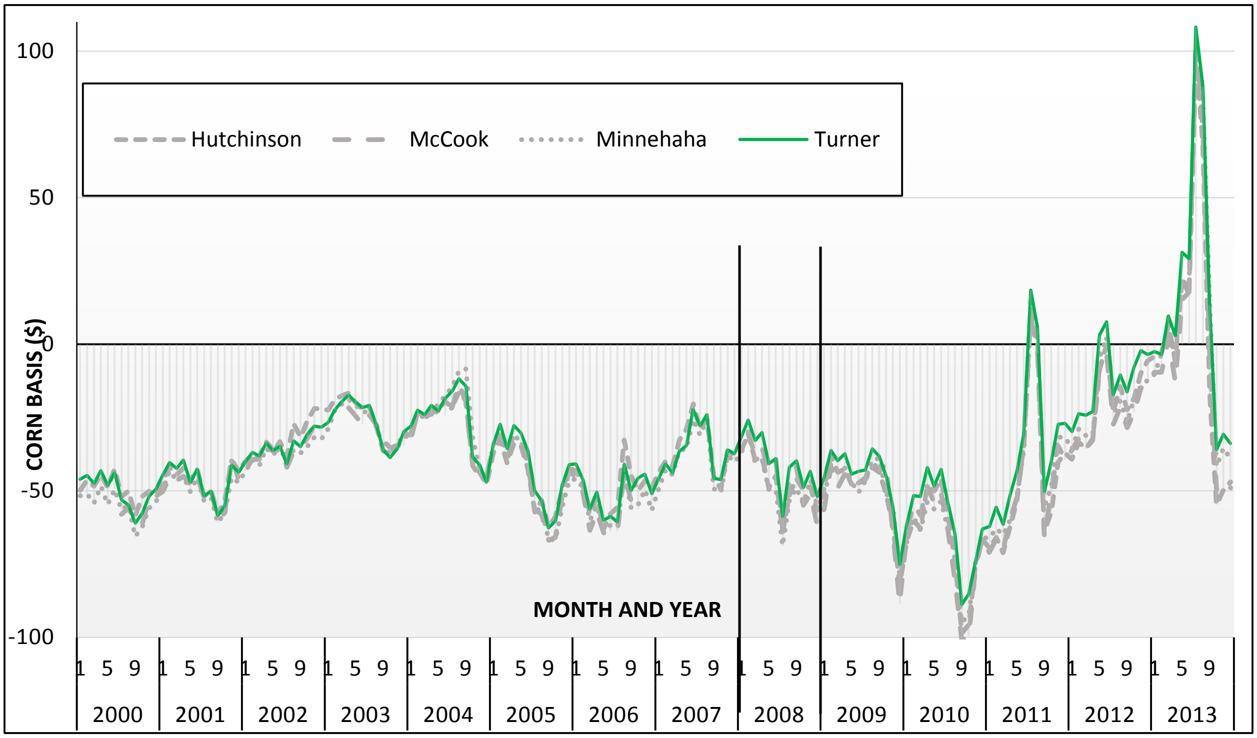
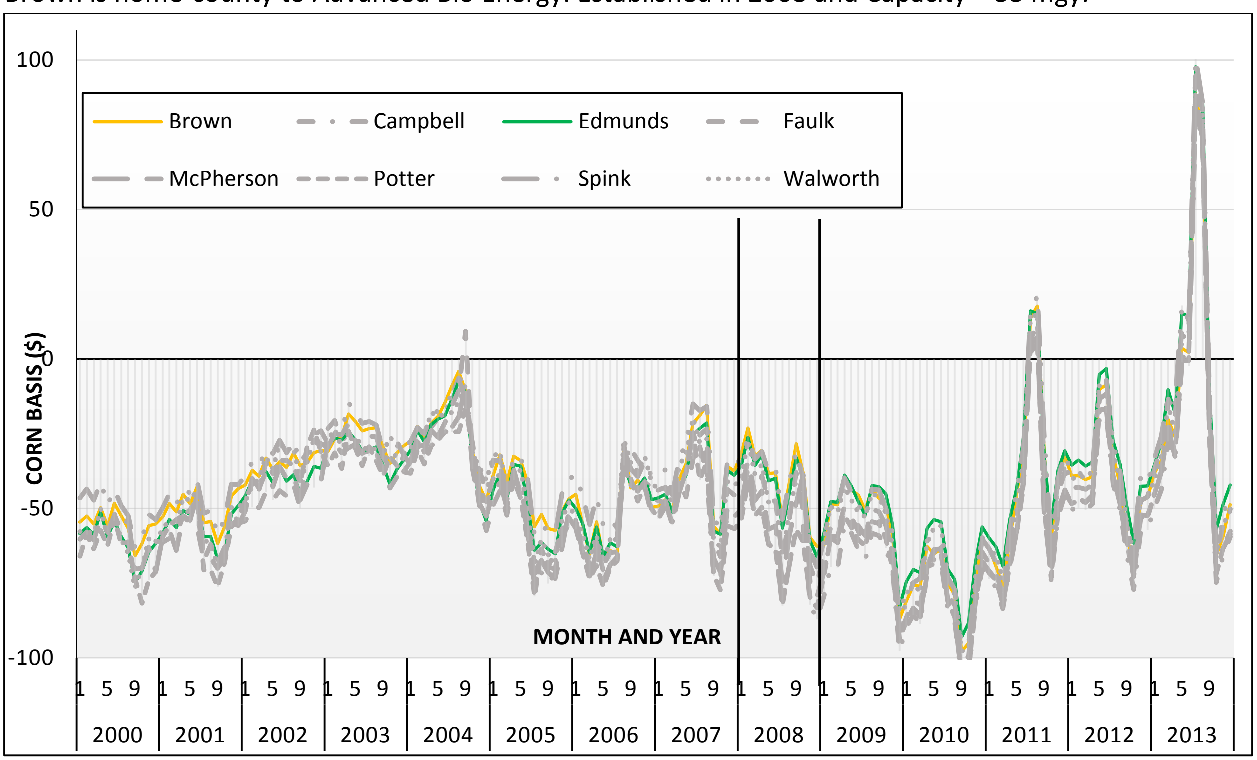


Figure 7: Cluster1: Spatially Comparative Corn-Basis Trends for Edmunds and Brown vs. Neighboring Counties

Edmunds is home-county to Glacial Lakes Energy. Established in 2008 and Capacity = 100 mgy.

Brown is home-county to Advanced Bio Energy. Established in 2008 and Capacity = 53 mgy.



APPENDIX

Moving Away from the Parallel Paths Assumption

Figure 8: Placebo Schematics

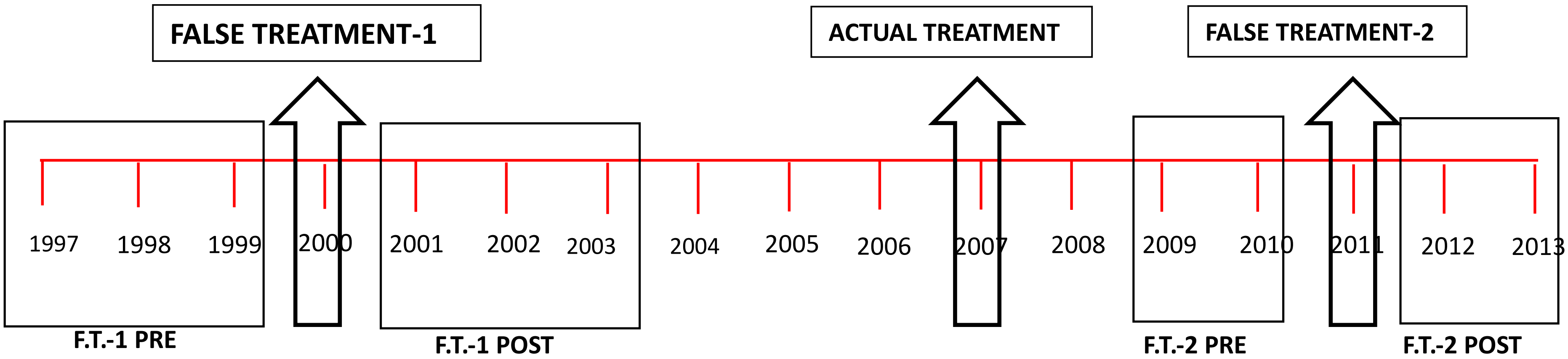


Table 17: Placebo Estimates with 'Logarithm of CS' as dependent variable

	Red Trail Energy	*Blue Flint Ethanol*	*Tharaldson Ethanol*	*Hankinson Renewable Energy*
F.T. – 1 (2000)	*-2.52****	*0.36*	*-0.96****	*-0.29****
ACTUAL TREATMENT	***-0.61****	***-0.22***	***-0.44****	***0.06****
F.T. – 2 (2011)	*0.80**	*-0.39*	*-0.14****	*-0.27****

Figure 9: Pre-Treatment Trends for North Dakota Ethanol Plants

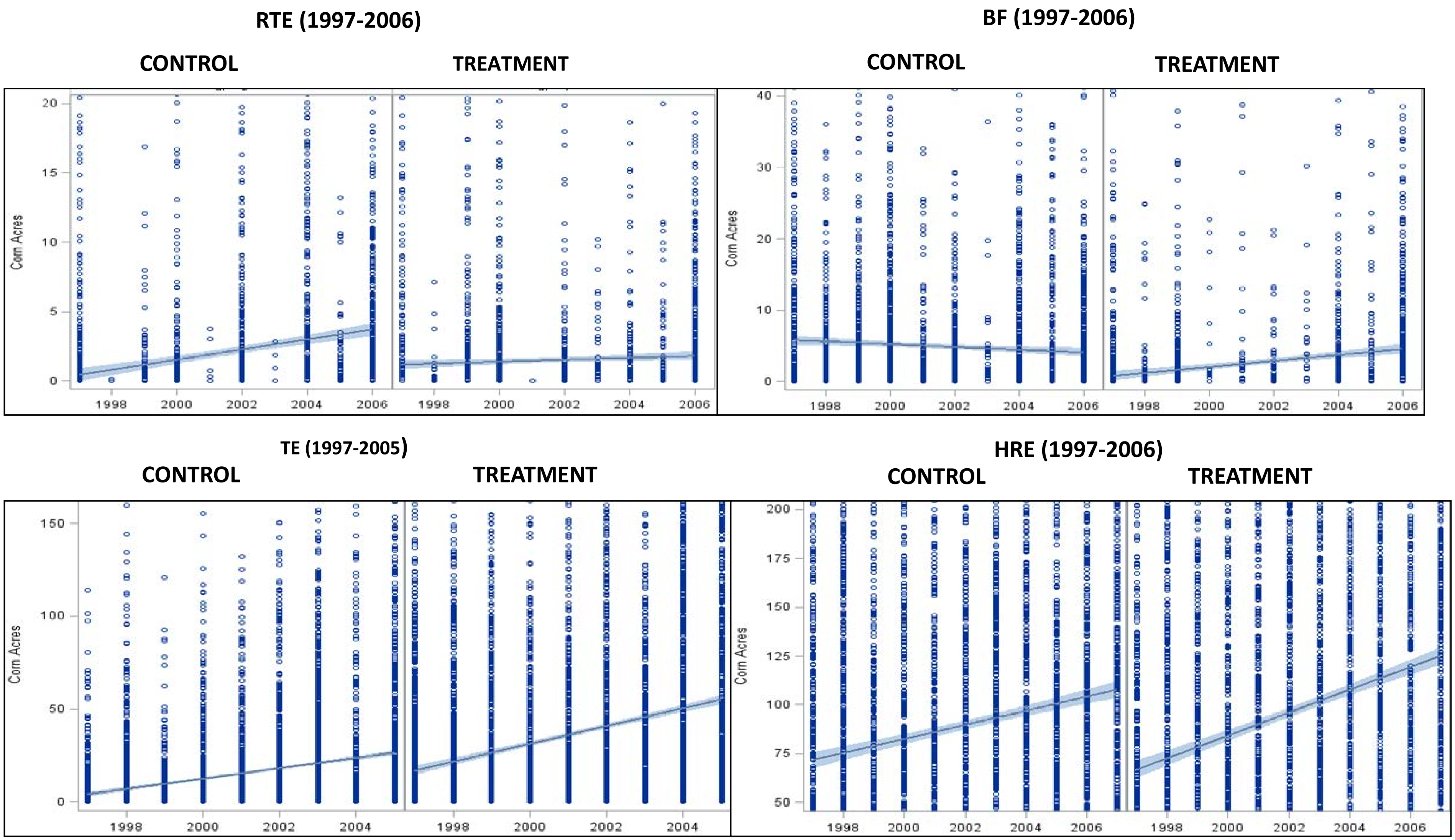
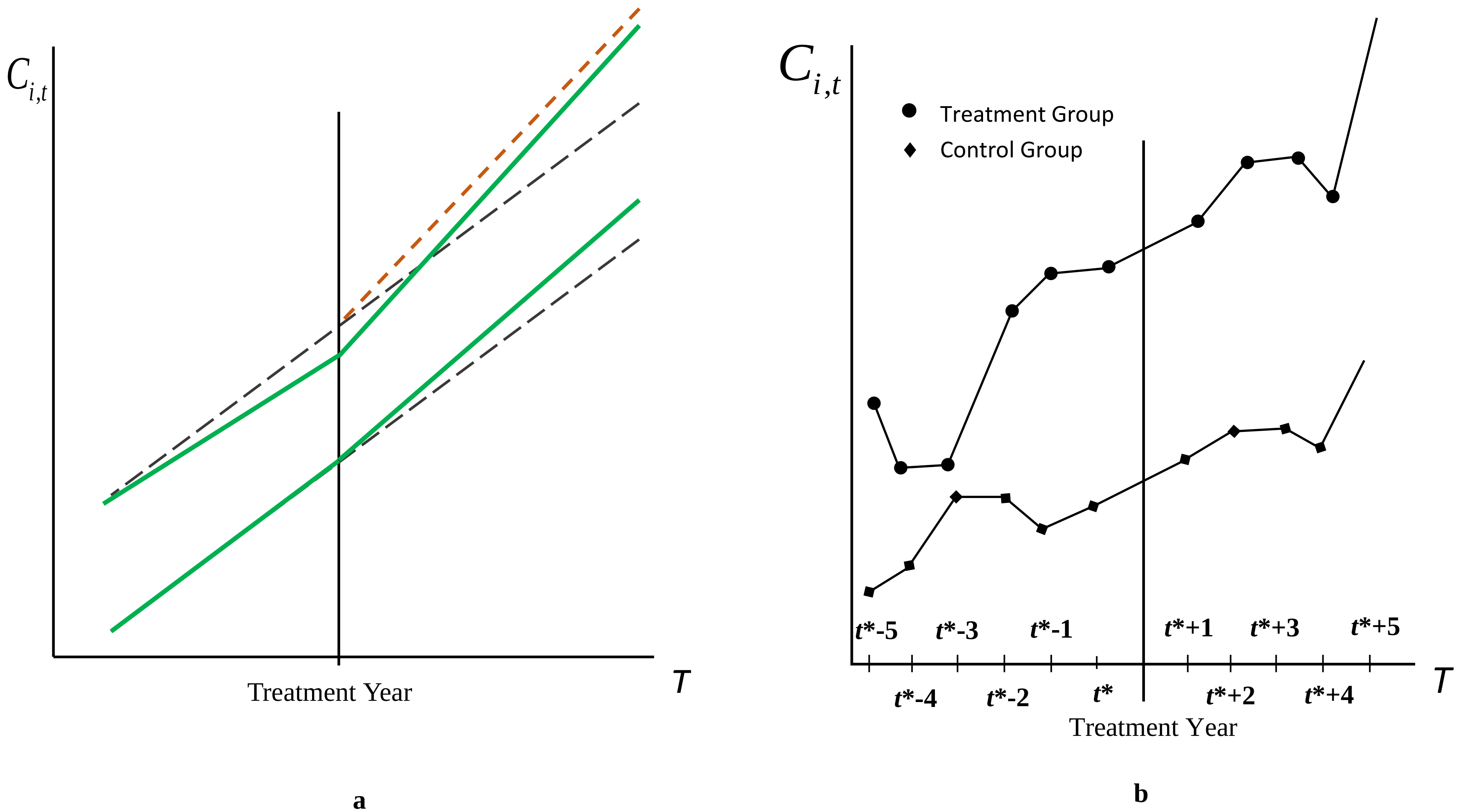


Figure 10 (a, b): The issue of Non-parallel trends among treatment and control groups.



Modelling Differentiated Trends into Our DID Framework

In this section we develop the DID framework to incorporate differentiated trends among treatment and control groups as well as between pre- and post-treatment periods. In the process, we will exploit the variations in corn acres in multiple periods before and after the advent of an ethanol plant. Capturing trends, by interacting trend variables with corresponding group and time-fixed effects of the original DID model, changes interpretation of regression coefficients that estimate the treatment effects and corresponding identification strategies (Mora and Reggio, 2012). We will first explain these implications of failed parallel paths assumption for pre-treatment years (figure 9) and then layout a ‘fully-flexible’ model, originally developed by Mora and Reggio (2012), to capture trends that could vary between different years and among groups. We also discuss a family of identifying assumptions tied to estimating treatment effects under a fully-flexible model. As stated before, this section will serve as the direction our analysis will take in future.

The standard DID framework and the role of Parallel Paths Assumption:

Reconsider our equation(1), that is $C_{i,t} = \beta_0 + \beta_1 d_t + \beta_2 d_i + \beta_3 d_i d_t + \beta_{4,t} Z_i + \varepsilon_{i,t}$, where the definitions of these variables and parameters are same as in the 'Methodology' section above.

Equation (2) suggests that $ATT = E[C_{i,t^+} - C_{i,t^-} \mid Z_i, d_i = 1] - E[C_{i,t^+} - C_{i,t^-} \mid Z_i, d_i = 0]$ and so mechanics of computing the treatment effects using regression equation (1) are as under:

$d_i = 1;\ d_t = 1 \rightarrow E[C_{i,t^+} \mid Z_i] = \beta_0 + \beta_1.1 + \beta_2.1 + \beta_3.1 + \beta_{4,t}\overline{Z_{i|d_i=1}}$,

$d_i = 1;\ d_t = 0 \rightarrow E[C_{i,t^-} \mid Z_i] = \beta_0 + \beta_1.0 + \beta_2.1 + \beta_3.0 + \beta_{4,t}\overline{Z_{i|d_i=1}}$,

$d_i = 0;\ d_t = 1 \rightarrow E[C_{i,t^+} \mid Z_i] = \beta_0 + \beta_1.1 + \beta_2.0 + \beta_3.0 + \beta_{4,t}\overline{Z_{i|d_i=0}}$,

$d_i = 0;\ d_t = 0 \rightarrow E[C_{i,t^-} \mid Z_i] = \beta_0 + \beta_1.0 + \beta_2.0 + \beta_3.0 + \beta_{4,t}\overline{Z_{i|d_i=0}}$. Note that $\overline{Z_{i|d_i}}$ is an unconditional average.

Hence, $ATT = [\beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_{4,t}\overline{Z_{i|d_i=1}} - \beta_0 - \beta_2 - \beta_{4,t}\overline{Z_{i|d_i=1}})] - [\beta_0 + \beta_1 + \beta_{4,t}\overline{Z_{i|d_i=0}} - \beta_0 - \beta_{4,t}\overline{Z_{i|d_i=0}}] = \beta_3$

It is, however, critical to note that by definition the ATT equals $E[C^T_{i,t^+} - C^U_{i,t^+} \mid d_i = 1]$ (where superscripts T (U) represent corn acres in presence (absence) of ethanol plant in $t \in t^+$) and needs the parallel paths assumption to hold for β_3 to estimate the impact of ethanol plants on corn acres. Figure 10 provides a visualization of the implications when parallel paths assumption fails. Basically, this assumption ensures that the treatment and control groups grow in a parallel fashion (grey-dashed lines) and any difference in their trends (orange- versus grey-dashed lines after the treatment year) after the advent of an ethanol plant is purely due to its existence. This difference is then captured by our estimate of β_3. However, in reality it seems that the process that we need to model is better depicted by green-solid lines in figure 10. That is, we are dealing with potentially different pre- and post-treatment trends, and also treatment and control group-specific trends. We incorporate these differences in trends in the standard DID model below.

The DID framework with Differentiated Trends:

We utilize this subsection to motivate the implication of incorporating trends into the standard DID model through a specialized example. We will discuss the mechanics involved in estimating the treatment effects within a new framework, including the underlying identifying assumptions, and show how these are different from the standard case. We will ultimately move towards a generalized model proposed by Mora and Reggio's (2012) working paper, discussing its applicability for our analysis.

To incorporate the differences in trends as depicted by figure 10 a., consider the following econometric model.

(7) $C_{i,t} = \beta_0 + \beta_0' t + \beta_1 d_t + \beta_1' t.d_t + \beta_2 d_i + \beta_2' t.d_i + \beta_3 d_i d_t + \beta_3' t.d_i d_t + \beta_{4,t} Z_i + \varepsilon_{i,t}$,

Where variable t represents time trends such that $t = 1$ for year =1997 (2006) for North (South) Dakota ethanol plants, and it increases by one for each subsequent year. While the standard DID model in equation (1) allows distinct intercepts for treatment/control groups and pre-/post-treatment periods, the updated model in equation (7) allows for distinct linear trends (slopes), as well as intercepts, for these groups and periods. Repeating our exercise of the previous subsection for computing treatment effects from equation (7), we get

$d_i = 1;\ d_t = 1 \rightarrow E[C_{i,t^+} \mid Z_i] = \beta_0 + \beta_0' t + \beta_1 + \beta_1' t + \beta_2 + \beta_2' t + \beta_3 + \beta_3' t + \beta_{4,t} \overline{Z_{i|d_i=1}}$,

$d_i = 1;\ d_t = 0 \rightarrow E[C_{i,t^-} \mid Z_i] = \beta_0 + \beta_0' t + \beta_2 + \beta_2' t + \beta_{4,t} \overline{Z_{i|d_i=1}}$,

$d_i = 0;\ d_t = 1 \rightarrow E[C_{i,t^+} \mid Z_i] = \beta_0 + \beta_0' t + \beta_1 + \beta_1' t + \beta_{4,t} \overline{Z_{i|d_i=0}}$,

$d_i = 0;\ d_t = 0 \rightarrow E[C_{i,t^-} \mid Z_i] = \beta_0 + \beta_0' t + \beta_{4,t} \overline{Z_{i|d_i=0}}$. And again, $\overline{Z_{i|d_i}}$ is an unconditional average.

So, $E[C_{i,t^+} - C_{i,t^-} \mid Z_i, d_i = 1] - E[C_{i,t^+} - C_{i,t^-} \mid Z_i, d_i = 0] = \beta_3 + \beta_3' t,$ which notably changes with *t*.

However, we already know that $\beta_3 + \beta_3' t$ does not identify the ATT due to advent of an ethanol plant. Now see that, if we subtract equation (7) from its one-period lagged counterpart, we have

(8) $\Delta C_{i,t} = \beta_0' + \beta_1' d_t + \beta_2' d_i + \beta_3' d_i d_t + \Delta\beta_{4,t} Z_i + \Delta\varepsilon_{i,t}$,

where $\Delta C_{i,t} = C_{i,t} - C_{i,t-1},\ \Delta\beta_{4,t} = \Delta\beta_{4,t} - \Delta\beta_{4,t-1}$ and $\Delta\varepsilon_{i,t} = \varepsilon_{i,t} - \varepsilon_{i,t-1}$.

It is evident that the mechanics of equation (8) to compute the treatment effects due to the advent of an ethanol plant are similar to that of equation (1), with pertinent differences in notations of variables and parameters. So, our 'new' average treatment effect for the treated (ATT') is given as:

(9) $ATT' = E[\Delta C_{i,t} - \Delta C_{i,t'} \mid Z_i, d_i = 1] - E[\Delta C_{i,t} - \Delta C_{i,t'} \mid Z_i, d_i = 0] = \beta_3' \ \forall\ t \in t^+,\ t' \in t^- \ \&\ t > t'.$

Here, it is important to realize that the interpretation of ATT' is not same as our standard ATT . Expanding the mathematical expression of ATT' from equation (9) gives

(10) $$\begin{aligned}ATT' = \{&E[C_{i,t} - C_{i,t'} \mid Z_i, d_i = 1] - E[C_{i,t} - C_{i,t'} \mid Z_i, d_i = 0]\} - \\ &\{E[C_{i,t-1} - \Delta C_{i,t'-1} \mid Z_i, d_i = 1] - E[C_{i,t-1} - C_{i,t'-1} \mid Z_i, d_i = 0]\} \ \forall \ t \in t^+, \ t' \in t^- \ \& \ t > t'\end{aligned}$$

We can now re-write our 'new' average treatment effect for the treated as a function of ATT,

$ATT'(t,t' \mid Z) = ATT(t,t' \mid Z) - ATT(t-1,t'-1 \mid Z) \triangleq \Delta ATT(t,t' \mid Z) \ \forall \ t \in t^+, \ t' \in t^- \ \& \ t > t'$, which in turn suggests that ATT' measures the impact of treatment as change in the standard treatment effects (ATT) between a specific post-treatment period and a specific pre-treatment period. In the context of ethanol plants, ATT' would measure a one-period change in corn acres from a post-treatment year relative to a one-period counterpart from a pre-treatment year.

One other dimension of our updated DID framework to incorporate trends is an identification assumption. The identification issue with ATT' would, however, remain consistent with the one in the standard DID model. That is, by definition, ATT' equals $E[\Delta C_{i,t}^T - \Delta C_{i,t}^U \mid d_i = 1, Z_i]$, where superscripts T (U) represent corn acres in presence (absence) of ethanol plant in $t \in t^+$. As with the standard DID model, since $\Delta C_{i,t}^U$ is not observed for the post-treatment years, we would need an identification assumption to be able to estimate ATT' as an estimate of β_3' in equation (8) above. This identification assumption for ATT' is a modified version of equation (1) above,

(11) $E[\Delta C_{i,t}^U - \Delta C_{i,t'}^U \mid Z_i, d_i = 1] = E[\Delta C_{i,t}^U - \Delta C_{i,t'}^U \mid Z_i, d_i = 0] \ \forall \ t \in t^+ \ \& \ t' \in t^-$.

Note that the new identifying assumption compares first-differences in outcome levels among treatment and control groups, as opposed to the outcome levels as in the identifying assumption for the standard ATT (see equation (1)). Based on equations (9) and (11) we can term our new estimator as a difference-in-first-difference estimator (following Moora and Reggio, 2012).

An aspect of the updated model and its identifying assumption is that it allows estimating a (change in) treatment effects for each of the multiple post-treatment periods, i.e. for every $t \in t^+$. Alongside, it also allows using multiple pre-treatment years, i.e. each $t' \in t^-$. However, it would suffice to estimate the impact of treatment from the last pre-treatment period, say $t*$. To see this, consider $ATT'(s \mid Z_i)$ defined s periods ahead of $t*$ such that that $t = t' + s$ and $t' = t*$. Hence, the identifying assumption and $ATT'(s \mid Z_i)$ are given by equations (12) and (13) respectively.

(12) $E[\Delta C_{i,t*+s}^U - \Delta C_{i,t*}^U \mid Z_i, d_i = 1] = E[\Delta C_{i,t*+s}^U - \Delta C_{i,t*}^U \mid Z_i, d_i = 0]$.

(13) $ATT'(s \mid Z_i) = E[\Delta C_{i,t*+s} - \Delta C_{i,t*} \mid Z_i, d_i = 1] - E[\Delta C_{i,t*+s} - \Delta C_{i,t*} \mid Z_i, d_i = 0]$

We can write $ATT'(s \mid Z_i)$ as a function of the original

(14) $$\begin{aligned}ATT'(s \mid Z_i) &= \{E[C_{i,t*+s} - C_{i,t*} \mid Z_i, d_i = 1] - E[C_{i,t*+s} - C_{i,t*} \mid Z_i, d_i = 0]\} - \\ &\quad \{E[C_{i,t*+s-1} - C_{i,t*-1} \mid Z_i, d_i = 1] - E[C_{i,t*+s-1} - C_{i,t*-1} \mid Z_i, d_i = 0]\} \\ \therefore \quad ATT'(s \mid Z_i) &= ATT(s \mid Z_i) - ATT(s-1 \mid Z_i)\end{aligned}$$

Now, to evaluate the impact of ethanol plants our primary interest still lies in estimating ATT from the standard model. Since $ATT'(s|Z_i)=\beta_3'$, independent of s, the ATT can be recursively calculated for each post-treatment year as s increases by 1. That is, $ATT(s+1|Z_i)= ATT(s|Z_i)+\beta_3'$ for $s\geq 2$. For $s=1$, first see that $ATT(0|Z_i)=0$ because $E[C^T_{i,t*}-C^U_{i,t*}|d_i=1,Z_i]=0$ [9], which in turn yields that $ATT'(1|Z_i)= ATT(1|Z_i)$. Since $ATT'(1|Z_i)$ is identified by (12) and $ATT(1|Z_i)$ is not, we compute $ATT'(1|Z_i)$ below.

We know that,

$$ATT'(1|Z_i)=\{E[C_{i,t*+1}-C_{i,t*}|Z_i,d_i=1]-E[C_{i,t*+1}-C_{i,t*}|Z_i,d_i=0]\}-$$
$$\{E[C_{i,t*}-C_{i,t*-1}|Z_i,d_i=1]-E[C_{i,t*}-C_{i,t*-1}|Z_i,d_i=0]\}$$

We explicitly write-out the expressions for $C_{i,t*+1}$, $C_{i,t*}$ and $C_{i,t*-1}$ below because $d_t=1$ only for $t*+1$.

$$C_{i,t*+1}=\beta_0+\beta_0'(t*+1)+\beta_1+\beta_1'(t*+1)+\beta_2 d_i+\beta_2'(t*+1).d_i+\beta_3 d_i+\beta_3'(t*+1).d_i+\beta_{4,t*+1}Z_i+\varepsilon_{i,t*+1}$$
$$C_{i,t*}=\beta_0+\beta_0'(t*)+\beta_2 d_i+\beta_2'(t*).d_i+\beta_{4,t*}Z_i+\varepsilon_{i,t*}$$
$$C_{i,t*-1}=\beta_0+\beta_0'(t*-1)+\beta_2 d_i+\beta_2'(t*-1).d_i+\beta_{4,t*-1}Z_i+\varepsilon_{i,t*-1}$$

It can now easily be shown that $ATT(1|Z_i)= ATT'(1|Z_i)=\beta_3+\beta_3'(t*+1)$. The way $ATT(1|Z_i)$ depends on $t*$ also justifies the use of last pre-treatment period as sufficient to compute $ATTs$ for all post-treatment periods. If we were to use the penultimate pre-treatments period instead of the last pre-treatment period, only $(t*+1)$ would be replaced by $(t*+2)$ in the expression for $ATT(1|Z_i)$ as the base period has changed. However, doing this would require at least 3 pre-treatment years which may not be practically available (as is the case of South Dakota for this article).

Hence, the recursive solution to estimate treatment effects, using a DID framework that incorporates differentiated trends, by estimating equation (8) is given as:

(15) $ATT(s|Z_i)=\beta_3+\beta_3'(t*+s)\ \forall s\geq 1$.

Now that we have motivated the idea of incorporating trends into the standard DID framework, we address two further issues addressed by Mora and Reggio (2012). First, that the parallel first-difference assumption that identifies our 'new' average treatment effects for the treated can be generalized into a family of parallel n-differences assumptions. The formulation and interpretation of the average treatment effects in those cases would, however, differ. Second, the authors provide a 'fully-flexible DID model' by incorporating trends through indicator variables for each time period. This model has its two advantages, when compared to our linear-trends model here: (A.) it incorporates flexible trends visualized in figure 10(b.), and (B.) it allows

---

[9] $E[\Delta C^T_{i,t'}-\Delta C^U_{i,t'}|d_i=1,Z_i]=E[C^T_{i,t'}-C^U_{i,t'}|d_i=1,Z_i]=0\ \forall\ t'\leq t*$. This is one of the reasons why it would suffice to consider only the last pre-treatment period to evaluate the treatment effects. Given a recursive formulation to compute ATT for each subsequent post-treatment period, the periods prior to $t*$ would not matter.

testing for equivalence between the parallel n-differences assumptions. The linear-trends DID model that we have developed in this sub-section is essentially a special case of the fully-flexible DID model' presented hereafter. A more intuitive way to incorporate trends into our model that vary each period for both groups is introducing non-linear functional forms for the trend-variable (for example, quadratic trends). Since the fully-flexible version includes dummy variables for each time-period these non-linear trends are only special cases of Mora and Reggio (2012)'s model.

Before presenting the mechanics of a fully-flexible DID model we will motivate the specifics of the family of generalized parallel n-differences assumption using our updated DID model in equation (7). Consider the parallel first-difference assumption in equation (12) that identifies $ATT'(s \mid Z_i)$, s periods ahead of the last pre-treatment period t^*, and re-write it as follows:

(16) $E[\Delta_s \Delta C^U_{i,t^*+s} \mid Z_i, d_i = 1] = E[\Delta_s \Delta C^U_{i,t^*+s} \mid Z_i, d_i = 0]$,

Where, U represents the case of no treatment (or no ethanol plant) and $\Delta_s \triangleq (1 - L^s)$ so that we compute the treatment effect s periods ahead of t^* relative to the first difference in outcome levels at t^*. A generalized parallel n-differences assumption including higher-order differences of outcome levels to identify ATT' for all post-treatment periods similar to that in equation (16). A parallel n-differences assumption, notated as parallel (n-s) assumption by Mora and Reggio (2012) is given as:

(17) $E[\Delta_s \Delta^{n-1} C^U_{i,t^*+s} \mid Z_i, d_i = 1] = E[\Delta_s \Delta^{n-1} C^U_{i,t^*+s} \mid Z_i, d_i = 0]$

See that for $n = 1$ equation (17) reduces to a parallel paths assumption and for $n = 2$ it is the parallel first-difference assumption. For $n > 2$, however, we move towards higher order differences. For example, $n = 3$ implies a $\Delta^2 [= (1 - \mathrm{L}) - (\mathrm{L} - \mathrm{L}^2)]$ operator on the s period ahead outcome variable. We will require at least 3 pre-treatment years in our dataset to exploit such an operator due to the parallel double-differences assumption. Thus, the generalizations introduced by $n > 2$ cases are only applicable to the cases of North Dakota ethanol plants. The generalized average treatment effects from parallel n-differences assumption is given as[10]

(18) $ATT'(s, n \mid Z_i) = \Delta^{n-1} ATT(s \mid Z_i) = E[\Delta_s \Delta^{n-1} C^U_{i,t^*+s} \mid Z_i, d_i = 1] - E[\Delta_s \Delta^{n-1} C^U_{i,t^*+s} \mid Z_i, d_i = 0]$

For the $n = 3$ case of our linear-trends model, $ATT'(s, 3 \mid Z_i) = \Delta^2 ATT(s \mid Z_i) = ATT(s \mid Z_i) - 2ATT(s-1 \mid Z_i) + ATT(s-2 \mid Z_i)$, which will recursively identify $ATT(s \mid Z_i) = ATT'(s, 3 \mid Z_i) + 2ATT(s-1 \mid Z_i) - ATT(s-2 \mid Z_i)$. Similar to the $n = 2$ case, for $s = 1,\ 2$ we will have $ATT(s \mid Z_i) = ATT'(s, 3 \mid Z_i)$. It is quite evident here that the treatment effects estimated under parallel double-differences assumption will not equal those under parallel first-difference or parallel paths assumptions. It is, however, interesting to note that the treatment effects estimated using an exactly same model in equation (11) can be very different in magnitude as well as interpretation depending on the identifying assumption used.

[10] See Theorem 1 in Mora and Reggio (2012).

Note that these updated assumptions for incorporating trends into DID cannot be validated since they are defined as nth-order difference in outcome variable including the post-treatment periods. However, these assumptions can be tested for equivalence using the fully-flexible model discussed next. A parallel n-differences assumption is equivalent to a parallel (n-1)-differences assumption (OR $ATT'(s,n|Z_i) = ATT'(s,n-1|Z_i)\ \forall\ s$) if and only if $E[\Delta^{n-1}C^U_{i,t^*} | Z_i, d_i = 1] - E[\Delta^{n-1}C^U_{i,t^*} | Z_i, d_i = 0]$ [11].

The Fully-Flexible DID Model

A fully flexible model by Mora and Reggio (2012) is as follows:

(19) $C_{i,t} = \beta_0 + \sum_{\tau=T(i)+1}^{T(l)} \beta_\tau I_{[t=\tau]} + \beta^d d_i + \sum_{\tau=T(i)+1}^{T(l)} \beta^d_\tau \times I_{[t=\tau]} \times d_i + \varepsilon_{i,t}$,

Where, $T(i)$ is the first pre-treatment period and $T(l)$ is the last post-treatment period. The model in equation (19) captures flexible time-trends for pre- and post-treatment periods and allows them to differ between treatment and control groups, thus capturing a fully-flexible situation visually depicted by figure 10(b.). $ATT'(s,n|Z_i)$, that is using the generalized parallel n-differences assumption s periods ahead of t^*, using the model in equation (19) is given as

(20) $ATT'(s,n|Z_i) = \Delta^{n-1}ATT(s|Z_i) = \Delta_s\Delta^{n-1}\beta^d_{t^*+s}$ [12]

An early application of the fully-flexible DID model in equation (19) can be found in Reber (2005) to assess impact of court-ordered desegregation plans for schools in 108 U.S. districts on school enrolments. An availability of many pre-treatment years for our analysis (at least for the four North Dakota ethanol plants) makes this model amenable for our analysis. However, an opportunity to implement multiple assumptions and estimating corresponding treatment effects for each case comes with a challenge of choosing among these estimates. We can test the equivalence, say $ATT'(s,n|Z_i) = ATT'(s,n-1|Z_i)\ \forall\ s$, by estimating (19) and testing the null: $\Delta^{n-1}\beta^d_{t^*} = 0$. These tests may help us find more parsimonious specification than (19). An alternative, but more brute force, way would be to evaluate treatment effects for each underlying assumption and seek differences in implications of these estimates. This paper will incorporate these flexible trends in the future.

[11] See Theorem 2 in Mora and Reggio (2012).
[12] See Theorem 3 in Mora and Reggio (2013)