Supply from many: Studies on heterogeneous US land use decisions at the extensive and intensive margins

by

Nicholas Jon Pates

B.A., Minnesota State University - Moorhead, 2011
M.S., Iowa State University, 2013

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Agricultural Economics
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2018

Abstract

Price changes affect the profitability of agricultural land use at the intensive margin (i.e. crop choice) and the extensive margin (i.e. land devoted to crop production). Understanding how prices impact localized land use decisions is important for predicting how production and its allocation across producers change with prices. Due to its wide expanse and diverse geography, the productivity US land differs across space and uses. Understanding the drivers of land use decisions while accounting for such diversity is essential for accurately modeling supply response at the regional and national level. This dissertation contains two studies that provide insight into how price changes impact land use decisions at the extensive and intensive margins.

In the first chapter examine the corn supply-price relationship in the United States. I perform this analysis using field-level data across the contiguous US (CONUS). This study is unique in that it incorporates micro-level data from over 3 million fields to estimate region-specific supply response and then aggregates results to the national level. The dataset used in this study is nearly comprehensive, representing field-level decisions across fields that accounted for over 88% of national corn production between 2009 and 2016.

The findings from this study illustrate the importance of incorporating heterogeneity in supply response models. Supply response to price differed substantially across regions with high supply sensitivity in the north-central US and Mississippi River Delta, moderate sensitivity in Corn Belt states, and

low sensitivity in the western and Gulf Coast states. The relative importance of corn production in the in the Corn Belt states of Iowa, Illinois, Indiana, and Nebraska meant that it was far less sensitive and, in the long-run, more stable to price changes than national corn supply as a whole. Including heterogeneity in supply response also provided policy relevant context to supply response studies. Overall supply response was negatively correlated with area yields. This meant that price changes have a larger effect on planted corn acres and a smaller effect quantity of corn itself.

In the last chapter I examine the impact that ethanol plant capacity has on local land use at the extensive margin. The Renewable Fuel Standard (RFS) has been one of the most influential agricultural policies in the past 20 years, increasing general US crop prices by over 20% and inducing a substantial in US ethanol production capacity (Carter et al., 2016; Roberts and Schlenker, 2013). Its effect on cropland extensification was a concern before it was passed since the policy includes a stipulation forbidding ethanol production on cropland converted after 2007. Lands at the extensive margin tend to be less productive and more environmentally sensitive. Extensive transitions also tend to be less frequent than transitory breaks in crop rotations making their impacts longer-lasting.

The goal of this final analysis is to isolate the impact of ethanol expansion on cropland transitions from the general price changes. The concurrent increase in general crop prices and ethanol construction from the RFS complicates the estimation of plants' effects. I isolate these effects using difference-in-differences (DID) which removes impact from common price trends between the treatment and control group. The standard DID approach results show significant pre-treatment effects stemming from non-random ethanol plant construction. Treatment is likely non-random since ethanol plants locate in areas that provide better returns. Factors that impact the returns to plants confound the analysis since they likely also impact cropland transition decisions. To address this, I use propensity score matching to ensure these confounding factors are identically distributed between the treatment and control groups. Under the matched DID models, the expansion of ethanol plants tended to increase cropland retainment and reduce lands transitioning from non-cropland to cropland. While these results seem contradictory, they are consistent with the findings in recent literature. These impacts are thought to arise due to higher program retention in the major US cropland retirement program CRP due to changes that disproportionately impacted major ethanol production areas.

Supply from many: Studies on heterogeneous US land use decisions at the extensive and intensive margins

by

Nicholas Jon Pates

B.A., Minnesota State University - Moorhead, 2011
M.S., Iowa State University, 2013

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Agricultural Economics
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas
2018

Approved by:

Major Professor
Nathan P. Hendricks

Abstract

Price changes affect the profitability of agricultural land use at the intensive margin (i.e. crop choice) and the extensive margin (i.e. land devoted to crop production). Understanding how prices impact localized land use decisions is important for predicting how production and its allocation across producers change with prices. Due to its wide expanse and diverse geography, the productivity US land differs across space and uses. Understanding the drivers of land use decisions while accounting for such diversity is essential for accurately modeling supply response at the regional and national level. This dissertation contains two studies that provide insight into how price changes impact land use decisions at the extensive and intensive margins.

In the first chapter examine the corn supply-price relationship in the United States. I perform this analysis using field-level data across the contiguous US (CONUS). This study is unique in that it incorporates micro-level data from over 3 million fields to estimate region-specific supply response and then aggregates results to the national level. The dataset used in this study is nearly comprehensive, representing field-level decisions across fields that accounted for over 88% of national corn production between 2009 and 2016.

The findings from this study illustrate the importance of incorporating heterogeneity in supply response models. Supply response to price differed substantially across regions with high supply sensitivity in the north-central US and Mississippi River Delta, moderate sensitivity in Corn Belt states, and

low sensitivity in the western and Gulf Coast states. The relative importance of corn production in the in the Corn Belt states of Iowa, Illinois, Indiana, and Nebraska meant that it was far less sensitive and, in the long-run, more stable to price changes than national corn supply as a whole. Including heterogeneity in supply response also provided policy relevant context to supply response studies. Overall supply response was negatively correlated with area yields. This meant that price changes have a larger effect on planted corn acres and a smaller effect quantity of corn itself.

In the last chapter I examine the impact that ethanol plant capacity has on local land use at the extensive margin. The Renewable Fuel Standard (RFS) has been one of the most influential agricultural policies in the past 20 years, increasing general US crop prices by over 20% and inducing a substantial in US ethanol production capacity (Carter et al., 2016; Roberts and Schlenker, 2013). Its effect on cropland extensification was a concern before it was passed since the policy includes a stipulation forbidding ethanol production on cropland converted after 2007. Lands at the extensive margin tend to be less productive and more environmentally sensitive. Extensive transitions also tend to be less frequent than transitory breaks in crop rotations making their impacts longer-lasting.

The goal of this final analysis is to isolate the impact of ethanol expansion on cropland transitions from the general price changes. The concurrent increase in general crop prices and ethanol construction from the RFS complicates the estimation of plants' effects. I isolate these effects using difference-in-differences (DID) which removes impact from common price trends between the treatment and control group. The standard DID approach results show significant pre-treatment effects stemming from non-random ethanol plant construction. Treatment is likely non-random since ethanol plants locate in areas that provide better returns. Factors that impact the returns to plants confound the analysis since they likely also impact cropland transition decisions. To address this, I use propensity score matching to ensure these confounding factors are identically distributed between the treatment and control groups. Under the matched DID models, the expansion of ethanol plants tended to increase cropland retainment and reduce lands transitioning from non-cropland to cropland. While these results seem contradictory, they are consistent with the findings in recent literature. These impacts are thought to arise due to higher program retention in the major US cropland retirement program CRP due to changes that disproportionately impacted major ethanol production areas.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to extend my deepest thanks to all those who have helped me through the Ph.D. program in thought and action. First and foremost I thank my parents, Mikkel and Barbara Pates, my sister and brother-in-law Jessica and Joe Beckett for their love and emotional support throughout my time in Manhattan. They instilled and appreciation for an education and a sincere responsibility to do good in the world. They are truly the finest people I know and I am forever thankful for their love and support. I would also like to thank my major professor Dr. Nathan Hendricks for his patience and dedication towards me as a student. Nathan has been a great role model to me and is dedicated to the education and research in his state and discipline and to the success of his students. I have throughly enjoyed working with Nathan over these past several years and I hope to continue to collaborate as I transition out of the Ph.D. program. I would also like to extend my sincere thanks to the members on my committee for their ideas and contributions as mentors, instructors, and friends. In particular I would like to acknowledge the contributions of Drs. Jason Bergtold and Jesse Tack. My work as a researcher has presented challenges and personal rewards and I have been grateful to share both with Jason, Jesse, and Nathan.

More broadly I would like to extend my thanks to the Department of Agricultural Economics here at Kansas State, the University as a whole, and the community of Manhattan, Kansas. Before I entered the program I was unsure about going back to school after receiving a Masters degree at Iowa State. From the moment I was accepted and took a tour of the campus the department made me feel wanted. I appreciate the opportunities the department has given me to show my appreciation for them and build my skills as an academic. The Agricultural Economics faculty has been world-class and take their roles as mentors to graduates students seriously. I would like to especially thank our Department Head Dr. Allen Featherstone who, in addition to being a fine instructor, gave me the opportunity to instruct my own courses and has been an advocate for our department on campus and the profession more broadly. I would also like to thank Dr. Keith Harris for reaching out to collaborate with me on other projects throughout the course of my time at Kansas State. I would also like to thank the departmental staff who do great work in their day-to-day jobs by ensuring that we as students and instructors have what we need to succeed.

Outside of the department I would like to thank researchers at other institutions for their help on this and other projects I've been involved with. I would especially like to thank the members of the department of geography at the University of Wisconsin including Tyler Lark, Matthew Bougie, and Holly Gibbs, and Joshua Woodard at Cornell University for their collaboration and generosity. I would like to also thank my outside committee member Dr. Marcellus Caldas for his words of support, comments and adding the perspective as a geographer to this dissertation. I would like to extend a special thanks to members of the Kansas State Computer Science Department Drs. David Turner and Dan Andresen and the rest of the team that oversees the Beocat cluster as well Dr. William Hsu for his service as my outside committee member.

Lastly I would also like to thank the generous support of the National Wildlife Fund for their funding which supported this research.

Chapter 1

Introduction

E pluribus unum, "Out of Many, One", the motto of the United States and the inspiration for the title of this dissertation. While an important reminder to the origins of the United States, this motto also parallels agricultural production in the US. Many individual producers contribute to the nation's total crop production and while facing differing constraints under different environments. In this dissertation I study the nature of this heterogeneity, how producers respond differently to changing prices and local demand and how these differences correspond to aggregate changes. The following two chapters in this dissertation cover how changes in local prices and demand influence farmers' extensive and intensive land use decisions.

In the context of this dissertation, the choice of crops is considered an intensive margin decision, and the amount of land devoted to crop production is an extensive margin decision. In the next chapter I study how changes in crop prices influence intensive land use decisions. Specifically I estimate the impact of price changes on the probability of planting corn across the United States. This study utilizes a large field-level dataset consisting of over 32 million observations, which accounted for over 80% of the total US corn acreage between 2009 and 2016. To incorporate heterogeneous price response the across the country, I divide the national sample into Major Land Resource Areas (MLRAs) and estimate separate across these areas. I find a high degree of supply response heterogeneity across the country. Supply in some areas was over three times more sensitive to price changes relative to the Corn Belt. In other areas, the supply of corn barely responds to prices at all. When results are aggregated across the country, I find that models that do not account for heterogeneity in supply response tend to underestimate supply elasticities. Similarly, studies that proxy nationwide supply elasticities with their Corn-Belt counterparts also underestimate nationwide supply elasticities.

In the final chapter of the dissertation I study how changes in local ethanol plant capacity impact extensive land-use decisions, the conversion to or retainment of cropland. Cropland conversion is important as lands at the extensive margin tend to be more environmentally sensitive and less productive as cropland. The remote sensed data used to identify field-level land uses across the US often sensed using satellite imagery and often identified with error. These errors compound when identifying sequences of choices such as more permanent extensive land-use transitions (Donaldson and Storeygard, 2016). In this study I use a unique land conversion dataset to measure whether land was converted to cropland and the year of conversions. This dataset specially constructed to address potential errors with validation against surveyed National Resources Inventory (NRI) observations, filtering by consistent observations, and removing lands that are more prone to measurement error (Lark et al., 2017).

I use these data to estimate the effect of ethanol plant construction on local cropland conversion or retention. The "treatment" of interest is whether a field was in an area of expanding ethanol capacity. I define these treatment groups using a geocoded dataset of over 200 ethanol plants across the country. Treatment and control groups were then defined using the proximity of the plant after accounting for each plants' capacity, county yields, and assumed conversion rates on land surrounding the plants. The goal is to isolate the effect of the plant from general price changes. I do this using a variety of difference-in-differences (DID) procedures which difference out the effects of changes common to the treatment and control groups. Like the previous chapter, I control for potential heterogeneity in the responsiveness to price. In this study I do this using propensity score matching with the difference-in-differences approach. Like other producers, plants are profit-seeking and therefore choose non-random locations that will produce higher expected profits. For instance, plants may choose to locate in areas with higher crop productivity to retain a consistent supply of raw inputs. With matching, comparable treatment and control groups can be constructed so that the DID estimates are more likely attributable to the treatment effect from the ethanol plant. This helps correct for potential bias of the DID approach. An event study reveals statistically significant treatment effects four years prior to plant construction. This suggests that the assumptions needed for the DID approach to yield a causal estimates do not hold and justify the use of the matching procedure. The matched DID estimates were generally more statistically significant giving evidence that ethanol plants increase cropland retainment but decrease cropland conversion. These results are contradictory since ethanol plants generally strengthen basis which should, ceteris paribus, make crop production more profitable for area farmers. Recent literature with similar findings suggest that concurrent changes in land retirement programs could have discouraged cropland conversion and disproportionately impacted areas with higher ethanol production. This and balancing problems from the matching procedure suggest that adding more matching variables, particularly ones relevant to the Conservation Reserve Program (CRP), could improve the analysis.

In the following chapters in this dissertation I utilize large amounts of field-level data across the major growing areas in the United States. Remote sensed datasets have been used in a variety of applications across economics and useful for providing data that is difficult to observe otherwise and provide micro-level surveillance over wide expanses (Donaldson and Storeygard, 2016). The agricultural economics discipline has benefited greatly from the use of field level data, allowing for accurate the estimation of the incidences of subsidies to assessing the environmental impact of crop production (Hendricks et al., 2014b; Kirwan and Roberts, 2016). In this paper, this higher spatial resolution allows for more accurate measurement of regional crop choices at the field-level and broader land use within unique neighborhoods of ethanol plants. Without these disaggregated datasets, researchers often rely on datasets that average over political areas such as states or counties.

Averaging observations across arbitrary lines can mask correlations in features of interest and lead to aggregation bias (Blundell and Stoker, 2005). Observing behavior closer to the level where individuals make decisions can help identify these correlations that would otherwise be lost due to averaging. For instance, I found that fields experiencing extremely wet planting conditions were less likely to plant corn in a given year. Widespread wet planting conditions generally increase expected harvest prices due to reductions in expected aggregate supply. With remotely sensed data, weather can be more accurately measured at the field-level, accounting for fields with more moderate planting conditions and controlling for omitted variable bias.

Chapter 2

Estimating Heterogeneous Corn Supply Response to Price

2.1 Introduction

Accurate supply elasticities are critical when estimating the distribution of economic surplus, the reallocation of production across producer groups, and the impact of production externalities brought about by price changes. While they represent general supply movements, elasticities ultimately arise from the responses of individual producers as prices change. In this paper, I construct corn supply response estimates for the Contiguous United States (CONUS) by constructing a set of 122 heterogeneous regional supply response estimates using observations across 3.6 million fields. Because corn production takes place over an environmentally, climatologically, and economically diverse area, accounting for heterogeneity is important to accurately estimate supply response in the US as a whole. I find that there is high degree of heterogeneity in corn supply response across the US and that models that do not allow for coefficient heterogeneity tend to underestimate nationwide elasticities and misrepresent supply dynamics.

Tractably producing accurate national supply elasticities is challenging. Using national or state average prices and quantities to estimate elasticities can average out potentially important correlations within the country or state. Using smaller regional samples may not represent the country as a whole. With remote sensed data from the Cropland Data Layer (CDL), I model supply response over a field-level dataset that is reasonably representative of the US as a whole. This dataset accounted for over 80% of the total planted corn acreage and over 90% of corn production between 2009 and 2016. I estimate supply response by modeling crop choices conditional on past choices. Modeling crop choice separately across more than 100 areas of the country and conditioning on past crop choices allows for variable supply response across areas, parsimoniously incorporates desirable rotational concepts of crop choice modeling, and reduces computational complexity across the large dataset.

Under this framework, I assume that supply responds to prices through only crop choice. To construct acreage response estimates, I multiply the estimated planting probabilities by field acreage. To compute quantity response I multiply by county yield estimates. Marginal effects and elasticities are then computed using estimated crop choice marginal effects with respect to crop prices under the assumption that yields are not correlated with prices variations. However, this does not mean quantity response should equal the acreage response. Corn planting decisions were found to be more sensitive to price changes in areas with lower yields which drove acreage response below quantity response.

Modeling over separate subnational samples of the data helps identify areas of the country where crop choices are more sensitive to price fluctuations. In general, I find that there is a high degree of supply response heterogeneity across the country. Supply in some areas was elastic while supply in other areas was uninfluenced by price changes. The diversity of supply response across the United States is pertinent to the study of regional effects of policies that impact domestic prices. In particular, I found that producers in the environmentally sensitive Prairie Pothole Region were more far responsive to price changes than the average farmer in Corn Belt states. Environmentalists are especially concerned with the expansion of cropland in this area as it serves as an important breeding and habitat area for many species of migratory waterfowl (Wright and Wimberly, 2013). While assessing the environmental impact of national corn supply response is beyond the scope of this paper, these findings suggest that there are potential benefits to using models that allow for more localized supply response heterogeneity. The high variability in the US corn supply response across regions also means that the supply a given region is less representative of the country as a whole. Regional elasticities particularly Corn Belt estimates are often produced in the literature. The Corn Belt elasticities in this paper were approximately half of their national counterparts.

To understand the benefits of using more micro-level data, it is useful to discuss why modeling with average aggregated data can produce different results. Modeling supply response with a single set of coefficients is appropriate when either producers respond to prices identically or when price variation is independent across individuals. While restrictive, homogeneity in response across all agents within a group will, by definition, imply that the mean response will equal the individual response. This was often an assumption in early supply response studies e.g. (Chavas and Holt, 1990; Lee and Helmberger, 1985) The second less restrictive assumption is that price variation occurs randomly across the sample. Modeling over average aggregate variables when price fluctuations are non-random can bias estimate in a fashion similar to omitted variable produces bias. This bias arises when non-random price fluctuations are correlated with omitted field-level features that also impact crop choices (Blundell and Stoker, 2005). There are a variety of field-level features which can impact cropping decisions. At the national-level, soil quality on average is static but is highly heterogeneous within the country. For this reason, a national aggregate model effectively omits the influence of soil quality. While this can be partially addressed by modeling with state or county-level data, these areas are not delineated by agriculturally relevant features and can similarly omit variation in relevant field-level variables. Areas with higher quality soil may be more apt to plant corn as prices change than areas with more marginal soils. If price fluctuations were non-random and tended to be more extreme in areas with higher soil quality, then models estimated using aggregated variables would overestimate the supply response relative to estimates built up from separately modeled-individual responses.

For similar reasons, models that allow for heterogeneous response across regions should be incorporated to reduce bias. Using models that restrict the responsiveness of producers to the country average is biased when prices do not change uniformly across the country. If prices are more variable in areas where producers are more sensitive to price variation, models that do not allow for heterogeneous coefficients will tend to underestimate the true price response of the country since the model will underestimates price responsiveness in areas with higher price variability. Using disaggregated field-data to estimate models with heterogeneous response is ideal since it allows observations that are likely to respond similarly to price to be grouped to improve the overall estimation of supply response.

In this paper I compare national elasticity measures between two models. While both were estimated using field-level observations, the first model pools observations over the entire national sample to produce a single set of coefficients for the nation. The second model allows for coefficient heterogeneity by separately estimating supply response over different subsamples of the national sample. The result of this analysis shows significant differences between the pooled and heterogeneous models which suggests that behavioral homogeneity and random price assignments do not likely hold. This indicates that there could be pooling bias and models that allow for supply response heterogeneity are preferred.

Provided that yields are do not respond to prices, homogeneous supply response implies that the acreage response should, on average, equal the quantity response. Homogeneity in planting response implies that planting decisions are made similarly across the entire sample therefore should be uncorrelated with yields. Homogeneity in supply response also implies that the impact of production externalities (e.g. environmental) brought about by price changes is also homogeneous since cropping decisions are not correlated with impact potential.[1] I find that this generally does not hold as supply response was correlated with several variables of interest. For example, crop choice was more sensitive to prices in areas with lower yields which lead to a quantity response that was below the acreage response. This is consistent with the results of the conceptual model of Lubowski et al. (2006) in which producers allocate land across uses to maximize profits. Generally speaking, crop choices will be more responsive to price changes when crop producers are more indifferent between choosing one crop or another as when two crops provide similar profits for a given set of prices. This was more likely when corn yields are relatively low.

Incorporating heterogeneous supply response across an area as vast and variable as the United States is computationally challenging. To retain tractability, oftentimes compromises in observations or scope need to be made to estimate supply response models under more complex contemporary modeling frameworks. Smaller samples may not be as representative to the national-level and more aggregated models may suffer from aggregation bias. The current supply response literature places more emphasis on incorporating heterogeneity (Haile et al., 2016; Lacroix and Thomas, 2011; Motamed et al., 2016). Many do so using a fixed effect or additive separable

[1]While the task of attributing environmental harm to changing agricultural practices due to price changes is beyond the scope of this paper, environmental impact is an area where knowledge of heterogeneous response can be applied.

heterogeneous effect frameworks. Models allowing for heterogeneous supply response coefficients are also becoming popular as more evidence of non-additive heterogeneity in supply response comes to light (Koutchadé et al., 2018). Seen as a major component of supply response, more of the literature now incorporates rotational frameworks. To include the influence of rotations while allowing for heterogeneity many researchers utilize variants of multinomial discrete choice models. In particular, random parameter and latent class multinomial discrete choice models are popular as they can consider sequential choices with many alternatives (Claassen et al., 2017; Langpap and Wu, 2011). While these models are highly flexible, they are also computationally burdensome and many of these studies utilize regional data and are limited to datasets on the order of less than 1,000 observations to a little over 100,000 observations.

Estimating regional supply response, Hendricks et al. (2014b,c) provide a promising alternative to multinomial discrete choice modeling frameworks when working with larger datasets. They utilize a Markov chain framework and model conditional crop transitions using two or more individual discrete choice models to characterize crop transition probabilities. Heterogeneity is introduced in these models using control variables within the models and by estimating these models over different subsets of the dataset. Modeling across separate regions of the country introduces heterogeneity in the model. Modeling crop transitions conditional on previous choices allows for the analysis of rotations. This is desirable since crop choices, especially in the short-run, are often motivated by the benefits of rotating crops. This framework is also convenient as the total sample is divided first by region, and then again when conditioning the sample over previous crop choices. While this methodology requires more regressions, each regression is computationally simple. This helps retain tractability even over large datasets (Hendricks et al., 2014b,c). As such, this study uses the Markov chain transition framework and focuses on the impact that prices have on the probability of planting corn and is able to model over a dataset with over 30 million observations.

2.2 Conceptual Framework

In this section I provide an overview of the Markov transition regression approach. Markov chains are used for a variety of purposes and describe sequences of state probabilities. In this model "state" probabilities refer to probabilities that a particular crop is grown in a particular period. Conditional transition probabilities are probabilities that a crop choice is made conditional on crop choices that precede it. Conditional transition probabilities are a primary component of Markov chains. Since, in rotations, the benefits of planting a given crop are conditional on past crop choices, Markov transition regressions naturally incorporate rotational concepts that are desirable in crop supply response models.

Suppose farmers choose among K alternative crops. These conditional crop choice probabilities help describe the Markov chain shown in equation 2.1. The $K \times K$ matrix ($\mathbf{T}$) is the conditional transition matrix. This matrix is composed of conditional transition probabilities $\mathbb{P}_t^{i|j}$ equal to the probability that crop i is chosen at time t given crop j was chosen in period

$(t-1)$. The $1 \times K$ vector $(\mathbf{p}_t)$ is a vector of state probabilities at time t. Unlike the transition matrix, the state probability vector is composed of *unconditional* probabilities that a farmer will plant a particular crop $\left(\mathbb{P}_t^k\right)$.

$$\mathbf{T} \times \mathbf{p}_{t-1} = \mathbf{p}_t \tag{2.1}$$

$$\mathbf{T} = \begin{bmatrix} \mathbb{P}_t^{1|1} & \mathbb{P}_t^{1|2} & \cdots & \mathbb{P}_t^{1|K} \\ \mathbb{P}_t^{2|1} & \mathbb{P}_t^{2|2} & \cdots & \mathbb{P}_t^{2|K} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{P}_t^{K|1} & \mathbb{P}_t^{K|2} & \cdots & \mathbb{P}_t^{K|K} \end{bmatrix}; \quad \mathbf{p}_t = \begin{bmatrix} \mathbb{P}_t^1 \\ \mathbb{P}_t^2 \\ \vdots \\ \mathbb{P}_t^K \end{bmatrix} \tag{2.2}$$

The goal of the analysis is to estimate the probability that a particular crop is planted in each period and on each field. Formally the goal is to characterize $\mathbf{p}_t$ and its relationship with crop prices. A complication of using the Markov chain in equation 2.1 is that the state probability is dependent upon all previous state probabilities. Therefore, to estimate the state probability of planting some crop in a given period would require all of the previous state probability estimates. To avoid the complication of nesting probability estimates, this study focuses on the steady state probability, the state probability vector after it has stabilized over time. This is represented in equation 2.3 as the steady-state version of the Markov Chain where $\mathbf{p}$ is stable over all time periods ceteris paribus.

$$\mathbf{T} \times \mathbf{p} = \mathbf{p} \tag{2.3}$$

In this case, $\mathbf{p}$ can be characterized entirely by the elements in the conditional transition matrix, in other words, rotational incentives.

With the importance of rotations in mind, Hennessy (2006) provides important theoretical insights on how one might characterize these rotation choice incentives. In particular, he shows that if the practical impact of rotation crops comes through changes in yield or input requirements, then the profitability of a given rotation simplifies to a linear function of the yield and the input benefits.[2] Hennessy simplifies the approach with the concept of a rotation "memory". The memory of a rotation refers to the maximum length of time that a lagged choice can influence contemporaneous yields or input requirements. For instance, if rotations have a memory of two, then a crop choice made two periods in the past could influence contemporaneous yield or effective input. Under crop rotations with a one-period memory, only the previous crop choice could influence contemporaneous yields. This study assumes that rotations have one-period memory since it is implicit to the simple Markov chain in equation 2.3. Hennessy found evidence for a one-period memory rotations in the state of Iowa (Hennessy, 2006).

Crop rotations with a single-period memory greatly simplify the analysis as it means that only the one-period-lagged choices need to be accounted for in the modeling process. For instance, if crop rotations only involved corn (crn) and soybeans (soy) and had a one-period memory, the modeling framework would need to account for the crop sequences $\{crn, soy\}$, $\{soy, crn\}$, $\{crn, crn\}$, and $\{soy, soy\}$ but not, for instance, $\{crn, crn, soy\}$. If $\{crn, crn, soy\}$ were selected, then the first two letters of the rotation imply that the optimal crop to follow a corn decision would be corn. However, the last two letters of the rotation imply that the optimal crop to follow corn

[2]While the functional form of the model is simplified, the results of Hennessy (2006) do not imply that these benefits will be the same across different observations and hence heterogeneous marginal effects are still relevant.

would is soybeans. This is because only the previous letter matters to the contemporary decision for rotations with one-period memory. While assuming that rotations have a single period of memory restrict the set of rotations that need to be modeled, these rotations may involve more than two crop choices. For instance if I added another land use choice of fallowing (fal), then $\{crn, fal, soy\}$ would be a relevant crop rotation. This would imply that it is optimal to fallow for a season following a corn-production year, plant soybeans following a fallow year, and plant corn following a soybean year. Even with a single-period memory assumption the modeling framework can still become complex when more crops are added.

Corn supply response is the primary focus of this analysis. Because there are many relevant alternatives to corn and this relevancy differs across the country, I simplify the rotations by considering two crops, corn (crn) and some "other" competing crop (oth). Under the assumption of one-period memory rotations, I only need to consider the set of rotations $R = \{\{crn, crn\}, \{oth, oth\}, \{oth, crn\}\}$ where $\{crn, crn\}$ is continuous corn, $\{oth, oth\}$ is continuous other crop selection, and $\{oth, crn\}$ is a other crop - corn rotation[3]. Since the steady-state probabilities are all functions of the transition probabilities, the initial modeling goal is to estimate the conditional transition probabilities and their relationships with crop prices.

[3] Here a $\{oth, crn\}$ rotation is equivalent to a $\{crn, oth\}$ rotation since it specifies an other crop follows a corn crop selection and vice versa.

2.3 Methodology

To estimate heterogeneous supply response across the United States, a set of Markov transition regression models were run over different subsets of the national sample. Across regressions, data were divided into groups delineated by both Major Land Resource Areas and soil texture classifications where separate regressions were estimated for each group. The procedure is equivalent to estimating a regression over the entire national sample and including first-order group interactions with a series of group-specific dummy variables. Following Hendricks et al. (2014b), I utilize two Markov transition equations to define the elements of the conditional transition matrix. One equation models the probability of planting corn given the previous planting decision was corn and the second models the probability of planting corn given some other crop was previously planted. These two models can be characterized as first-order Markov transition probability equations describing the probability that a farmer plants corn conditional on the previously planted crop. These choices are characterized by the indicator variable y_{it}^{k} where:

$$y_{it}^{k} = \begin{cases} 1 \mid & \text{Crop } k \text{ planted on field } i \text{ at time } t \\ 0 \mid & \text{Otherwise.} \end{cases} \tag{2.4}$$

Equations 2.5 and 2.6 show the structure of the Markov transition probabilities of planting corn given corn or some other crop was planted in the previous year respectfully.

$$y_{it}^{CC} = \mathbb{E}\left[y_{it}^{C} \mid y_{it-1}^{C}\right] = \Lambda\left(\beta_{10} + \beta_1^C P_{it}^C + \beta_1^O P_{it}^O + \gamma_1 \mathbf{X}_{it} \mid y_{it-1}^C = 1\right) \tag{2.5}$$

$$y_{it}^{OC} = \mathbb{E}\left[y_{it}^{C} \mid y_{it-1}^{O}\right] = \Lambda\left(\beta_{20} + \beta_2^C P_{it}^C + \beta_2^O P_{it}^O + \gamma_2 \mathbf{X}_{it} \mid y_{it-1}^O = 1\right) \tag{2.6}$$

Here y_{it}^C is an indicator variable if corn were grown on field i at time t, P_{it}^C is the expected harvest-time corn price, P_{it}^O is the expected harvest-time other price index, and $\mathbf{X}_{it}$ contains field-level controls. These controls include extreme pre-plant precipitation, field i's slope, soil productivity, irrigation status, and a time trend. $\Lambda(\cdot)$ is the logistic distribution. Only fields raising either corn or other crops between the two consecutive periods are in the sample used to estimate these models. Functionally, these two equations are identical with the only difference between them being the samples used to estimate the relationship.

Equation 2.7 shows the structure of the Markov chain that is being estimated. Since this analysis admits only two crop choices, the conditional transition probabilities can be simplified using the compliment of the non-corn transitions. Here $(\mathbb{P}_{it})$ is the steady state probability of planting corn. In this way, the left hand side variables for the two models are conditional probabilities of $\{crn, crn\}$ and $\{oth, crn\}$ planting sequences.[4] The estimated coefficients β^C and β^O are of interest as they are used to estimate the marginal effect that crop prices have on the conditional probability of planting corn in the contemporary period.

[4]Note that sequences are not the same as rotations. While rotations are repeated, sequences need not be. For instance a crop sequence $\{oth, crn\}$ occurs when a farmer plants the other crop in the previous period and corn in the following period. An $\{oth, crn\}$ rotation is when a farmer plants some other crop in the previous period and corn in the following period *and then* immediately repeats this sequence. In other words, if a farmer adopts the $\{oth, crn\}$ rotation, half of the time, he will plant an $\{oth, crn\}$ sequence and the other half he will plant a $\{crn, oth\}$ sequence. Conversely, continuous rotations (e.g. the $\{crn, crn\}$ and $\{oth, oth\}$ rotations) are entirely comprised of their respective continuous sequences and are therefore identical to their respective sequences.

$$\begin{bmatrix} y_{it}^{CC} & y_{it}^{OC} \\ \left(1 - y_{it}^{CC}\right) & \left(1 - y_{it}^{OC}\right) \end{bmatrix} \begin{bmatrix} \mathbb{P}_{it} \\ (1 - \mathbb{P}_{it}) \end{bmatrix} = \begin{bmatrix} \mathbb{P}_{it} \\ (1 - \mathbb{P}_{it}) \end{bmatrix} \tag{2.7}$$

The steady-state probability of planting corn in period t takes the form of equation 2.8. Assuming steady-state probabilities exist, they can be estimated using Cramer's rule over the steady-state Markov process.[5] The steady state probability is completely characterized by the probabilities of switching from crop to another. The steady state probability will be above 0.5 if transitions from non-corn to corn are more likely than switches away from corn.

$$\mathbb{P}_{it} = \frac{y_{it}^{OC}}{1 - y_{it}^{CC} + y_{it}^{OC}} = \frac{y_{it}^{OC}}{y_{it}^{CO} + y_{it}^{OC}} \tag{2.8}$$

With the structure of the Markov chain in mind, I define short and long-run effects of price changes on the probability of planting corn. From the Markov chain, the two state probabilities are shown in equations 2.9 and 2.10.

$$\mathbb{P} = \underbrace{\mathbb{P} y_{it}^{CC}}_{\mathbb{P}^{CC}} + \underbrace{(1 - \mathbb{P})\, y_{it}^{OC}}_{\mathbb{P}^{OC}} \tag{2.9}$$

$$(1 - \mathbb{P}) = \underbrace{\mathbb{P} y_{it}^{CO}}_{\mathbb{P}^{CO}} + \underbrace{(1 - \mathbb{P})\, y_{it}^{OO}}_{\mathbb{P}^{OO}} \tag{2.10}$$

The state probabilities are made up of two sequential probabilities. Since the data in this analysis is restricted to those planting corn or some other crop, any field planted in corn would must have transferred from corn to corn,

[5]The steady-state probability of planting the other crop is the complement of the steady-state probability of planting corn.

or from the other crop to corn. The probability of a field planting a corn-to-corn sequence $\left(P^{CC}\right)$ plus the probability of field planting an other-to-corn sequence $\left(P^{OC}\right)$ therefore equals the steady-state probability of planting corn. The short-run marginal effect assumes price changes impact the state probability of planting corn only through the transition matrix. Since the impact from the transition matrix alone does not contain any influence from the change in the steady state variable, it is considered a transitory influence from price and is therefore called the "short-run" marginal effect. Any marginal effect that incorporates influence from changes in the steady-state relationship is therefore referred to as as a "long-run" marginal effect. Equation 2.11 shows the short-run marginal effect when the price of crop k changes.

$$\left.\frac{\partial \mathbb{P}}{\partial P^k}\right|_{SR} = \underbrace{\mathbb{P}\frac{\partial y^{CC}}{\partial P^k}}_{\left.\frac{\partial \mathbb{P}^{CC}}{\partial P^k}\right|_{SR}} + \underbrace{(1-\mathbb{P})\,\frac{\partial y^{OC}}{\partial P^k}}_{\left.\frac{\partial \mathbb{P}^{OC}}{\partial P^k}\right|_{SR}} \tag{2.11}$$

The long-run marginal effect allows influence from changes in the steady state probability as well as the transitory effects of the transition matrix. Equation 2.12 shows the long-run marginal effect of a price change and can be derived by inserting the derivative of equation 2.8 as a second term. Like the short-run marginal effects, the long-run state marginal effects will be a sum of the long-run marginal effects of prices on the sequence probabilities $\mathbb{P}^{OC}$ and $\mathbb{P}^{CC}$.

$$\left.\frac{\partial \mathbb{P}}{\partial P^k}\right|_{LR} = \underbrace{\left.\frac{\partial \mathbb{P}}{\partial P^k}\right|_{SR} + \left[y_{it}^{CC} - y_{it}^{OC}\right] \frac{\left[1-y_{it}^{CC}\right]\frac{\partial y^{OC}}{\partial P^k} + y_{it}^{OC}\frac{\partial y^{CC}}{\partial P^k}}{\left[1-y_{it}^{CC}+y_{it}^{OC}\right]^2}}_{\left.\frac{\partial \mathbb{P}^{CC}}{\partial P^k}\right|_{LR} + \left.\frac{\partial \mathbb{P}^{OC}}{\partial P^k}\right|_{LR}} \tag{2.12}$$

The long-run marginal effect shares much of its functional form with the short-run effect and differs only by the second component in equation 2.12.[6]

Understanding this component is important for understanding how short and long-run elasticities and marginal effects differ. If farmers that planted corn in the previous period had the same probability of planting corn as those planting the other crop in the previous period, then the long-run effect would equal the short-run effect. Since there are typically returns to rotating crops, it is more likely that a field not in corn plant would plant corn in the next period than a field continuously plants corn $\left(y^{OC} \geq y_{CC}\right)$. If this is the case, then the additive term will take on the opposite sign as the last ratio component. Since the "y" terms are all conditional probabilities, they take values between zero and one. Also note that the denominator is squared and is therefore always positive. Therefore, the sign of the ratio term will take the signs of the derivatives in the numerator. It is expected that these derivatives have consistent signs for a given crop k. If the derivative is with respect to corn prices, these derivatives are expected to be positive since higher corn prices make producing corn more profitable regardless of the previous crop choice. If the crop price k is for some other crop, these derivatives are expected to be negative since higher "other" prices raises the profitability of the other crop relative to corn ceteris paribus. Since the returns to rotations are likely to imply $(y_{OC} \geq y_{CC})$, the long-run marginal effect will be smaller in magnitude than the short-run effect.

[6]This model can also be characterized as a standard dynamic model with a lagged dependent variables. Under this model short-run marginal effects involve the effect contemporary prices while long-run effects also incorporates the influence from the lagged dependent variable.

The effect that prices have on rotations can be found using the sequential marginal effects as well. In the case continuous corn or continuous other-crop rotations, the sequential marginal effects will equal the rotational effects. This is because individuals in the continuous corn (other) rotation perform a corn-corn (other-other) planting sequence in every year. This is not true for those in the other-corn rotation. Half of the years, individuals in these rotations perform a other-corn planting sequence and the other half they perform a corn-other planting sequence. As such, the probability of a other-corn rotation is the sum of these sequential probabilities. Equations 2.13, 2.14, and 2.15 show the relations between each rotational probability and the sequential probabilities. Since every rotational probability is a linear function of the sequential probabilities the rotational marginal effects will be linear functions of the sequential marginal effects.

$$\mathbb{P}^{\{CC\}ROT} = y^{CC} \tag{2.13}$$

$$\mathbb{P}^{\{OO\}ROT} = 1 - y^{OC} \tag{2.14}$$

$$\mathbb{P}^{\{OC\}ROT} = y^{OC} + 1 - y^{CC} \tag{2.15}$$

Under this framework, supply responds to price only through crop choice. To compute the expected acreage for a given field, I simply multiply the state probability by the field acreage to obtain the field's expected planted corn acres. To obtain the expected quantity I multiply this last term by the field's county-level acreage. This is shown in equations 2.16 and 2.17 respectively.

$$acres_{it}^{crn} = \mathbb{P}_{it} \times acres_i \tag{2.16}$$

$$qty_{it}^{crn} = \mathbb{P}_{it} \times acres_i \times yield_{it}^{crn} \tag{2.17}$$

Since only the crop choice probability is modeled as a function of prices, price fluctuations influence planted acreage and quantities through this probability term as shown in equations 2.18 and 2.19.

$$\frac{\partial acres_{it}^{crn}}{\partial P^k} = \frac{\partial \mathbb{P}_{it}}{\partial P^k} \times acres_i \tag{2.18}$$

$$\frac{\partial qty_{it}^{crn}}{\partial P^k} = \frac{\partial \mathbb{P}_{it}}{\partial P^k} \times acres_i \times yield_{it}^{crn} \tag{2.19}$$

From the summation rule of differentiation, aggregating these marginal effects to the national level is as simple as summing them across all i fields. Equations 2.20 and 2.21 show the estimated national aggregate planted corn acreage and 2.22 and 2.23 show the aggregate marginal effects for corn planted acres and quantity respectively.

$$acres_t^{crn} = \sum_i acres_{it}^{crn} \tag{2.20}$$

$$qty_t^{crn} = \sum_i qty_{it}^{crn} \tag{2.21}$$

$$\frac{\partial acres_t^{crn}}{\partial P^k} = \sum_i \frac{\partial acres_{it}^{crn}}{\partial P^k} \tag{2.22}$$

$$\frac{\partial qty_t^{crn}}{\partial P^k} = \sum_i \frac{\partial qty_{it}^{crn}}{\partial P^k} \tag{2.23}$$

Weighted acreage elasticities were calculated according to equation 2.24. Here P^k is the price of crop k, $\bar{P}^k$ is the national average of the k^{th} crop's price, and $acres_i$ is size of field i. Equation 2.25 shows the formula for the quantity weighted elasticity where $yield_{it}^{crn}$ is the estimated yield for each field estimated at the county-level using NASS data.[7]

$$\varepsilon^{acres} = \sum_i \left(\frac{\partial \mathbb{P}_{it}}{\partial P^k} \times acres_i \right) \frac{\bar{P}^k}{\sum_i \mathbb{P}_{it} \times acres_i} \tag{2.24}$$

$$\varepsilon^{qty} = \sum_i \left(\frac{\partial \mathbb{P}_{it}}{\partial P^k} \times acres_i \times yield_{it}^{crn} \right) \frac{\bar{P}^k}{\sum_{i=1} \mathbb{P}_{it} \times acres_i \times yield_{it}^{crn}} \tag{2.25}$$

2.3.1 Standard Error Calculation

Supply's response to price is the relationship of interest. Estimates of both the effects *and* their standard errors are needed to identify statistically significant effects. To characterize these standard errors, it is important to understand how prices vary over time. If futures prices varied over time while basis patterns remained perfectly fixed, then only the temporal variation would be used in estimating coefficients. Stable spatial correlation in prices violates the independence assumption in regression analysis and biases standard errors downward. Figures 2.1 and 2.2 show the box plots for the corn price and other price index for each year of the analysis. These plots show evidence of consistent basis patterns arising over time since the variance is similar in each year. This illustrates that the price variation is larger between years than it is within years. Taken together this means there are likely within-year dependence problems in the dataset.

The dependence of observations in the dataset complicates the estimation of the standard errors of the coefficients since it effectively shrinks the

[7] A simple linear trend model was used to estimate county yields.

degrees of freedom. A common way to correct the estimation of the standard error is to first suppose that observations can be clustered such that there is observational independence across clusters while allowing observational dependence within clusters. If the number of clusters is large, then the coefficient variance can be estimated with the "sandwich" matrix, an average of the variance-covariance matrix across the clusters. The goal in averaging is to produce a unbiased estimator for the coefficient variance. Therefore, when the number of clusters is "small" an unbiased estimate of variance cannot be obtained by computing the sandwich matrix. While the minimum number of clusters needed to obtain an unbiased estimator is debated in the literature, generally 30 to 50 clusters are considered minimally sufficient (Cameron and Miller, 2015; Cameron et al., 2008).

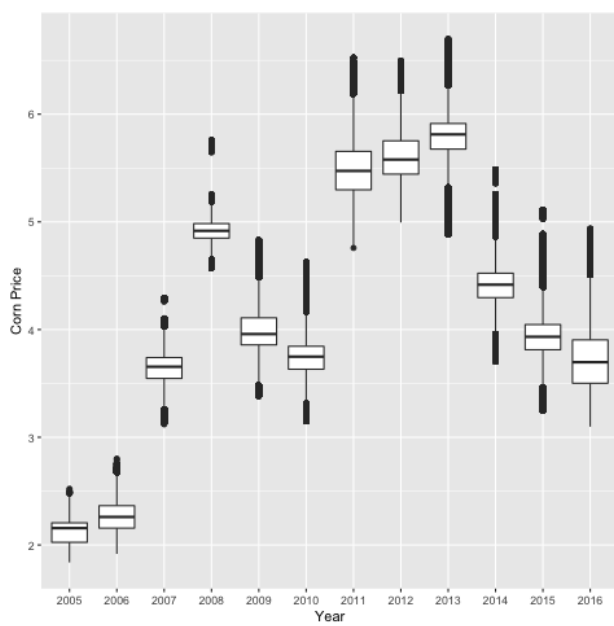

Figure 2.1: *Corn Price Distributions by Year*

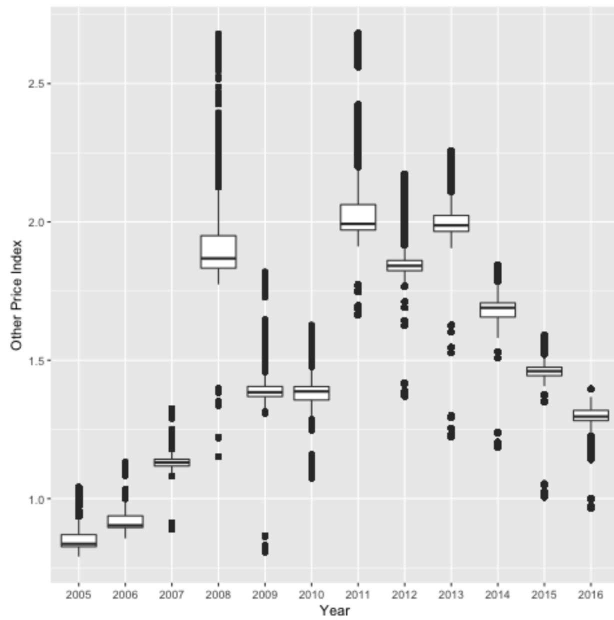

Figure 2.2: *Other Price Distributions by Year*

Under the standard criteria, the number of clusters in this analysis would be considered small since at most there are 13 clusters, one for each year between 2004 to 2016. In this case, the standard error can be estimated using the wild bootstrap. Unlike the usual paired bootstrap procedure the wild bootstrap preserves the distribution of the original error terms and independent variables. The wild bootstrap creates pseudo-error terms by multiplying the errors from the original regression by some random perturbation with a mean of zero and a unit standard deviation. This effectively ensures the pseudo-error terms have the same distribution as the original error terms. In the wild bootstrap, the pseudo-error terms are added to the corresponding dependent variable estimates to produce pseudo-dependent variables which are then used to obtain a new set of coefficients. The standard wild bootstrap is not practical when using the logit model or any other model estimated through maximum likelihood since errors are not observed. Instead I use the wild score bootstrap from Kline and Santos (2012) recommended by Cameron and Miller (2015).

The wild score bootstrap is similar to the standard wild bootstrap except the score function components are perturbed instead of the error terms. The score function of the logit model is shown in equation 2.26. Here y_{it} is the dependent binary variable, $\mathbf{Z}_{it}$ is a $(1 \times k)$ vector of independent variables, and θ is a $(1 \times k)$ vector of coefficients where i indexes observations. The wild score bootstrap creates coefficient replications by perturbing the original score function components by some random variable (W_t) where $\mathbb{E}[W_t] = 0$ and $\mathbb{V}[W_t] = 1$ and then performing an additional Newton-Raphson iteration using the perturbed score function, shown in equation 2.27. Here (n) is the total number of observations, $\left(\hat{\theta}\right)$ is the original coefficient vector and $(-\mathbb{H}^{-1})$ is the Fischer Information matrix. This bootstrap procedure has a key advantage that once the original model is estimated, bootstrapped coefficients can be produced analytically without re-estimating the model or computing the Fischer Information matrix at each iteration (Kline and Santos, 2012). This attractive feature of the score bootstrap keeps the 1,000 iteration bootstrap over the 32 million observations in this dataset tractable.

$$\frac{\partial \ell}{\partial \theta} = \sum_{i} \left[y_{it} - \left(1 + \exp\left\{-\mathbf{Z}_{it}'\theta\right\}\right)^{-1} \right] \mathbf{Z}_{it} \tag{2.26}$$

$$\hat{\theta}^{wild} = \hat{\theta} - \mathbb{H}^{-1}\frac{1}{n}\sum_{i}\left[y_{it} - \left(1 + \exp\left\{-\mathbf{Z}'_{it}\hat{\theta}\right\}\right)^{-1}\right]\mathbf{Z}_{it}W_t \tag{2.27}$$

To compute cluster robust statistics using the wild score bootstrap, a common perturbation value W_t is applied to every observation within each cluster by year (t). To simplify the wild bootstrap, the random perturbation (W_t) usually takes on discrete values. Popular choices of W_t include the Rademacher distribution where W_t takes on values of either 1 or -1 each with probability $\frac{1}{2}$, or the values suggested by Mammen (1993) where $W_t = -\frac{(\sqrt{5}-1)}{2}$ with probability $\frac{(\sqrt{5}+1)}{2\sqrt{5}}$ and $W_t = \frac{(\sqrt{5}-1)}{2}$ with probability $\frac{(1-\sqrt{5}-1)}{2\sqrt{5}}$. While simple, the small number of potential values for W_t restrict the number of samples that can be generated over the limited number of clusters. Regressions with a minimum of year of 2008 have only 9 clusters which means that there will only be ($2^9 = 512$) possible randomized samples using the traditional weighting. I instead use an alternative weighting scheme proposed by Webb (2013) which allows W_t to take on six values according to equation 2.28. This method helps retain a simple sampling distribution while increasing the number of potential random samples when there are only a few clusters (Webb, 2013). As a result, Webb's distribution expands the potential random samples of the nine-year clustered bootstrap to ($6^9 = 10,077,696$).

$$W_t = \begin{cases} -\sqrt{\frac{3}{2}} \text{ with probability } \frac{1}{6} \\ -\sqrt{\frac{2}{2}} \text{ with probability } \frac{1}{6} \\ -\sqrt{\frac{1}{2}} \text{ with probability } \frac{1}{6} \\ \sqrt{\frac{1}{2}} \text{ with probability } \frac{1}{6} \\ \sqrt{\frac{2}{2}} \text{ with probability } \frac{1}{6} \\ \sqrt{\frac{3}{2}} \text{ with probability } \frac{1}{6} \end{cases} \tag{2.28}$$

A series of 122 models were separately estimated to produce the original coefficients. While these models were estimated separately, their results need to be aggregated to produce national wide elasticity statistics. To ensure that for each group, clusters were treated uniformly across different replications, the clustered weights were applied identically over every model. That is, for each of the 1,000 replications, weights were randomly drawn for each year between 2004 and 2016 and identical replication-year weights were used in all of the 122 estimated models.

Note: I am currently in the process of computing the standard errors for the statistics to follow and will appear in future updates.

2.4 Data

Field Boundaries and Crop Choice

The goal of this study is to estimate the effect that price changes have on the likelihood that farmers plant corn relative to other crops while allowing for potential heterogeneity. Data were broadly used to (1) identify crop choices of individual producers over time, (2) incorporate relevant heterogeneity across the United States, and (3) serve as independent price variables and regression controls. The USDA's Cropland Data Layer (CDL) provides a set of farm-level crop choice observations from as far back as 1997 to present. The available crop price data restricts the analysis between 2004 and 2016. The CDL identifies crop-choice using remote sensing, primarily via satellite. This provides a categorical raster image of the CONUS at a 30m resolution. The 2008 Common Land Unit (CLU) shapefile dataset provides field-level boundaries across the United States.[8] Using these boundaries, I crop choices and other field-level attributes to a given field using raster cells that overlap with an off-centroid point in each field boundary. Areas of the country with missing common land unit boundaries were filled in with boundaries from Yan and Roy (2016).

Regions

To incorporate heterogeneity in this study, I estimate separate models over a set of geographic boundaries known as Major Land Resource Areas (MLRAs) that were established by the Natural Resource Conservation Service (NRCS). The set of MLRAs consist of 278 subregions within the US that are broadly categorized using regional characteristics such as physiography, geology, climate, water, soils, biological resources, and historic land use. These features are relevant for agricultural productivity making these regions a convenient way of incorporating heterogeneity in row crop agriculture across the coun-

[8]These boundaries were generously provided by Josh Woodard and Ag-Analytics (Woodard, 2016a,b).

try. Physiography considers the general elevation of the area above sea level in feet and relief. These features are associated with drainage properties. The geology criteria refers to general geologic properties of the land such as rock age. Climate delineations were produced using Parameter-Elevation Regressions on Independent Slopes Model (PRISM) data based on ranges of annual precipitation, seasonal precipitation distribution, annual temperature ranges, and seasonal freeze statistics. The water criteria considers water resources available, its quality and quantity, and water use within the region. This includes seasonal effects and intertemporal usage such as drought-year water usage. The soil criteria characterizes the soil taxonomy to the "great group" as defined in the Soil Survey Geographic (SSURGO) and SSURGO2 database. Soils at the great group level are defined by characteristics such as salinization, wetness, and other important soil properties such as fragipan which impacts water and root penetration (Soil Survey Staff, 2014). Biological resources involve the descriptions of the dominant flora and fauna in the area. The last category, land use, was produced using the 1997 National Resources Inventory (NRI) data on land use. The NRI consists of survey data collected at five year intervals at over 800,000 sample sites in the 50 United States, Puerto Rico and the US Virgin Islands. Land use categories used for MLRAs include cropland, grassland, forest, urban development, water, and other (Natural Resources Conservation Service, 2001; NRCS, 2006).

The use of MLRAs as a control for spatial heterogeneity is convenient since they simultaneously control for many important agricultural productivity predictors. The NRCS defines smaller subregions within MLRAs. However, most of these subregions do not have enough observations to model separately and therefore MLRAs were used instead.

MLRAs are attractive alternatives to using political boundaries such as counties since they are characterized by important agricultural features. However, a challenge of using CDL observations with MLRAs is that not all states entered the dataset in the same year. For example, North Dakota entered the CDL in 1997, its inaugural year, while Texas observations first became available in 2008. Models were run at the MLRA-level and to avoid issues where a subset of states were represented due to CDL data availability, I removed observations before the MLRA had full coverage. For instance, Iowa entered the dataset in 2004 while Minnesota entered the CDL in 2006. The dataset for an MLRA that overlaps with Iowa and Minnesota and no other state will begin in 2006 as this is the latest year between 2004 and 2006. Figure 2.3 shows the MLRA map and earliest year of analysis for each MLRA dataset.

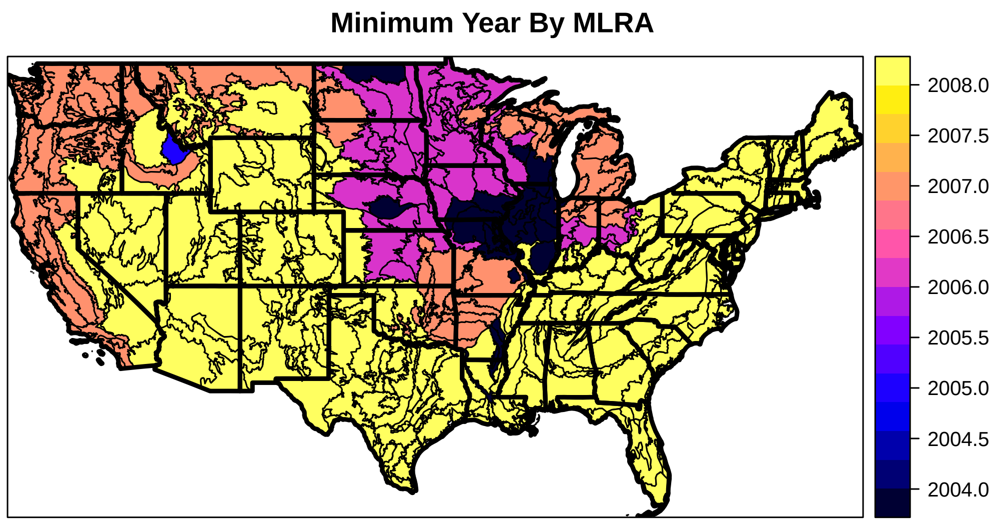

Figure 2.3: *Minimum Observation Year by Major Land Resource Area*

Soil texture statistics were used with MLRA boundaries to characterize model heterogeneity. To do this I constructed the commonly used 12 texture classification groups using SSURGO designations and soil texture percentages (Benham et al., 2009). I then aggregated the 12 classes into 5 classes according to table 2.1. The 5-group classification tended to retain more observations within each group which ensured there were an adequate number of observations in each regression. Even so, there were many instances where the within-MLRA texture classification groups were too small to reliably estimate the models. In this case, the soil texture classifications were aggregated yet again. If a soil texture group within an MLRA had less than 20,000 observations, it was combined into the next closest group within an MLRA. The "closeness" of these groups was determined by the distance between the mean values of silt and clay percentages that define the texture groups. For instance, the "clayey" five group texture classification has a mean proportion of 22.3% silt and 58.6% clay and the medium class has a mean proportion of 25.5% silt and 13.6% clay. This means that the two groups have a distance of $(0.223 - 0.255)^2 + (0.586 - 0.136)^2 = 0.204$. Table 2.2 shows the soil texture distances.

Table 2.1: *Soil Texture Classifications*

Group 12 Desig.	Group 5 Desig.
Sand	Sandy
Loamy Sand	Sandy
Sandy Loam	Moderately Sandy
Silt	Medium
Silt Loam	Medium
Loam	Medium
Clay Loam	Moderately Clayey
Sandy Clay Loam	Moderately Clayey
Silty Clay Loam	Moderately Clayey
Sandy Clay	Clayey
Silty Clay	Clayey
Clay	Clayey

Table 2.2: *Soil Texture Group Distances*

Texture Group	Clayey	Medium	Mod. Clayey	Mod. Sandy	Sandy
Clayey	0	0.204	0.104	0.416	0.681
Medium	–	0	0.0445	0.158	0.365
Mod. Clayey	–	–	0	0.113	0.286
Mod. Sandy	–	–	–	0	0.0435
Sandy	–	–	–	–	0

Prices

The importance of including and expectations in prices has been a persistent issue in agricultural supply response literature (Gardner, 1976; Haile et al., 2016; Miao et al., 2016; Nerlove, 1956; Roberts and Schlenker, 2013). A common theme is that, when modeling planting decisions, it is important to use prices that reflect harvest-time expectations at or before the time planting. The earliest and simplest forms of expected prices were simply the previous years lagged harvest price (Lee and Helmberger, 1985; Whittaker and Bancroft, 1979). In Nerlove's famous supply response paper, he assumed that price expectations come only from past realizations of prices and used lagged prices and lagged dependent variables as right hand side variables(Nerlove, 1956). Others have used futures prices since, under the efficient market hypothesis, these prices should reflect information about expected future price changes (Gardner, 1976; Haile et al., 2014). Roberts and Schlenker (2013) argue that futures prices are endogenous to expected plantings and argue that they should be instrumented using purely exogenous variables such as past weather, and yield shocks. Hendricks et al. (2014a) replicate this study and find that instrumenting futures prices is unnecessary and harmed model precision (Hendricks et al., 2014a; Roberts and Schlenker, 2013).

I employ a mixture of pre-plant spot and futures prices to form the expected price series. To be clear, the expected prices are the harvest prices that farmers expect as they make their planting decisions. Under the efficient market assumption, the price of the harvest-time futures contract represents the expected price of the commodity at harvest time in a delivery location. The nearby futures price is the expected price of the commodity at the delivery location at some imminent date. The difference between these two contracts gives the market's expected cost of carrying the product between the nearby date and harvest time. That is, it is the market's expected cost that a farmer would incur if she were to store grain until harvest time. Adding the current local spot price to the cost of carry gives the expected harvest-time price for a given location.[9]

Much of the nation's corn acreage is planted in the month of April with planting starting in early April or late March. I assume that the planning process begins in the months of January and February as this gives time for required crop specific land preparation before planting begins. To construct expected prices, I first average local daily spot prices over the course of the months of January and February which I call the planting price $\left(C_{it}^{P}\right)$. Next, I average daily the nearby futures contract price $\left(F_{t}^{P}\right)$ and the harvest-time futures contract price $\left(F_{t}^{H}\right)$ for the respective commodities in January and February. I construct the expected harvest-time spot price expectations according to equation 2.29.

$$E_{it}\left[P_{it}^{H}\right] = F_{t}^{PH} + \left[C_{it}^{P} - F_{t}^{PN}\right] = \underbrace{\left[F_{t}^{PH} - F_{t}^{PN}\right]}_{\text{Expected Cost of Carry}} + C_{it}^{P} \tag{2.29}$$

In this analysis I use the prices of corn, soybeans, hard red winter wheat, hard red spring wheat, soft red winter wheat, rice, and cotton. These prices are all quoted in dollars per bushel with the exceptions of rice and cotton

[9] An alternative interpretation of this expected price comes from the perspective of a farmer performing a short hedge. In a short hedge, farmers sell harvest-time futures before planting. They would then buy back the futures contract and sell their commodity at harvest time. As a result they would receive the harvest-time basis (the spot price minus the buyback futures price) and initial future sale price. If the pre-plant basis for the nearby contract is consistent with the nearby harvest-time basis, then this also equals the expected harvest-time price for the farmer.

which were quoted in dollars per pound. Daily values for futures prices for these commodities are readily available from the Data Transfer Network (DTN). Spot price datasets were constructed using a combination of data from the Data Transfer Network (DTN) and Cash Grain Bids (CGB). The pre-plant price was the market's average expected price in January and February. To ensure that the estimates are not dictated by a small number of observed prices, I remove markets with less than 10 spot-price observations over January and February. Figure 2.4 shows the price coverage by crop. Between 2004 and 2016 there were 1,367 corn, 1,252 soybean, 84 hard red spring wheat (HRSW), 96 hard red winter wheat (HRWW), and 123 soft red winter wheat (SRWW) price locations with consistent observations. Rice and cotton prices were collected from the National Agricultural Statistics Service's (NASS) at the national level from 2004 to 2016. Coverage for the continuously observed local markets was rather good, and densely covers most of the major field crop production areas.

Using the center of the market city as a reference, I used these prices to construct annual basis maps for each commodity over the contiguous United States. After estimating the expected commodity prices over the markets I interpolated these estimates using ordinary kriging. Ordinary kriging has advantages over other interpolation procedures such as inverse distance weighting since it takes the spatial correlation of the observations into account to minimize the variance of the estimates.[10] Basis map values were estimated

[10]Ordinary kriging requires the estimation of spatial dependence and the variogram. While estimation problems can persist, much of the observations were within areas with dense market coverage. It is therefore unlikely that error from variogram misspecification would influence the bulk of used areas of the map.

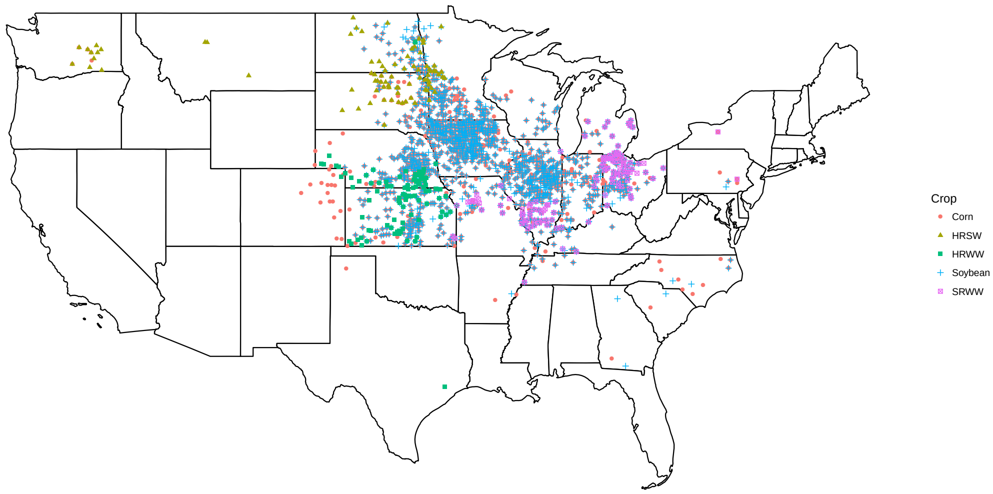

Figure 2.4: *Commodity Price Locations Continuously Observed Between 2004 and 2016*

over a raster image with a resolution of 0.01 square degrees (or approximately 36 square miles). Figure 2.5 shows an example of the corn basis map in 2009 and their respective standard errors for the localized markets of corn. To maintain consistency of the price expectations estimates over time, the original observation set contains only markets with continuously observed price averages in every year from 2004 to 2016.

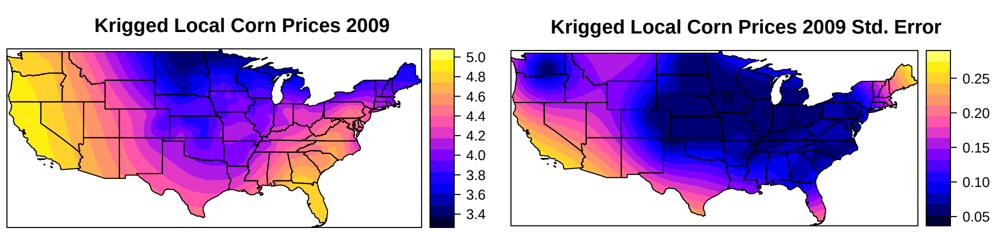

Figure 2.5: *2009 Corn Expected Price Map*

In this study I characterize the other crop price using a weighted average of soybeans, HRWW, SRWW, HRSW, cotton, and rice prices. Using a common weighting scheme for every field is problematic since the set of relevant alternative crops to corn production will differ by the area of the country. For instance, cotton may be a relevant alternative crop in the Mississippi River Delta area but an irrelevant choice in Wisconsin. To account for this, I construct the other crop price as a Laspeyres Index. This is the same index that is used to compute the Consumer Price Index. To track price changes within a given region I create a "basket" of commodities indexed from $k = 1, \ldots, K$ using produced quantities of these crops in some base period (period 0). Equation 2.30 shows the functional form where P_{it}^k is the price for commodity k at time t and q_0^k is the total quantity of crop k produced in period 0. Using unique price indices for each MLRA ensures that the other crop price will largely consist of crops grown in the region. For instance, the dominant alternative crop to corn in the state of Iowa is soybeans. If the alternative acreage only consists of soybeans, then $q_0^{k'} = 0 \; \forall k'$ where k' is not soybeans, which means that only soybeans prices would enter the price index.

$$P_{it}^O = \frac{\sum_{k=1}^K P_{it}^k q_0^k}{\sum_{k=1}^K P_{i0}^k q_0^k} \tag{2.30}$$

There are several complications to using the standard Laspeyres index in this study. First, crop choices are subject to change over time so it is unclear whether using the observed quantity produced in a single period q_0^k sensibly represents the typical crop choice basket over the course of the study. Second, the analysis is at the CLU level, each of which has a single observation in each year. A particular crop planted at the beginning of the analysis does not preclude another crop from being considered in the future. Properly defining of the initial quantity q_0^k is important so the alternative crop price index properly represents the prices of corn alternatives across different areas of the country. Lastly while I observe crop choices on some fields before 2008, the CDL dataset did not gain full coverage until 2008. To address these issues, I modified the Laspeyres index. In particular I define q_0^k as the MLRA-specific total production of crop k from 2008 to 2016. I compute total production by each MLRA for each crop by merging the field-level data with a county-level yield dataset, multiplying the yield with the field-level acreage choices, then summing over the MLRA from 2008 to 2016. Average annual yield data for each of these crops are available at the county-level through the National Agricultural Statistics Service (NASS).[11]

Controls

The final component to the data are the field-level controls. These controls incorporate individual field heterogeneity and reduce potential bias in the price coefficient estimates. To control for soil differences at the field-level, I include the National Commodity Crop Productivity Index (NCCPI) and the field's slope as regressors. The NCCPI index takes many facets of the soil pertaining to productivity into account. More detail of the NCCPI values can be found in Dobos et al. (2008). The slope of the field is a key determinant

[11]The computation of this crop basket assumes that every crop grown in an area from 2004-2016 was a relevant crop over the entire course of these years. While technological changes could introduce some of these crops over time, incorporating technological trends into the production basket weakens the effectiveness of the price index since temporal variation in the index will not entirely be due to price variation.

of runoff, and erosion and is a common control in the literature (Wang et al., 2015; Wu et al., 2004a).

Weather is also a factor that can influence crop choice. Since the analysis is over planting decisions, pre-plant weather is of interest. Extremely wet conditions can potentially delay planting and could cause farmers to plant alternative crops such as soybeans with later planting dates. Extremely dry conditions may have similar effects and may persuade farmers to switch to more drought resistant crops. I incorporate extreme planting precipitation conditions with two binary indicator variables. These variables were calculated using planting precipitation data percentiles from the field's respective MLRA between 1983 and 2016. The first variable indicates whether the field experienced exceptionally dry pre-planting conditions in a given year and equaled one if the April-May precipitation was at or below the MLRA's the 25th percentile. [12] The second variable indicates wet planting conditions and equals one if the field's April-May precipitation was at or above the MLRA's 75th percentile precipitation. Lastly, I control for irrigation by including a 2012 Moderate Resolution Imaging Spectroradiometer Irrigated Agriculture Dataset (MIrAD) dataset which gives information on the irrigation status over the conterminous United States. This variable equals one if the field was irrigated and zero if non-irrigated (Brown and Pervez, 2014).

[12] Corn planting for many of the largest corn producing states is most active in these months (NASS, USDA, 2010).

Crop Definitions and Data Filtering

Due to the diversity of suitable crops, constructing a parsimonious supply response model over an area as large as the United States is challenging. The Cropland Data Layer distinguishes a variety of crops, however their respective prices are more difficult to find. In addition, many crops do not have associated futures contracts which makes expected prices difficult to construct. Since I account for heterogeneity using pooled regressions at the MLRA-texture group level, I needed to ensure each of the regressions have enough observations to produce reliable estimates. The Markov transition probability regression modeling strategy requires that the training data consists of only observations with two consecutive corn or other-crop choices which can constrain the MLRA-texture group sample size.

For these reasons I divided the observation strata into 5 groups: corn, priced crops, other crops, cropland, and non-cropland. Table 2.3 shows the CDL observation designations. Corn consists of observations where the CDL signifies only a conventional corn observation. That is, "corn" does not include double-cropping observations involving corn (e.g. double-cropping corn and soybeans), or less conventional varieties such as sweet corn, or popcorn. Priced crops are crops with expected prices that enter the "other" crop price index value (soybeans, rice, non-Durham wheat varieties, and cotton) and associated double-cropped observations with these crops (e.g. winter wheat-cotton double cropped observations). Other crops are crops that are assumed to be substitutes in production to the priced crops. This category includes double-cropped observations containing these crops. The

"Cropland" category contains other crops that are less substitutable to the priced crops. This category includes specialized fruit and vegetable crops and perennial crops such as alfalfa. The final category, non-cropland, contains land uses that are not immediately suitable for crop production including marshland, pasture, and developed lands.

Table 2.3: *Cropland Data Layer Observation Designations*

Corn	Sweet Potatoes	Greens	Squash
Cotton	Triticale	Herbs	Strawberries
Rice	Alfalfa	Honeydew Melons	Sugarcane
Soybeans	Almonds	Lettuce	Sweet Corn
Spring Wheat	Apples	Mint	Switchgrass
Winter Wheat	Apricots	Misc Vegs	Tobacco
Barley	Aquaculture	Nectarines	Tomatoes
Buckwheat	Asparagus	Olives	Turnips
Camelina	Blueberries	Onions	Vetch
Canola	Broccoli	Oranges	Walnuts
Dry Beans	Cabbage	Other Crops	Watermelons
Durum Wheat	Caneberries	Other Hay/Non Alfalfa	Barren
Fallow/Idle Cropland	Cantaloupes	Other Tree Crops	Clouds/No Data
Flaxseed	Carrots	Peaches	Deciduous Forest
Hops	Cauliflower	Peanuts	Developed (All Levels)
Lentils	Celery	Pears	Evergreen Forest
Millet	Cherries	Peas	Forest
Mustard	Chick Peas	Pecans	Grassland/Pasture
Oats	Christmas Trees	Peppers	Herbaceous Wetlands
Other Small Grains	Citrus	Pistachios	Mixed Forest
Potatoes	Clover/Wildflowers	Plums	Nonag/Undefined
Rape Seed	Cranberries	Pomegranates	Open Water
Rye	Cucumbers	Pop or Orn Corn	Perennial Ice/Snow
Safflower	Eggplants	Prunes	Shrubland
Sorghum	Fruits	Pumpkins	Water
Speltz	Garlic	Radishes	Wetlands
Sugar Beets	Gourds	Shrubland	Woody Wetlands
Sunflower	Grapes	Sod/Grass Seed	
Legend			
Priced Crop	Other Crop	Cropland	Non-Cropland

Using the CDL classifications in table 2.3, I remove MLRA-texture groups using a set of three hurdles. The first hurdle filters out MLRA-texture groups with less than 20% of its total acreage in corn, priced crops, other crops, or cropland to remove areas with low agricultural activity such as desert, and mountainous or developed areas. The second hurdle ensures that the price index reasonably applies to relevant alternatives to corn production. Since the price index is constructed using prices from the "priced" crops and need to represent the prices of the "other" crops, the second hurdle removes MLRAs where less than 50% of the other crop acreage is composed of priced acreage. Lastly, since corn is the main crop of interest, the third hurdle ensures the MLRA-texture groups have enough corn observations to effectively model the relationship between corn plantings and price. This threshold removes MLRAs with less than 10% of their total other crop acreage in corn. In addition to these three hurdles, I also remove MLRA-texture groups with less than 50,000 total observations and MLRA-texture groups where less than 20,000 observations enter either of their respective Markov transition regressions.[13]

Data Summary and Probability Estimates

Table 2.4 shows the summary statistics for corn plantings, and the field controls. After filtering, the data include a total of 69 MLRAs and over 32 million individual observations across 3.6 million fields. I estimated a total of 244 separate Markov chain transition regressions across 122 total

[13] Observations also exclude CLUs that are smaller than 15 acres.

MLRA-texture group subsamples. According to the Cropland Data Layer and county-level NASS estimated yields, these MLRAs accounted for over 90% percent of the national corn production between 2009 and 2016 and over 80% of corn planted acres when compared with NASS estimates. While this may suggest that the sample is biased toward areas with higher yields, the sample omits fields that are less than 15 acres since these fields make up a relatively small portion of overall supply and area. Smaller fields are known for being less productive (Lubowski et al., 2008). However, adding these fields would increase the sample size and, by extension, the computational burden considerably. The average field in the sample had a 40% chance of planting corn in a given year from 2009 to 2016. A mean corn planting probability under 50% should be expected since, under many rotations, corn is rotated out for another crop every other year.

Table 2.4: *Summary Statistics and NASS Comparison*

Statistic	Dataset Value	NASS Value	Pct. of NASS Value
Total Corn Acreage*	596,968,262	731,786,000	81.58%
Total Corn Qty. (bu.)*	95,040,212,033	105,300,000,000	90.26%
Mean Prob. of Corn Planting*	40.90%		
No. of Obs.	32,127,647		
No. MLRAs	69		
No. MLRA-Texture Groups	122		
No. of Fields	3,694,282		
Total Field Acreage	217,310,028		
Mean Field Size (acres)	58.82		
Share of Irrigated Fields	8.52%		
Mean Field Slope	3.09%		
Mean NCCPI Soil Index	0.58		

* Indicates values from 2009 to 2016

I also present the computed the mean sequential and state probabilities by year. The state probability of planting corn in a particular year is equal to the probability of performing an other-corn planting sequence plus the probability of performing a corn-corn sequence. Figure 2.6 shows the mean values for the sequential and steady-state probabilities by year with the accompanying price variables.[14] This figure shows that probabilities have the expected correlations with prices, the probability of planting a continuous corn or a other to corn sequence, and the steady state probability of planting corn are all positively correlated with the corn-other price ratio while continuous other sequences are negatively correlated with the corn-other price ratio. Rotations are a disproportionately important component of the steady state probability of corn. Recall that this is equal to the probability of planting a continuous corn sequence plus the probability of planting a other-to-corn sequence. Figure 2.6 shows that roughly two-thirds of fields planted to corn transitioned from the other crop in the previous year.

[14] Note that the probability of performing a (Corn-Other) planting sequence will equal the probability of the (Other-Corn) sequence since these probabilities are estimated at the steady-state.

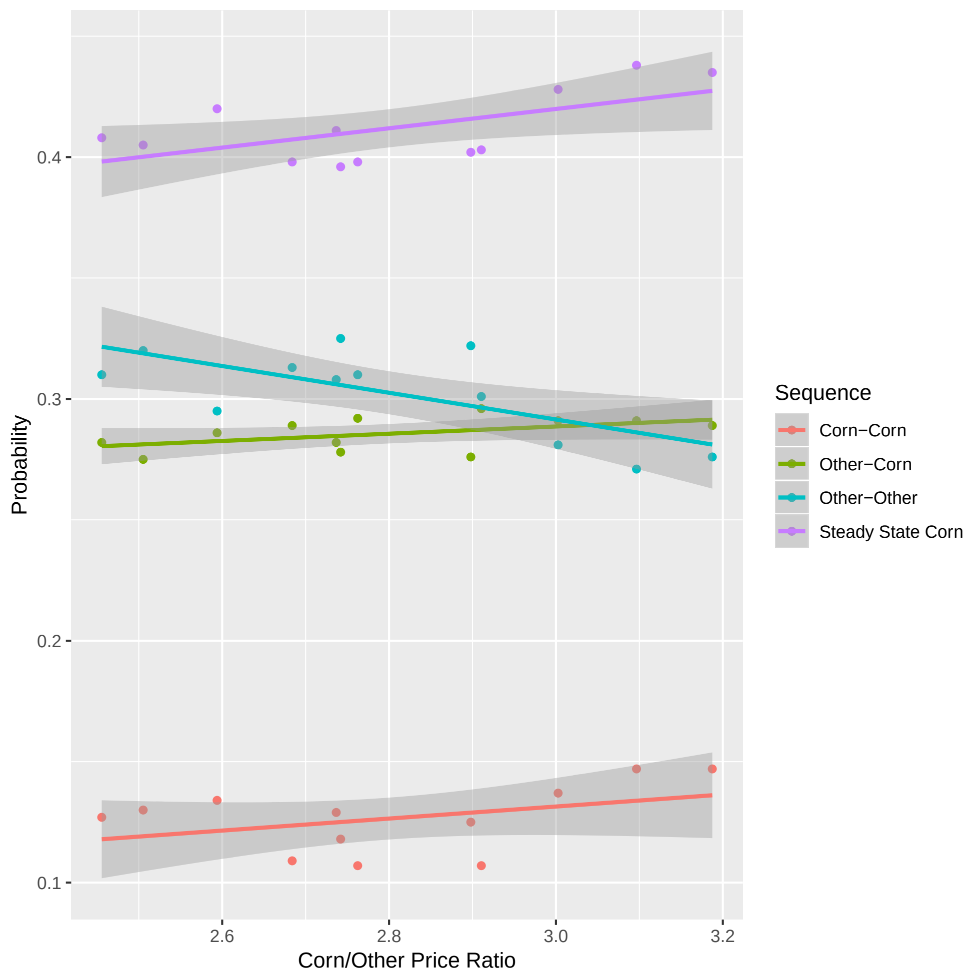

Figure 2.6: *Annual Sequential Probabilities and Corn-to-Other Price Ratios*

2.5 Results

2.5.1 Elasticity Comparison

In this first subsection of the results I compare the short and long-run own-price corn in this paper acreage elasticities with others found in the literature. Since the goal of this paper is to compare regions of the country and national-level elasticities. I compare elasticities from studies that conducted analysis over a multi-state area summarized in table 2.5. Many of the earlier studies employed simple models over short state-level panels and had elasticities around 0.2. Whittaker and Bancroft (1979) used a simple log-log OLS model with a lagged price variables across a 44 observation state-level panel. Lee and Helmberger (1985) estimated separate models over different years to study the impact of farm programs. They used 3SLS procedure across four states to take advantage of autocorrelation across time and contemporaneous correlation across states. When accounting for farm programs, they get a similar own-price acreage elasticity similar to Whittaker. It was popular over the 1990s to incorporate risk preferences into elasticity estimates, Chavas and Holt (1990) and Chavas et al. (1996) used national-level data to estimate corn and soybean acreage elasticities. To incorporate risk preferences, they used the expected utility model framework using mean and variance of per acre returns, and farm values as farm wealth proxies. Holt (1999) also utilized a risk-preferences framework assuming that producers maximize the certainty equivalent when making crop choices. Holt (1999) had by far the largest elasticities in this set finding elastic planted corn acreage response. He suggests that a the relatively short panel in part attributed to the unusually elastic estimate. Langpap and Wu (2011) model crop choice at the Natural Resources Inventories (NRI) site level. NRI site data is collected and distributed every 5 years. In their model, farmers make crop decisions to maximize utility and are modeled using a multinomial logit framework which they used to distinguish to non-crop and crop decisions and between different crops (e.g. corn,soybeans,wheat,hay). While they had a rich dataset over a large area of the corn belt, only 20% of their dataset could used for modeling due to the computational complexity of the multinomial logit model.

Huang et al. (2010) model corn, soybeans and wheat production and incorporate the influence of yield response in their estimates. Kim and Moschini (2018) use a similar model but exploit the temporal price variation that resulted from demand changes due to Renewable Fuel Standard legislation between 2005 and 2007. Both Huang et al. (2010) and Kim and Moschini (2018) utilize single differenced Arellano-Bond estimators which estimates supply response with a fixed-effect framework and includes a differenced lagged independent variable using the generalized method of moments. In their paper Hendricks et al. (2014c) used the Markov transition regression used in this paper to estimate supply response over Iowa, Illinois, and Indiana. They compared their estimates with the estimates of the fixed-effects models including the Arellano-Bond estimator and found that fixed effect approaches tend to to underestimate supply elasticities and generally had considerably larger standard errors. They also found that modeling over county-level variables tended to bias long-run effects upward and short-run effects downward. Both Hendricks et al. (2014c) and Kim and Moschini (2018) found that the short-run elasticity was larger than the long-run elasticity. Crop production is unique in that oscillating production between different products improves yields, reduces input usage, and reduces production risk. These rotational benefits encourage producers to resume standard cropping patterns after a price shock leading to a smaller longer-run effect of prices.

The nationwide results in this paper were estimated over a representative field-level sample which tracks the decisions and field-level controls of over 3.6 million fields. As such, the results from this model should not suffer from pooled bias that others modeling with county or state-level statistics. Because the total area consists of over 80% of the corn-growing area of the country over the years of the analysis, the data is also likely representative at the national level. In this paper I provide two sets of elasticities estimates, the first is over four states within the corn belt, Iowa, Illinois, Indiana, and Nebraska and the second is at the national level. The Corn Belt elasticities were similar to the ones found by Hendricks et al. (2014c). This is unsurprising since this paper uses a similar method over a similar area. Furthermore, this procedure allows for heterogeneity in planting response across the country and likely provides a good quantity and acreage response estimates. The Corn Belt estimates were below the estimates in Kim and Moschini (2018) but this likely arises from differences in the growing areas. The national elasticity estimates were large relative to many of the other estimates. The expanded and more heterogeneous area considered in this study is likely a contributer to the larger estimates. Areas that are less dependent on consistent corn production may be more likely to plant corn after a price increase since corn is a less competitive crop choice. A smaller change in the relative price of corn in these areas may be more decisive in determining the crop choice relative to areas where corn is grown on a regular basis.

Table 2.5: *Own-Price Corn Acreage Elasticity Estimates From the Literature*

Study	Years	Coverage	Resolution	SR Elast.	LR Elast.
Whittaker and Bancroft (1979)	1963-1974	IA,IL,IN,OH	State	0.221	–
Lee and Helmberger (1985)	1948-49, 1951-53, 1980	IA,IL,IN,OH	State	0.118	–
	1961-73, 1978-79	IA,IL,IN,OH	State	0.249	–
Chavas and Holt (1990)	1954-1985	United States	Country	0.166	–
Chavas et al. (1996)	1954-1985	United States	Country	0.249	–
Holt (1999)	1991-1995	IA,IL,IN,MI,MN,MO,OH,WI	State	1.04	–
Lin and Dismukes (2007)	1991-2001	IA,IL,IN,MI,MN,MO,OH,WI	State	0.345	–
Langpap and Wu (2011)	1979-1997	58,579 NRI sites in Corn Belt	NRI Site	0.246	–
Huang et al. (2010)	1977-2007	U.S. counties	County	0.510	0.980
Hendricks et al. (2014c)	2000-2010	IA,IN,IL	Field	0.40	0.29
Kim and Moschini (2018)	2005-2015	12 Midwest States	County	0.50	0.39
This Paper	2009-2016	IA,IL,IN,NE	Field	0.390	0.259
	2009-2016	United States	Field	0.674	0.526

2.5.2 Control Coefficient Summary

While the coefficient values for the control variables are not the primary interest of this paper, they help characterize why differences in planting probabilities arise across the country. The coefficient values in logit models do not correspond to marginal effects however their signs are consistent with the signs of the corresponding marginal effects. Since many of the variables act as controls and are not the primary focus of the analysis, coefficient statistics are shown in tables 2.6 and 2.7 to illustrate the general relationship between the transition probabilities and the controls. Coefficient values varied widely across the different models. Across every control, the minimum coefficient value was negative and the maximum coefficient value was positive. The coefficient signs within the quartiles were largely consistent between the transition probability regressions. The price coefficients conform to economic theory. Corn transitions tend to be positively correlated with corn price and negatively correlated with the other price index. The probability of planting corn was negatively correlated with the slope of the field. This is consistent with expectations since fields with steeper grades require more fertilizer and herbicide due to runoff and are generally more difficult to farm. The positive dry planting indicator coefficients and the negative wet planting indicator coefficients were also expected. Planting is time consuming and requires a large amount of workable field hours within specific windows of the year. Dry conditions during the corn planting season ensure farmers that planting can conclude within the optimal seasonal window. Conversely wet planting conditions create a host of problems including reduced yields and severe soil compaction and planting delays also lead to yield penalties (Farnham, 2001; MacKellar and Anderson, 2016). Since corn is generally planted earlier in the year, it is not surprising that wet pre-plant conditions were negatively correlated with corn planting. The productivity of the soil also had a positive effect on the probability of planting corn. The positive coefficient on the soil productivity index is expected since, from a nutrient standpoint, corn is a relatively demanding crop. Irrigated fields were more likely to plant corn. Corn tends to have a more demanding evapotranspiration requirements. It is therefore not surprising that transitions to corn are more probable on fields with irrigation (Stone and Schlegel, 2006).

Table 2.6: *Other-Corn Markov Transition Regression Coefficient Summary*

Coefficient	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Intercept	-264.76**	-136.747***	-73.468***	-68.154***	-14.969	138.297***
	(107.705)	(9.043)	(5.843)	(8.969)	(9.754)	(29.582)
Corn Price	-0.752***	0.089*	0.378***	0.236***	0.52***	2.203***
	(0.169)	(0.052)	(0.045)	(0.056)	(0.073)	(0.309)
Other Price	-5.253***	-1.084***	-0.708***	-0.425**	0.101	2.228***
	(0.709)	(0.190)	(0.129)	(0.168)	(0.164)	(0.610)
Slope	-0.471***	-0.052***	-0.012***	-0.009	0.036***	0.237***
	(0.091)	(0.003)	(0.004)	(0.006)	(0.008)	(0.037)
Precip. Q1	-0.705***	-0.055	0.098**	0.107***	0.192***	1.037***
	(0.175)	(0.048)	(0.040)	(0.037)	(0.042)	(0.216)
Precip. Q3	-0.605*	-0.23***	-0.103***	-0.103***	0.042	0.572***
	(0.326)	(0.034)	(0.026)	(0.024)	(0.027)	(0.212)
NCCP Soil Index	-10.379***	0.236***	1.25***	1.09***	2.238***	8.058***
	(0.369)	(0.035)	(0.032)	(0.042)	(0.063)	(0.870)
Irrigation Status	-1.151***	-0.099	0.221***	0.194**	0.46***	2.62***
	(0.169)	(0.075)	(0.075)	(0.085)	(0.082)	(0.039)

Note: Std. Errors in (), *** 1%, ** 5%, * 10% Sig. Lvl.

Table 2.7: *Corn-Corn Markov Transition Regression Coefficient Summary*

Coefficient	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Intercept	-408.784***	-5.457	40.08***	55.217***	90.063***	206.754***
	(71.910)	(11.801)	(13.278)	(11.765)	(16.177)	(34.704)
Corn Price	-0.373**	0.083**	0.271***	0.243***	0.413***	1.329***
	(0.158)	(0.039)	(0.046)	(0.047)	(0.054)	(0.162)
Other Price	-3.004***	-1.111***	-0.59***	-0.465***	-0.043	1.511**
	(0.476)	(0.164)	(0.135)	(0.140)	(0.115)	(0.639)
Slope	-0.318***	-0.036***	-0.005***	-0.007***	0.024***	0.408***
	(0.034)	(0.002)	(0.002)	(0.002)	(0.002)	(0.099)
Precip. Q1	-0.533***	-0.097***	0.054***	0.029	0.18***	1.011***
	(0.097)	(0.025)	(0.020)	(0.021)	(0.023)	(0.169)
Precip. Q3	-0.411**	-0.132***	-0.016	0	0.088***	0.456***
	(0.163)	(0.015)	(0.010)	(0.015)	(0.017)	(0.079)
NCCP Soil Index	-4.462***	-0.88***	-0.293***	-0.19***	0.633***	3.528***
	(0.320)	(0.055)	(0.024)	(0.051)	(0.055)	(0.453)
Irrigation Status	-10.806***	-0.256**	-0.066	-0.015	0.293***	1.331***
	(0.307)	(0.121)	(0.099)	(0.102)	(0.074)	(0.083)

Note: Std. Errors in (), *** 1%, ** 5%, * 10% Sig. Lvl.

State Probability Elasticities and Marginal Effects

I now move to the primary focus of the paper, the relationship between corn plantings and prices. Table 2.8 shows the marginal effects and elasticities of the steady-state corn-plant probability with respect to prices. Like the coefficients in the transition equations, the price marginal effects and elasticities on corn probability were highly varied across the models. The 1st and 3rd quartiles of the elasticities are less than one suggesting that planting response is inelastic to prices. There were however, extremely sensitive areas with elasticities reaching as high as 4 which indicate that these areas are highly responsive to prices. Figure 2.7 shows that fields that were highly sensitive to corn prices were also highly sensitive to other prices. While extensification of corn production may occur due to price changes, the high degree of correlation between the corn and other-price elasticities suggests that cropland is not disproportionately added in areas due to higher corn prices.

The interquartile ranges of the marginal effects conform with economic theory. Higher corn prices tended to coincide with greater likelihood of planting corn and higher prices for other competing crops tended to reduce the likelihood of planting corn. The results from table 2.8 also show that the supply response in the short-run is larger than the supply response in the long-run which (Hendricks et al., 2014c) showed occurs due to temporary rotational changes from price changes. Since there are benefits to rotating crops, producers are more likely to perform short-run adjustments to the rotational pattern as a result of a price change than to alter their rotations over longer periods of time.

Table 2.8: *Summary of Marginal Effect and Elasticity Price Response Statistics*

Statistic	Min	1st Quartile	Median	Mean	3rd Quartile	Max
SR Corn ME	-0.066	0.024	0.045*	0.057***	0.079***	0.318***
	(0.107)	(0.027)	(0.026)	(0.020)	(0.021)	(0.084)
LR Corn ME	-0.077	0.018	0.033**	0.044**	0.057***	0.4**
	(0.102)	(0.029)	(0.016)	(0.018)	(0.005)	(0.172)
SR Other ME	-0.705***	-0.187***	-0.094***	-0.114***	-0.026***	0.237***
	(0.030)	(0.007)	(0.009)	(0.007)	(0.010)	(0.047)
LR Other ME	-0.944***	-0.137***	-0.065	-0.092**	-0.022	0.231
	(0.167)	(0.039)	(0.053)	(0.045)	(0.067)	(0.267)
SR Corn Elast	-0.541	0.269	0.461*	0.714***	0.815***	4.518***
	(0.911)	(0.293)	(0.165)	(0.251)	(0.054)	(1.839)
LR Corn Elast	-0.683	0.178	0.346**	0.605**	0.623***	4.261**
	(0.724)	(0.155)	(0.195)	(0.241)	(0.287)	(0.434)
SR Other Elast	-4.332***	-0.683***	-0.312***	-0.558***	-0.119***	0.474***
	(0.871)	(0.294)	(0.269)	(0.258)	(0.214)	(1.193)
LR Other Elast	-4.086***	-0.542***	-0.24	-0.488**	-0.094	0.375
	(0.183)	(0.024)	(0.029)	(0.035)	(0.045)	(0.095)

Note: Std. Errors in (), *** 1%, ** 5%, * 10% Sig. Lvl.

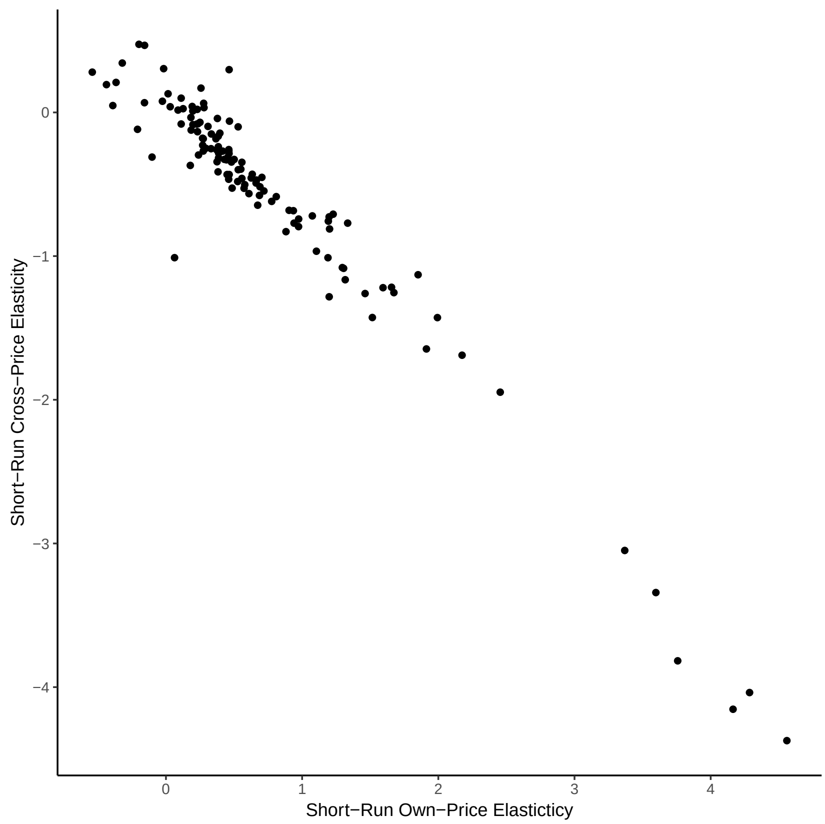

Figure 2.7: *Average Group Short-Run Elasticities*

2.5.3 Supply Response Across Space

Modeling at the subnational level allows for price response comparisons across the entire country. Heterogeneity in supply response was allowed across a total of 122 MLRA-soil texture groups subdivided over 69 MLRAs across the CONUS. Figure 2.8 and 2.9 show the field-level own and cross-price marginal effects on the probability of planting corn respectfully.[15] These results illustrate the response heterogeneity across the country. Generally speaking, supply inside the traditional Corn Belt tended to have a middling response to corn prices. In states like Iowa, Illinois, Indiana, and Nebraska, a $1 increase in the price of corn increases the steady-state probability of planting corn by approximately 10%. Corn supply in the eastern Dakotas, western Minnesota, southern Wisconsin, central Michigan, and the Mississippi River delta was more sensitive to price fluctuations since a $1 increase in the price of corn would increase corn plantings by nearly 30%. Not every area outside of the Corn Belt had such sensitive corn supply however. Corn planting decisions were less sensitive to prices in states like Kansas, and areas along the east coast. This could be due to heterogeneity in growing conditions. In particular, western Kansas is prone to droughts and relies heavily on irrigation.

The disproportionately high drought pressure in the production areas Kansas and Colorado could be the reason that production does not respond much to prices as general drought conditions in the US tend to elevate prices.

[15] The plotted results are a 3% random sample of the overall field-level results. These results were sampled randomly to reduce computational time. The sampled plot is nearly identical to the complete plot.

The moderate price response in the traditional corn states such as Nebraska, Iowa, Indiana, and Illinois is likely due to the fact that the most popular crop rotations in these states already include corn. Starting prices may be another explanation for high sensitivities in the North Central US. The basis patterns in figure 2.5 were generally consistent from year-to-year. Corn prices in the eastern Dakotas and western Minnesota tend to be lower relative to the rest of the country which may make growing corn in the region unprofitable without transient upward price fluctuations. If this is case, the region may respond more aggressively to an increase in prices than areas with consistently favorable prices.

Maps 2.8 and 2.9 also show that there is a degree of within-MLRA differences in price sensitivity. The MLRA on the border of North Dakota and Minnesota provides a clear example of within-MLRA heterogeneity by soil texture type. The Red River of the North provides the boundary for these two states. Soils in this area tend to contain more clay and the area is more susceptible to flooding in early springs. This could be a reason why corn plantings are less sensitive to price. Other researchers have noted that supply in this area of the country is especially sensitive to price changes, particularly among primary crop producers (Wang et al., 2017).

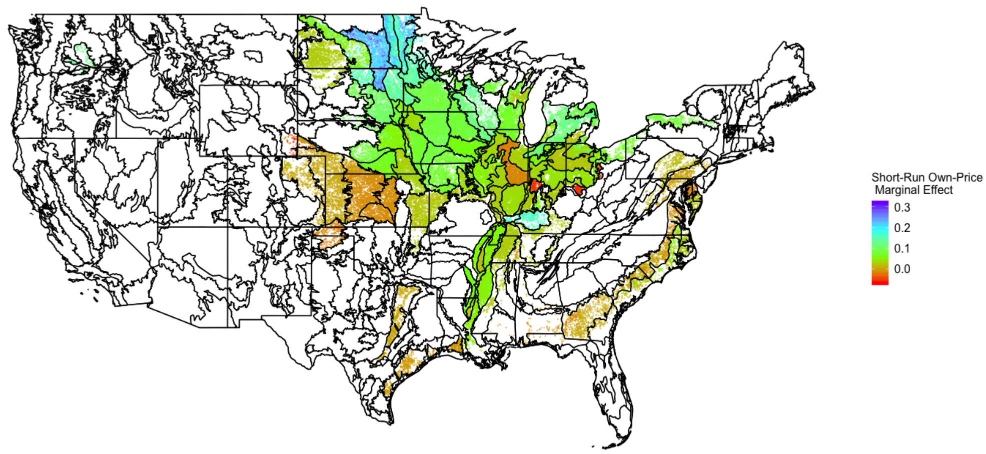

Figure 2.8: *Average MLRA-Group Short-Run Own-Price Marginal Effect over CONUS*

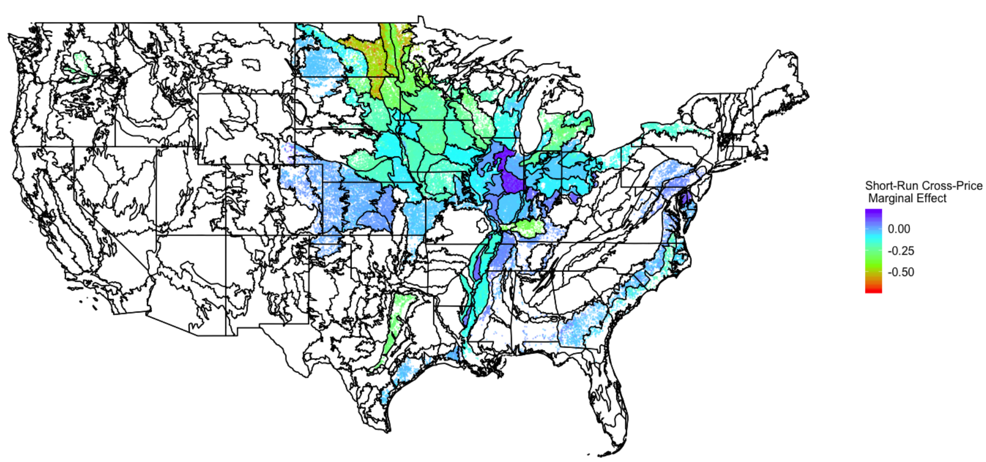

Figure 2.9: *Average MLRA-Group Short-Run Cross-Price Marginal Effect over CONUS*

2.5.4 Rotational Response

Since crop rotations describe the dynamic patterns of crop choices and have environmental benefits relative to continuously producing a single crop, price impacts on rotations are also important. Table 2.9 shows the distribution of average partial marginal effects affects across individuals within the MLRA groups. Continuous corn rotations were fairly rare among producers. On average fields had a 12% probability of planting to continuous corn. On average fields had a 30% probability of planting other crops continuously while other-corn rotations were dominant rotation as the average field had over a 50% likelihood of rotating between corn and the other crops.

Table 2.9 shows that rotational marginal effects generally had signs consistent with economic theory. Higher corn prices corn tended to increase the probability of planting continuous corn rotations and decrease the probability of continuous other rotations. The opposite was true for increasing the prices of other competing crops. Generally, prices did not impact the probability of performing a corn-other rotation as prices had a near-zero median marginal effect. The small effect that price changes had on rotating between crops is not surprising since on average half of farmers in a corn-other rotation plant corn and the other half plant some other crop. This means that a price fluctuation in a given year encourages half of the farmers that planted the other crop in the previous year to continue the rotation and discourages the other half that planted corn in the previous year to break the rotation. Unlike the state probabilities, the short-run rotational probabilities tended to be smaller in magnitude than their long-run counterparts. This is due to the fact that rotational probabilities are a multi-year concept. It would therefore make sense that a temporary fluctuation in prices would not have a large influence over multi-seasonal decisions of farmers relative to a more persistent price change.

I also include the elasticity estimates of the rotations in table 2.10. The results are similar to those by Hendricks et al. (2014b) and indicate that the rotations are quite sensitive to corn price fluctuations. On average fields the probability of planting continuous crop rotations were elastic to prices. Since mono-cropping is a detrimental to yields, and brings environmental concerns, these results indicate there could be significant environmental impact for policies that increase relative domestic price of corn.

Table 2.9: *Rotational Estimated Probabilities and Marginal Effects*

Statistic	Min	Q1	Median	Mean	Q3	Max
$\mathbb{P}^{CC}$	0	0.039	0.086	0.125	0.183	0.907
$\mathbb{P}^{OO}$	0.001	0.06	0.188	0.302	0.529	1
$\mathbb{P}^{OC}$	0	0.38	0.66	0.57	0.76	0.99
$\frac{\partial \mathbb{P}^{CC}}{\partial P^{C}}\|_{LR}$	-0.586***	0.009	0.024*	0.035***	0.058***	0.371*
	(0.109)	(0.008)	(0.013)	(0.013)	(0.018)	(0.214)
$\frac{\partial \mathbb{P}^{CC}}{\partial P^{C}}\|_{SR}$	-0.045*	0.007***	0.018***	0.024***	0.04***	0.125***
	(0.024)	(0.002)	(0.004)	(0.004)	(0.006)	(0.011)
$\frac{\partial \mathbb{P}^{OO}}{\partial P^{C}}\|_{LR}$	-0.81***	-0.068	-0.027	-0.053*	-0.014	0.73***
	(0.092)	(0.045)	(0.033)	(0.030)	(0.014)	(0.145)
$\frac{\partial \mathbb{P}^{OO}}{\partial P^{C}}\|_{SR}$	-0.321***	-0.04***	-0.021***	-0.033***	-0.01	0.077*
	(0.025)	(0.004)	(0.005)	(0.004)	(0.007)	(0.040)
$\frac{\partial \mathbb{P}^{OC}}{\partial P^{C}}\|_{LR}$	-0.082	-0.015	0	0.009	0.02	0.32*
	(0.100)	(0.026)	(0.023)	(0.021)	(0.016)	(0.179)
$\frac{\partial \mathbb{P}^{OC}}{\partial P^{C}}\|_{SR}$	-0.049**	-0.012***	0	0.004	0.014**	0.16***
	(0.025)	(0.003)	(0.004)	(0.004)	(0.006)	(0.025)
$\frac{\partial \mathbb{P}^{CC}}{\partial P^{O}}\|_{LR}$	-0.91***	-0.136***	-0.051***	-0.072***	-0.01	1.14***
	(0.109)	(0.008)	(0.013)	(0.013)	(0.018)	(0.214)
$\frac{\partial \mathbb{P}^{CC}}{\partial P^{O}}\|_{SR}$	-0.323***	-0.083***	-0.037***	-0.05***	-0.008	0.257***
	(0.028)	(0.015)	(0.009)	(0.010)	(0.005)	(0.087)
$\frac{\partial \mathbb{P}^{OO}}{\partial P^{O}}\|_{LR}$	-1.423***	0.013	0.052	0.108***	0.165***	1.846***
	(0.092)	(0.045)	(0.033)	(0.030)	(0.014)	(0.145)
$\frac{\partial \mathbb{P}^{OO}}{\partial P^{O}}\|_{SR}$	-0.15	0.01	0.038***	0.065***	0.088***	0.729***
	(0.095)	(0.018)	(0.014)	(0.012)	(0.012)	(0.077)
$\frac{\partial \mathbb{P}^{OC}}{\partial P^{O}}\|_{LR}$	-0.726***	-0.057**	-0.002	-0.018	0.04**	0.183
	(0.100)	(0.026)	(0.023)	(0.021)	(0.016)	(0.179)
$\frac{\partial \mathbb{P}^{OC}}{\partial P^{O}}\|_{SR}$	-0.364***	-0.038**	0	-0.008	0.034***	0.164**
	(0.061)	(0.016)	(0.011)	(0.011)	(0.008)	(0.082)

Note: Std. Errors in (), *** 1%, ** 5%, * 10% Sig. Lvl.

Table 2.10: *Rotational Price Elasticities*

Statistic	Min	Q1	Median	Mean	Q3	Max
$\frac{\partial \Pi^{CC}}{\partial P^C}\mid_{LR}$ Elas.	-0.904***	0.638	1.15*	1.534***	2.09***	6.47*
	(0.168)	(0.540)	(0.608)	(0.567)	(0.647)	(3.732)
$\frac{\partial \Pi^{CC}}{\partial P^C}\mid_{SR}$ Elas.	-0.685*	0.381***	0.818***	0.882***	1.361***	3.227***
	(0.370)	(0.099)	(0.164)	(0.130)	(0.190)	(0.271)
$\frac{\partial \Pi^{OO}}{\partial P^C}\mid_{LR}$ Elas.	-5.555***	-2.229	-1.093	-1.333*	-0.41	1.592***
	(0.629)	(1.468)	(1.321)	(0.751)	(0.415)	(0.315)
$\frac{\partial \Pi^{OO}}{\partial P^C}\mid_{SR}$ Elas.	-4.41***	-1.546***	-0.735***	-0.921***	-0.252	1.067*
	(0.345)	(0.138)	(0.162)	(0.115)	(0.161)	(0.559)
$\frac{\partial \Pi^{OC}}{\partial P^C}\mid_{LR}$ Elas.	-6.124***	-1.671***	-0.851***	-1.162***	-0.322	1.398***
	(0.734)	(0.097)	(0.210)	(0.208)	(0.579)	(0.262)
$\frac{\partial \Pi^{OC}}{\partial P^C}\mid_{SR}$ Elas.	-2.736***	-1.082***	-0.635***	-0.64***	-0.172	1.406***
	(0.235)	(0.189)	(0.149)	(0.126)	(0.111)	(0.476)
$\frac{\partial \Pi^{CC}}{\partial P^O}\mid_{LR}$ Elas.	-1.019	-0.179	0.04	0.29	0.365	4.403*
	(1.252)	(0.318)	(2.655)	(0.691)	(0.298)	(2.470)
$\frac{\partial \Pi^{CC}}{\partial P^O}\mid_{SR}$ Elas.	-0.746**	-0.138***	0.042	0.16	0.28**	2.332***
	(0.376)	(0.031)	(0.734)	(0.137)	(0.121)	(0.371)
$\frac{\partial \Pi^{OO}}{\partial P^O}\mid_{LR}$ Elas.	-2.303***	0.118	0.676	0.941***	1.637***	5.432***
	(0.148)	(0.410)	(0.426)	(0.260)	(0.139)	(0.426)
$\frac{\partial \Pi^{OO}}{\partial P^O}\mid_{SR}$ Elas.	-2.069	0.067	0.471***	0.628***	1.111***	2.836***
	(1.310)	(0.122)	(0.173)	(0.114)	(0.150)	(0.298)
$\frac{\partial \Pi^{OC}}{\partial P^O}\mid_{LR}$ Elas.	-4.23***	-0.343**	-0.033	-0.251	0.168**	1.217
	(0.585)	(0.158)	(0.425)	(0.301)	(0.068)	(1.192)
$\frac{\partial \Pi^{OC}}{\partial P^O}\mid_{SR}$ Elas.	-2.242***	-0.285**	-0.037	-0.137	0.119***	0.949**
	(0.379)	(0.123)	(2.145)	(0.191)	(0.030)	(0.474)

Note: Std. Errors in (), *** 1%, ** 5%, * 10% Sig. Lvl.

2.5.5 Comparing Heterogeneous and Pooled Models

Preventing bias is one of the primary motivations for estimating the model using heterogeneous coefficients. To illustrate the benefits of incorporating heterogeneity, I estimate a second model where each Markov transition equation is estimated once over the entire pooled national sample. The pooled model is nearly identical to the earlier model that incorporates heterogeneity through MLRAs and soil-texture groups. Under both models, the price variables are identical at the observation level. That is, I reuse the "other" price index from the earlier framework at the observation-level. Since the other price index was constructed using MLRA-level production baskets, I include MLRA fixed-effects in the pooled model by adding MLRA-specific dummy variables. Without the MLRA fixed-effects, the model produced results that were inconsistent with economic theory (e.g. increasing corn (other) prices decreased (increased) corn acreage). While the heterogeneous modeling framework estimates distinct coefficients over subsets of the national sample, the pooled model does not. Because of the large sample, computational problems arose when estimating over the entire national sample. For the sake of tractability, I estimate the pooled model using a 10% random sample of the total national sample. This 10% subsample had 3,212,765 observations.

In the final results section I compare corn quantity and acreage elasticity estimates using the national sample shown in table 2.11 and a subsample of the Corn Belt states of Iowa, Illinois, Indiana, and Nebraska shown in table 2.12. In each table, acreage and quantity elasticities are estimated using two different models. In the first model, elasticity estimates are constructed using the aggregated results from the heterogeneous regional elasticity estimates. The second set of elasticities were constructed using a single pooled logit model over the entire sample. Both estimates were constructed as a summation over field-level data and therefore the only differences between these estimates arise from the heterogeneous coefficients in the aggregated model. These tables help illustrate the benefits of modeling supply response with heterogeneous response, and how the response from even important production areas can differ from national estimates.

Result 1: At the national-level, acreage elasticities tend to be larger than quantity elasticities.

The first result is that acreage tended to be more sensitive to prices than quantity. This finding is consistent with an underlying assumption in Ricardian rent theory which posits that the marginal product of the lowest quality land used in production will be at or just above land rental rates. When land is fixed in supply but heterogeneous in productivity, as prices increase, less productive land can and will be profitably brought into production (Barlowe, 1972; Ricardo, 1891). These findings are also consistent with Lark et al. (2015) who found that the expansion of new acres tended to come from marginal land that was less suitable for cultivation. Under the Corn-Belt-only estimates, the acreage and quantity weighted elasticities are much closer together. This is likely due to more behavioral and yield homogeneity at the regional-level relative to the national-level. With less yield variation, the covariance between yield and planting decisions is likely to reduce. This can be shown using the definition of correlation in equation 2.31. If the correlation between two random variables Y and Z remains fixed, reductions in the variance either Y or Z reduce the covariance. In other words, the covariance between planting and yields potential will likely go down when yields are more homogeneous.

$$Cov\left[Y,Z\right]=Corr\left[Y,Z\right]\sqrt{Var\left[Y\right]}\sqrt{Var\left[Z\right]} \tag{2.31}$$

Result 2: Nationwide heterogeneous elasticities are larger than the pooled elasticities especially in the short-run.

The second result relates to differences between heterogeneous elasticities and pooled elasticities. At the national-level, the estimated elasticities that allow for heterogeneous coefficients tended to be larger than the pooled estimates. In the short-run, these differences were substantial, ranging from 7% to 11%. These differences were present but less pronounced in the long-run elasticities which only differed by around 3%. However, this is does not indicate that a lack of in the long-run estimates of the bias from the pooled model. The delta values in the tables are akin to the coefficient on a lagged dependent variable in a dynamic model. In this view, while the pooled model underestimates the short-run elasticity, it also underestimates the size of the lagged dependent coefficient which just happens to bring the long-run elasticities between the models closer together. This suggests that the pooled approach generally underestimates acreage and quantity response and that there is merit to incorporating heterogeneous coefficients. Under the national dataset, the pooled model tends to underestimate supply response but in the corn-belt sample, the pooled model overestimates supply response. The likely reason for this is the pooled model averages the impact of price changes over the entire national sample. Averaging over the pooled model reduces the influence from areas with larger supply-response and creates discrepancies between the heterogeneous aggregate and the pooled estimates.

Result 3: The difference between short run and long run effect is more pronounced in the Corn Belt than in nationwide estimates.

Tables 2.11 and 2.12, include "delta" statistics on the elasticities. These terms are the percent difference in the LR elasticity relative to the short-run elasticities and are analogous to the coefficient on the lagged dependent in a typical dynamic model (Hendricks et al., 2014c). Generally speaking, these delta terms indicate whether fields tend to continuously grow a single crop crop (positive delta) or transition from one crop to another (negative delta). From the functional form of the long-run marginal effect in equation 2.12, this will be proportional to the difference between the long and short -run marginal effects shown in equation 2.32.

$$\left[y_{it}^{CC} - y_{it}^{OC}\right] \frac{\left[1 - y_{it}^{CC}\right] \frac{\partial y^{OC}}{\partial P^k} + y_{it}^{OC} \frac{\partial y^{CC}}{\partial P^k}}{\left[1 - y_{it}^{CC} + y_{it}^{OC}\right]^2} \tag{2.32}$$

Relative to the rest of the country, the long-run marginal effect was much smaller than the short-run effect inside the Corn Belt. Provided that the coefficient values take on the same value as the marginal effect, long-run effect will be smaller than the short-run effect when the first bracketed term is negative. This occurs when fields are more likely to transition to corn when some other crop is planted in the previous period. While always positive, the denominator in the ratio term is conditional probability that a field will transition to a different crop between one period and the next. The denominator will be smaller when likely to consistently host the same crop from year to year will have a smaller denominator. Subsequently the long run price effect on these fields will be smaller than the short-run. The numerator is a measure of the effect that prices have after two or more periods. The $\left(1 - y^{CC}\right)$ term is the probability that a field planted to corn will switch to the other crop. If such a transition is made, marginal effects for fields that planted the other crop in the previous period will be relevant in predicting future crop choices. Similarly, the y^{OC} term is the probability that a field is planted to corn given some other crop was planted previously. If a field indeed moves from the other crop to corn then the marginal effect of prices conditional on corn being planted will be relevant for predicting the next crop choice.

The values in table 2.13 show the conditional and sequential probabilities in and outside of the Corn Belt. The conditional probability values in this table show that the consistent rotational patterns within the Corn Belt are likely the reason that the long-run effect is smaller relative to the the short-run effect. Constructing the bracketed term on the left of equation 2.32, from the conditional probabilities in the table shows that the term was nearly nearly nine times higher in the Corn Belt. The larger differences between the long and short-run elasticities within the Corn Belt relative to the national level is likely attributable to the popularity of the corn-soybean rotation within the Corn Belt. This could also be a result of the more established corn production within the Corn Belt relative to outside areas and that permanent price increases could lead to permanent expansions into new corn growing areas. High prices in the Corn Belt may persuade farmers to make a one-year short deviation in their rotations but, as table 2.13 shows, farmers that plant some other crop in the previous period were considerably more likely to plant corn in the subsequent period than farmers outside the Corn Belt.

Result 4: Nationwide elasticities were much larger than the Corn-Belt elasticities.

The nationwide heterogeneous aggregate elasticity estimates in table 2.11 were nearly twice the size of the Corn Belt heterogeneous elasticities in table 2.12. This that the corn supply in the Corn Belt is much less sensitive to price changes than the supply from country as a whole. Corn is generally better suited to the traditional Corn Belt states making it relatively profitable to produce each year. The conceptual model in (Wang et al., 2015) suggests that supply response will be larger in areas where the probability of a binary planting decision is close to 50%. If this is the case, under a binary choice framework, the probability of making one choice or the other are relatively close. Therefore, any change in the relative profitability of between the choices will influence planting decisions more than otherwise. In the case of the Corn Belt, the corn-soybean rotation extremely popular sequential crop choice. Table 2.13 shows conditional and sequential probabilities for the different pairs of planting decisions. Approximately 75% of the Corn Belt sample was in a other-corn rotation. From this table, I can infer that the probability of a producer deciding on a continuous corn rotation is 15.3% and the probability that a producer chooses a continuous other crop rotation is 9.7%. Adding the remaining sequential probabilities gives the probability that a farmer rotates crops each year. This implies that producers in the Corn Belt have a 75% probability. Produces inside he Corn Belt,that planted an "other" crop in the previous season had an 80% probability of planting corn. Outside of the Corn Belt these producers would have a 37% probability of planting corn.

This means that conditional on the crop choice in the previous year was not corn, farmers in the Corn Belt are very likely to plant corn in the subsequent period. The high degree of correlation in the marginal effects suggest that rotations make up a high degree of supply response decisions and that soybeans-to-corn transitions likely make up an important component in the state probability of planting corn. In this way it is not surprising that farmers in the Corn Belt are less sensitive to prices. Those that were in the soybean stage of their rotations in the previous year are already more predisposed to planting corn in the first place. If this were the case, then I would expect that the supply response in similarly productive areas outside of the Corn Belt would be more sensitive to prices.

Table 2.11: *Nationwide Corn Elasticities With and Without Heterogeneity*

Weighting →	Acreage		Quantity	
Statistic ↓ Model →	Het. Aggr	Pooled	Het. Aggr	Pooled
Own-Price SR Elasticity	0.674**	0.568	0.644***	0.545
	(0.266)	(–)	(0.094)	(–)
Own-Price LR Elasticity	0.526**	0.502	0.496	0.468
	(0.219)	(–)	(0.082)	(–)
Own-Price Delta	-0.281	-0.133	-0.299	-0.165
Cross-Price SR Elasticity	-0.496*	-0.413	-0.47***	-0.396
	(0.266)	(–)	(0.0346)	(–)
Cross-Price LR Elasticity	-0.398*	-0.365	-0.370	-0.340
	(0.231)	(–)	(0.219)	(–)
Cross-Price Delta	-0.246	-0.132	-0.269	-0.165

Note: Std. Errors in (), *** 1%, ** 5%, * 10% Sig. Lvl.

Table 2.12: *Corn-Belt Elasticities With and Without Heterogeneity (IA, IL, IN, NE)*

Weighting →	Acreage		Quantity	
Statistic ↓ Model →	Het. Aggr	Pooled	Het. Aggr	Pooled
Own-Price SR Elasticity	0.390*	0.437	0.391***	0.432
	(0.234)	(–)	(0.0835)	(–)
Own-Price LR Elasticity	0.259	0.313	0.260***	0.306
	(0.235)	(–)	(0.0839)	(–)
Own-Price Delta	-0.505	-0.396	-0.505	-0.411
Cross-Price SR Elasticity	-0.252***	-0.322	-0.253***	-0.318
	(0.0274)	(–)	(0.0274)	()
Cross-Price LR Elasticity	-0.169	-0.231	-0.170	-0.226
	(0.166)	(–)	(0.166)	(–)
Cross-Price Delta	-0.489	-0.396	-0.488	-0.410

Note: Std. Errors in (), *** 1%, ** 5%, * 10% Sig. Lvl.

Table 2.13: *Conditional and Sequential Probabilities In and Outside of the Corn Belt*

Region	Probability	Crop Choices			
		Corn→Corn	Other→Corn	Corn→Other	Other→Other
Outside Corn Belt	Sequential	0.111	0.241	0.241	0.407
	Conditional	0.315	0.372	0.685	0.628
Corn Belt	Sequential	0.153	0.375	0.375	0.097
	Conditional	0.29	0.795	0.71	0.205

2.6 Conclusions

Understanding nationwide crop supply response statistics are useful in a variety of areas. In this paper, I assumed that responds to price exclusively through crop choice. I estimate the effect that prices have on the corn acreage and quantity by multiplying field acres and county yields by the marginal effect of prices on the probability of planting corn. The relationship between crop choice and prices were modeled separately across 122 subregions of the country using a set of conditional Markov chain transition regressions. The data in the study sets itself apart from the literature by including crop choices at the field-level that account for over 80% of the land devoted to corn production in the US between 2009 and 2016. The results show that modeling supply response without accounting for heterogeneity can lead to bias in national elasticity estimates, particularly in the short-run. They also show that regional supply, even in the Corn Belt tended to be less sensitive than the national estimates. This cautions against extrapolating such estimates to the nation or to other regions. Regional corn supply response was found to be highly heterogeneous with the Northern Plains states and the Mississippi showing especially high sensitivity and western Kansas and the Gulf states exhibiting low sensitivity.

Heterogeneous supply response is useful in a variety of applications. For instance, since river systems can carry fertilizer contaminants across vast distances, the impact of runoff externalities are worse near the mouthes of major river systems (Wu et al., 2004b). Understanding where supply is especially responsive to prices can help policymakers better understand the secondary impacts of policies distorting domestic prices. Identifying areas where corn planting decisions respond to prices can also provide a richer analysis of the supply response of other crops. If corn supply response displaces the production in areas that specialize in other crops, researchers can provide a richer explanation of cross-price elasticities. For example, the corn supply in the upper Mississippi River Delta was highly sensitive to prices. This is an important production area and helps explain why some in the literature find a relatively large cross-price cotton supply elasticity on corn prices (Vorotnikova et al., 2014). Lastly, modeling at the field-level and allowing for heterogeneous response helps address aggregation bias when estimating nationwide supply response which are also used for a variety of purposes across agricultural economics.

There are a variety of extensions that can be made in the area of heterogeneous supply response. While the models attempt to embed heterogeneity in the production decisions, it is not clear at what point the relevant heterogeneity is fully captured. The work of Athey et al. (2016) provide promising, data-driven modeling options for incorporating latent heterogeneity over large datasets. Further analysis can be performed to establish the cause of the heterogeneous supply response. In particular, the results from this study could enter a second stage regression as a dependent variable with regional farmer characteristics as independent variables.

Chapter 3

Estimating the Local Impact of Ethanol Plants on Cropland Transitions

3.1 Introduction

The first chapter of this dissertation focused on how crop choice varies with prices, in this chapter, I shift the focus of supply response to the extensive margin and study how broader land usage changes with the intervention of new local ethanol demand. The 2007 Renewable Fuel Standard (RFS) had a dramatic impact on row crop agriculture within the United States. Roberts and Schlenker (2013) estimate that a third of corn production goes into ethanol production. Along with a precipitous rise in corn dedicated to ethanol production, the RFS also induced widespread investment in ethanol processing plants. To address the concerns of cropland extensification from environmental groups, policymakers included a stipulation in the policy to lessen its impact. This stipulation forbids land converted into farmland after

2007 from being used to produce ethanol (EPA, 2010). In this chapter I assess the local impacts of introducing ethanol plants on land use.

The RFS mandate has significantly impacted national and international markets across several primary crops. Over a third of corn in the US was diverted to producing ethanol and increased major commodity prices by 30% (Carter et al., 2016). As a major exporter of corn, the RFS redirected a third of US corn production and 5% of the world's caloric production to ethanol(Roberts and Schlenker, 2013). By substantially increasing the demand for ethanol, the RFS caused a dramatic expansion of US ethanol production, more than doubling production capacity within a few years after its passage. Since producers use land to produce profits, the plausible influence that ethanol plants have on land choice is channeled through its impact on crop prices. The introduction of ethanol plants can substantially change the local basis in an area by increasing local prices in the area surrounding the plants (McNew and Griffith, 2005). The expansion of ethanol capacity in the wake of the passage of the RFS of 2007 provides an ideal scenario to study the local impacts of ethanol plant on cropland conversion and retention. Since the RFS also produced a general price increase, I seek to isolate the local impact from ethanol plants using a set of difference in differences (DID) approaches.

Utilizing corn ethanol as a source of transportation fuel dates back to the 1920s. Through a series of increasing mandates, the Renewable Fuel Standard, (RFS) passed in 2007, was a major contributor to a surge in US ethanol plant investment. Starting from less than 6 billion gallons of ethanol, the RFS progressively mandated the use of over 36 billion gallons of biofuel for transportation purposes by the year 2022. Subsequently the total US production capacity boomed after 2008 with the surge of new investment. Figure 3.1 shows the total US ethanol plant capacity over time. Further fluctuations in capacity occurred after 2013 when policymakers, faced with mandate compliance issues due to economic and technological constraints, began relaxing the mandate schedule (Lade et al., 2018).

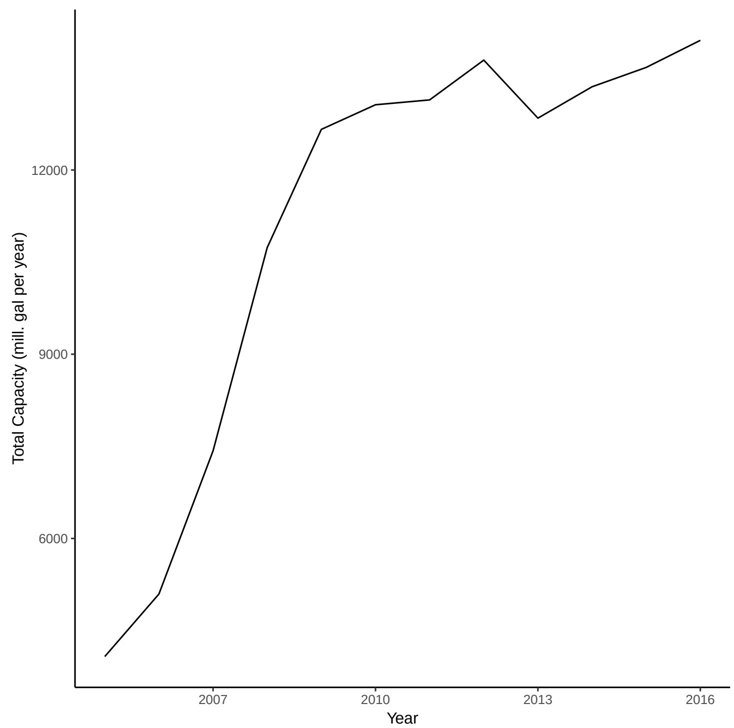

Figure 3.1: *US Ethanol Plant Capacity (Nebraska Energy Office, 2016)*

Cropland conversion attributed to the RFS as a whole has been a popular subject of research. Yearly cropland-to-grassland transitions are relatively less frequent and therefore even small changes in conversion incentives can have a large relative impact on the conversion rate (Carriazo et al., 2010). In an influential article, Searchinger et al. (2008) found that through direct price effects and indirect effects on global trade, the RFS caused a 10.8 million hectare expansion of cropland globally with 2.2 million hectare expansion from the US alone (Searchinger et al., 2008). To study the impact that the RFS had on nitrogen leaching into the Gulf of Mexico, Donner and Kucharik (2008) simulate the expansions cropland that would likely be made to comply with the 15 billion gallon ethanol production target for 2022 from the RFS. While they estimated that no expansion of corn land would be needed if yields improve at their current rate, shortfalls in yield improvements could lead to a 9% expansion in corn growing area between 2008 and 2022 to meed the mandate requirements Donner and Kucharik (2008). Using estimates derived from the CDL Wright and Wimberly (2013) found that corn and soybean area displaced 1.3 million acres of grassland in the Dakotas, Minnesota, Iowa, and Nebraska between 2006 and 2011.

When aggregated across the US, the long-run composition of land use has been remarkably stable. Between 1910 and 2004, overall cropland has only increased by around 3% (Lubowski et al., 2006; Ramankutty and Foley, 1999). However, there has and continues to be dynamic changes at the regional level with areas like Midwest gradually making up larger shares of the national cropland over time. Transitions between uses is also common. Over 15% of the cropland in 1997 was CRP or some other use in 1982. Transitioning land from non-crop uses such as hay production and pasture land to cropland tends to be detrimental to the environment. Lands at the extensive margin tend to be less productive as cropland than permanent cropland (Huang et al., 2010; Lubowski et al., 2006). These lands tend to be less productive and more difficult to work due to lower quality soils and higher slopes. Since land at the extensive margin generally requires more fertilizer to profitably raise crops, farming on newly converted cropland tends to have a greater environmental impact than long-established cropland (Lubowski et al., 2006).

While many have studied the impact the RFS has had on crop conversion as a whole, less attention has been given to studying the localized impacts of ethanol plant expansion. While the RFS produced general impacts in cropland conversion, the associated rise in ethanol capacity due to the RFS provides a useful dataset to estimating the impact of ethanol plant construction on cropland conversion and retention. In this study I use a series of difference-in-differences models to estimate these local impacts across a thirteen-state corn-growing region of the corn-growing United States.

Using current sources of remote sensed data to estimate broader land transitions is difficult. In the previous chapter, I used the Cropland Data Layer (CDL) to estimate the response of crop choices to changing prices. While the CDL is generally useful for determining land cover, it is less appropriate for identifying land use. For example, accurately classifies the land cover for major commodity crops at a rate of 90% or higher. However, due to methodological changes to the CDL, the base data layer has a tendency to over-estimate cropland expansions (Lark et al., 2017). To address this, I utilize data from Lark et al. (2015) which was specially designed to address potential cropland transition and classification bias in the CDL.

To cover the general methodology and findings of land use studies on ethanol plants, I'll discuss a few articles and then discuss my methodology for this chapter. With an influx of new corn demand for ethanol in local markets, many have studied the impact that the increases in local demand brought about by added ethanol capacity. There are several challenges to these types of studies. The first is an endogeneity issue, ethanol plant locations are endogenous to the supply of corn in the area. Much of the methodological effort in studies on the impact of ethanol supply has been to address the endogeneity of plant location. Motamed et al. (2016) used a instrumental variable approach to account for this. They noted that ethanol plants tend to locate in areas close to railroad lines since ethanol is often transported via rail. They argue that since many of these lines have been established since the 1970s and that grain transport from field is normally by truck and not rail. Therefore the only way that railway line proximity to the field influences contemporary planting decisions is through the ethanol plant proximity. They found that capacity changes had a significant and positive effect corn planting and agricultural acreage. Specifically found that a 1% increase in ethanol capacity within an area causes a 1.5% increase in corn acres within the area and a 1.7% increase in agricultural acres and found that these effects were larger in areas with less corn Motamed et al. (2016).

The majority of similar studies use some sort of a difference in differences (DID) design. To construct treatment and the control groups that are comparable, most studies either used some form of propensity score matching or fixed effects regressions. Generally studies relying on remote sensing data used the propensity score matching approach. Arora et al. (2016) studied how land-use decisions of farmers in the Dakotas were impacted by the construction of 15 individual ethanol plants in the area between 2005 and 2013. They used a combination of difference-in-differences (DID) framework with a psropensity score matching framework. In their analysis, they specified a control group and treatment groups by isolating areas closer to one of 15 ethanol plants (for the treatment group) and further away (for the control group) from the plants. They then estimated propensity scores over the treated and control group to estimate the probability that a control group observation would be in the treatment group. A major contribution of this paper was they used a flexible DID structure that relaxes the parallel trend assumption. While their DID step was sophisticated, their propensity score matching step was very simple. Matching used the quadratic form of only two variables, the Land Capability Classification (LCC) codes and slope. They then select for their treated and untreated samples using a one-to-one nearest neighbor algorithm with the added restriction that the observations used in the analysis had to a untreated match within a given radius. To perform the propensity score matching, they used a logit model with quadratic weighted land slopes and quadratic weighted LCC as predictors. Their results were mixed, indicating a positive treatment effect for some ethanol plants and a negative treatment effect for others (Arora et al., 2016). However, because their propensity score matching fit was so simple, bias in the matching process could be impacting their results (Smith and Todd, 2005).

There has been evidence for a positive impact of ethanol plants on regional extensive land changes. Brown et al. (2014) studied the effects that plants had on corn extensification and intensification in the state of Kansas. Using the Cropland Data layer, a county-level control for alternative corn demand from the cattle industry within the state, and spatial lags and errors to account for potential spillover effects from overlapping markets. They found that the proximity to ethanol plants positively impacted the likelihood of planting corn land extensification (Brown et al., 2014).

Ifft et al. (2018) studied the impact that county-level ethanol plant capacity had on the re-enrollment of expiring CRP land. They also used a DID strategy using non-ethanol producing counties as the control group in conjunction with fixed effects to ensure that county and time effects were stable across the treatment and control groups. Contrary to expectations, they found that CRP re-enrollment increased in ethanol producing counties relative the control group after the RFS and cite concurrent changes in the CRP program as the reason (Ifft et al., 2018).

I'll lastly discuss Towe and Tra (2012) who also used the DID approach. In their study, they estimated the impact ethanol capacity changes had on the farmland values for over 50 ethanol plants. They classified parcels as "close" to an ethanol plant as ones that were within 30 miles of a plant. Their method is similar to the approach of Arora et al. (2016) but they utilize a considerable number of controls during the propensity score matching step. In addition, they use a difference-in-difference-in-differences (DDD) approach, testing the difference between the treatment effects from areas that experience ethanol expansion before and after the 2005 RFS mandate. They found that there was a statistically significant treatment effect in the post 2005 RFS mandate but not before. Because both treated and untreated parcels were exposed to higher prices in the post RFS time period, they concluded that there was a land capitalization premium for being near ethanol plants when demand was high.

In this paper, I employ a method similar to Towe and Tra (2012), a DID approach with propensity score matching. Like Towe and Tra (2012), this study uses the decisions around multiple plants to estimate the a single treatment effect, and also emphasizes plant capacity as opposed to simply the construction of the plant itself. Like Towe and Tra (2012), I also found the vast majority of ethanol plants the dataset do not add capacity once they are built. That is, most of the increase in total ethanol capacity comes from the construction of new plants being built in 2007 and coming online in 2009. Unlike Towe and Tra (2012) I want to determine the impact of new ethanol going into the production regions and not the passage of RFS legislation. I therefore use changes in plant capacity and their corresponding construction years to delineate treatment and control groups over time. Generally construction on ethanol plants takes around two years, I therefore restrict control groups during the matching procedure in the propensity score matched DID approach and compare it with the traditional linear DID approach. This study expands on the dataset of Towe and Tra (2012) and refines the treatment assignment by considering multiple treatment years and using the capacity of the plant to determine the area being treated by new plant construction. Many of the controls used in Towe and Tra (2012) are already collected or are readily available and also apply to cropland conversion including the NCCPI, crop price basis, weather controls from PRISM, and irrigation information.

3.2 Data

This study covers a 13-state region in the north-central United States outlined in figure 3.2. This area was chosen for several reasons. The first is that the majority of the country's ethanol plants are located in the north-central US. Secondly, cropland transitions differ across the US. Crop transitions in this region of the country tended to be more homogeneous and are characterized by transitions between to cropland and uncultivated usage such as pasture and hay production as opposed to transitions to and from forestry land that is more popular in the south (Lubowski et al., 2006).

Theory demands that all variables that plausibly impact either the treatment assignment or changes in the outcome variables should be used to match treatment and control observations. Generally speaking, matching methods benefit the analysis by eliminating two sources of bias. The first source of bias is some treated and control individuals may not be comparable. That is to say, they may not have a common support over observable, pre-treatment variables. Another source of bias that matching methods help eliminate is a non-uniformity of the distribution of relevant time invariant variables. The process of matching explicitly connects treated and control individuals together by these observable features so that the matched groups are com-

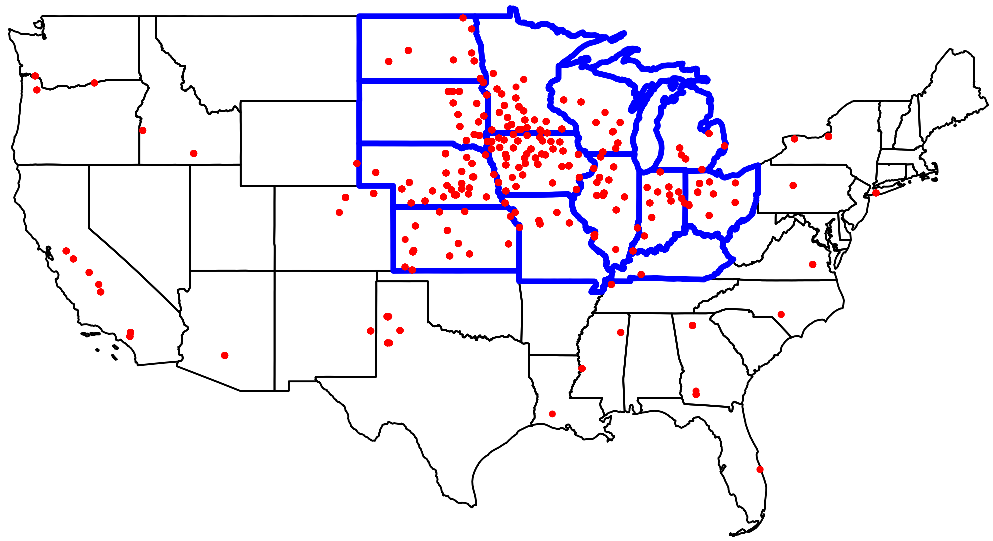

Figure 3.2: *Region of Interest and Ethanol Locations*

parable with one another. Failure to include relevant controls leads to poor matching and consequentially hinders the ability of matching to eliminate this bias (Heckman et al., 1997). It is therefore crucial that these controls are properly justified as relevant in treatment assignment or the evolution of cropland transitions.

3.2.1 Cropland Transitions

The dependent variables in this study are the conditional land-use transition probabilities that were constructed from a long-term cropland conversion dataset. These probabilities are approximated using a field-level raster dataset (Lark et al., 2015). The raster cell that intersects with the centroid of each field is used to estimate the broad land use. The dataset consists of a set of binary variables and accompanying observation years. This binary dataset contains four codes indicating: (1) the field was continuously cropped from 2008-2016, (2) stayed as non-cropland from 2008-2016, (3) transitioned between non-cropland to cropland, and (4) transitioned from cropland to non-cropland. Since the outcome variables consist of a set of binary variables, the model will estimate the probability of a field transitioning from one broad land use to another. By restricting the dataset into fields that were either always cropland or ones that converted out of cropland, the conditional probability of cropland retainment can be modeled. Likewise, by restricting the dataset to fields that were either always in non-cropland or transitioned to cropland, the conditional probability of cropland conversion can be modeled.

Lark et al. (2015) constructed this dataset using data from the Cropland Data Layer (CDL). Using remote sensed data to describe long sequential transitions is difficult since each observation is measured with some potential error which compounds the uncertainty of the estimate. First, to improve classification of the CDL, they divided the CDL land coverage indicators into crop and non-crop uses. Since idled or fallow lands were generally temporarily out of crop production, they were considered crops. These binary data layers were then "stacked" onto one another, assigning each cell a sequence of crop status indicators over each year. These sequences were called "trajectories". An example of a trajectory for a field in cropland in the final two years of a 6 year interval converted would be represented in binary is "000011" where zero indicates non-cropland and one indicates cropland. To improve the quality of these trajectory estimates, they applied spatial and temporal filters.

They then sorted these trajectories into the categories: (1) *No-Change*, where the land use was consistent over the entire study period (e.g. “000000” or “111111”), (2) *Change*, where a transition was made but remained consistent for the remainder of the study period, (e.g. “001111” or “111100”) (3) *Noise* where a single transitory change was made (e.g. “001000” or “101111”), and (4) *Flip-Flop* where a change was made consistently for at least two years and then reverted back to the original status (e.g. “011010” or “001100”). Flip flops were considered misclassified and removed from the analysis altogether. Noise trajectories were considered “No Change” trajectories with a measurement error and reassigned so that a “001000” noise trajectory was reclassified to a “000000” no change trajectory and a “101111” was reclassified as a “111111” no change trajectory.[1] Lastly, fields with less than 15 acres were removed since fields with smaller area are especially prone to measurement error.

The decisions to move land into of cropland is of interest in this study. I collectively refer to these decisions as cropland transitions. Fields can either transition from cropland to cropland or *retain* cropland, or transition from some non-crop use to cropland or *convert* to cropland. These are represented as two binary variables. The first measure I call *cropland retention* $(C \mid C)_{it}$. This is equal to one if field (i) was in cropped in the year t given it was cropped $t-1$. The second is called *cropland conversion* $(C \mid N)$ this is a binary variable equal to one if the field were cropped in year t given it was not in cropland in year $t-1$. Like the previous chapter, these transitions

[1] A converted field would not convert back to the original status.

can be thought of as elements in a Markov transition matrix.

By defining the dependent variables in a similar way to Chapter 1, the models estimate the change in the probability that land is being retained as cropland or transitioning to cropland from the expansion of an ethanol plant. One dependent variable is defined as the conditional probability that a given piece of land transitions into cropland given cropland was observed in the previous year $\left(\mathbb{P}^{C|C}\right)$. The second dependent variable is defined as the conditional probability that a given piece of land is converted into cropland given that it was observed as non-cropland in the previous period $\left(\mathbb{P}^{C|N}\right)$.

3.2.2 Matching Variables

Soil Controls

Many of the variables relevant to crop choices are also relevant to broader land use choices. For this reason many of the variables used in the previous chapter are utilized here. Soils characteristics are an important time-invariant feature of fields. Land that transitions from cultivated to non-cultivated cropland tends to be lower quality marginal land Lubowski et al. (2006); Wright and Wimberly (2013). Therefore, the slope of the field, soil texture, the NCCPI soil productivity index, and irrigation status likely correlate with transition benefits. There are soil features that are also likely to determine the relative profitability of broad land uses. In the SSURGO data, soils defined as hydric are soils that "are sufficiently wet in the upper part to develop anaerobic conditions during the growing season" (NRCS, 2018). Conceptually these are soil that are likely to retain excessive moisture in the topsoil and hamper plant development. The final soil variable I include as a matching variable is wind erodability. Wind and water erodability is inversely related to crop yields in general and a criteria for acceptance in the Conservation Reserve Program (CRP), a land retirement program. Slope is a field characteristic that is highly related to water erosion and fields with steeper slopes are generally more difficult to farm than fields with gentler grades.

Climate Controls

Some of the controls in the previous chapter are not appropriate for cropland conversion. Cropland conversion as it is defined in the dataset is a longer-term production decision and therefore transitory weather shocks aren't likely to impact these decisions. This is not to say that weather plays no impact however. The climate of an area may impact broader land use by restricting the suitable set of crops of an area. To control for climate I construct 30-year growing-season mean and variance estimates for growing-season extreme degree days, growing degree days, and precipitation. I assume a growing season to be between April 1st and September 30th for each year. Extreme degree days are a measure of exposure extreme temperatures at or above $30°C$ and growing degree days are a measure of exposure to suitable heat units for most crop throughout the course of a season. For each grid PRISM cell, observations on degree days and precipitation between 1984 and 2004 were used to construct mean and standard deviation statistics. These climate statistics were then applied to each field whose centroid lies inside each cell.

Price Controls

Crop prices can impact the relative profitability between crop and non-crop uses and therefore need to be controlled for. The introduction of ethanol plants have been shown to increase the basis within an area around them and would therefore using yearly prices would make it difficult to distinguish the effect of the introduction of a new plant (McNew and Griffith, 2005). However, the basis patterns before plants enter the market may be of interest. In the first chapter, I found that crop producers in areas that initially have exceptionally low prices were more receptive to price increases and the same could be true of local land conversion with ethanol demand. To control for relative local prices I include the 2004 field-level expected corn prices from the previous chapter. Since these prices correspond to a single crop year, variation it will only vary cross-sectionally by the basis pattern. Basis patterns may also help control for treatment assignment. Crop prices are input prices to ethanol plants and therefore areas with lower prices would be advantageous to an ethanol plant ceteris paribus.

Location Controls

Proximity to major population centers may also impact land use choices. The proximity to major cities potentially impacts both the likelihood of treatment and the land use choice itself. Fields closer to larger urbanized areas likely have different competing land uses than those farther away from major urban centers. Recreational and agri-tourism uses for farmland are more feasible for farmland closer to major urban centers. Farmland may be more likely to become urbanized itself as since developed land generally produce higher returns relative to cropland. The proximity to major urban centers could also be an indicator of transportation costs to final intermediate or final consumers for potential crop producers and ethanol plants. Ethanol is generally too corrosive to be transported by pipe, therefore plants generally utilize rail, barge or over-the-road trucking all of which are likely to be more developed closer to urban markets. Transportation costs differ by land-use, land that is in pasture or in uncultivated crops such as hay have little to no transportation costs while the transportation costs for raising crops and logging are relatively significant. To account for this I include a dummy variable indicating whether a field is within 30 miles of a major urban city defined as areas a population of 100,000 or more people.

3.2.3 Ethanol Plant Location and Capacity

In similar studies analyzing the impact of ethanol plant construction observations are generally assigned into treatment and control groups using a simple distance from the plant itself. While studies often consider several radii, a common radius are considered across every plant in the study (Arora et al., 2016; Ifft et al., 2018; Motamed et al., 2016; Towe and Tra, 2012). In this paper, I incorporate the impact of plant capacity in the treatment assignments. To assign observations into treatment and control groups I use Ethanol Production Capacity by Plant archived dataset provided by the Nebraska Department of Energy. This dataset provides a listing of over 200 ethanol plants and tracks each plant's the production capacity in millions of gallons per year by month. To estimate the treatment effect of the ethanol plant, I assigned fields to a control or treatment group based their proximity to the plants. To construct neighborhoods around the plants I first geocoded each plant using the centroid of the city in the plant's address.

The Nebraska Energy Office data tracked the capacity of US ethanol plants on a monthly basis. From 2005 to 2016 there were 497 ethanol plants in the dataset. To define my treatment and control groups, I use the field's Euclidean distance from the plant to delineate treatment and control groups. This study examines the impact of of ethanol capacity changes in a total of 217 market cities with annual nameplate ethanol plant capacity from 2008 to 2016. The literature on the effect that an ethanol plant has on local basis patterns is rather weak. McNew and Griffith (2005) found that the presence of an ethanol plant strengthened local basis but the capacity of of a plant does not influence how far this influence reaches. However, their dataset consisted of only 12 plants with a relatively uniform and small capacities McNew and Griffith (2005). The average market city in this dataset had just over 73 million gallons of ethanol per year in nameplate capacity and nameplate capacity ranged from 420 million gallons per year to only a 1 million gallon per year. This suggests that the analysis would benefit by considering different sized neighborhoods based on the size of the plant. I constructed four neighborhoods for each plant in the study based off assumptions on the land devoted to ethanol production around the market city. Neighborhoods were delineated by the radial distance away from the market city.

To construct these neighborhoods I assume that a certain percentage of the land immediately surrounding the plants is planted to corn that is subsequently converted into ethanol and that the plant fills its capacity. I assume constant productivity between plants such that one bushel of corn can produce 2.8 gallons of ethanol (Jackson, 2018). I use simple land conversion rates and yields to then estimate the treatment radius. A square mile (known as a section) has 640 acres. The radius can then be found using the NASS county yields and the formula for the area of a circle. For example, an ethanol plant with 300 million gallon per year capacity in a county with a yield of 160 bushels per acre requires $\frac{\text{300 million gal}}{\text{2.8 gal per bu}} = 107.14$ million bushels of corn. Which would require $\frac{\text{107.14 million bushel}}{\text{160 bu. per acre}} = 669,625$ acres, or $\frac{\text{669,625 acres}}{\text{640 acres per sq. mile}} = 1,046.29$ square miles. This creates a neighborhood about the ethanol plant with a radius of $\left(\frac{1{,}046.29}{\pi}\right)^{\frac{1}{2}} = 18.25$ miles. The neighborhood radius can therefore be computed as equation 3.1. Here cap_{it} is the market city's total capacity in gallons per year, y_i is the average estimated yield measured at the county-level over 2009 to 2016, and *prop* is the proportion of crop acres used for ethanol production. Since the exposure to close ethanol production capacity is considered the treatment, I allow the neighborhoods to vary over time with capacity changes.[2]

$$r_{it}^{prop} = \left(\frac{cap_{it}}{2.8}\frac{y_i}{640}\frac{1}{prop}\frac{1}{\pi}\right)^{\frac{1}{2}} \tag{3.1}$$

Figure 3.3 shows a random sample of the treatment groups under the different distance thresholds in 2016. Observations in blue, red, green, and

[2] I hold yields fixed at the average of the county yields from 2009 to 2016. Since transitions are long-run, I do not consider individuals that go from treated to untreated. That is, I only consider expansions.

black are treated CLUs under the assumptions that 10%, 25%, 50%, and 100% of the area is planted to corn which was then used by the plant to produce ethanol. On average the 100% radius was 8 miles, the 50% radius was 11 miles, the 25% radius was 15 miles, and the 10% radius was 24 miles. While these radii vary by plant, the average distances are in line with the distances considered by (Towe and Tra, 2012) who used radii of 10, 20 and 30 miles. The total sample consists of about 5 million CLU fields. Treatments were assigned by year.

The goal of this study is to determine the impact of an ethanol plant entering a local market. The decisions from fields that had been near an ethanol plant for many years would not be of interest since their decisions could be influenced by post-treatment price variation that is unrelated to the plant itself. Therefore a "treated" field is one that was newly introduced to ethanol plant capacity. In other words, a field that was within a treatment group for the first time in a given year.

Table 3.1 show the proportion of the sample, that was treated in each year under each ethanol land conversion assumption. The relative sizes of the treatment groups span a wide range from only less than 1% of all CLUs being treated in a given year to over 10% all CLUs being treated. The share of treated observations fluctuates according to the incoming ethanol capacity entering in each year.

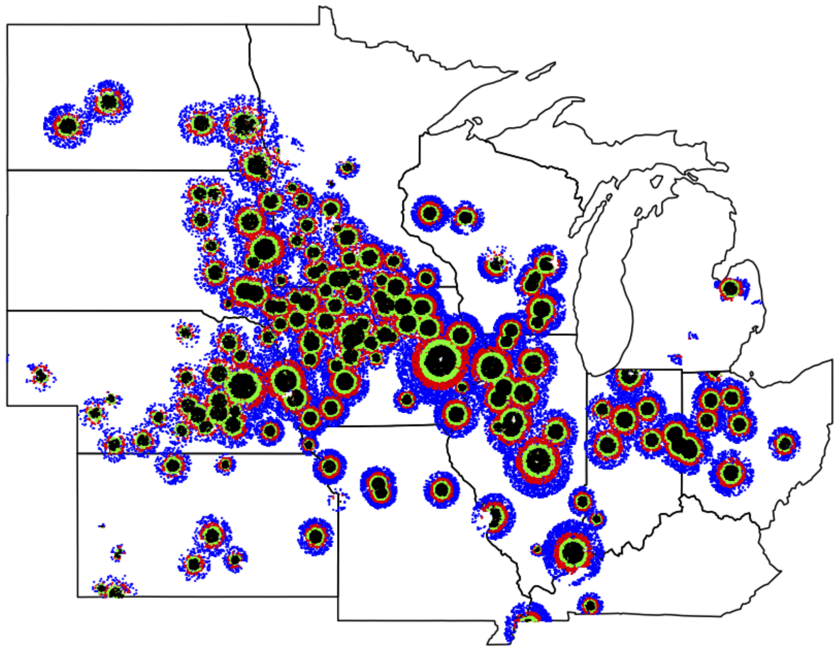

Figure 3.3: *2016 US Ethanol Plant Neighborhoods*

Table 3.1: *Newly Treated Sample Proportions by Year*

	100% Radius		50% Radius		25% Radius		10% Radius	
Year	Treat. Obs.	Treat. Frac.	Treat. Obs.	Treat. Frac.	Treat. Obs.	Treat. Frac.	Treat. Obs.	Treat. Frac.
2009	34746	1.70%	64491	3.40%	111311	6.80%	144610	12.60%
2010	10864	0.50%	23081	1.20%	35410	2.30%	40573	4.00%
2011	4625	0.20%	8622	0.50%	11537	0.80%	13394	1.40%
2012	6206	0.30%	10496	0.60%	19063	1.30%	31979	3.20%
2013	334	0.00%	557	0.00%	632	0.00%	758	0.10%
2014	7019	0.30%	12453	0.70%	17834	1.20%	24260	2.60%
2015	5089	0.30%	8590	0.50%	13393	0.90%	22601	2.50%

3.3 Conceptual Model

This paper does not explicitly study how the quality of cropland change as prices change. However, it is a motivation for understanding extensive decisions. Previous literature has shown land transitioning to and from crop production is less suitable for crop production (Lubowski et al., 2006; Wright and Wimberly, 2013). As a consequence, inputs used on these lands are not as productive and are over-allocated relative to more productive lands. Because of this, and the fact that many agricultural inputs such as fertilizer have environmental externalities, researchers cite that crop production on converted land could be relatively environmentally harmful (Lubowski et al., 2006).

Suppose farmers allocate their land across two different uses, crop production (C), and some other use (O) in order to maximize profits. Land is assumed to vary by some crop suitability index (θ) where higher values of the index indicate higher crop productivity. This suitability varies according to a density distribution ($g(\theta)$) with the proportion of land below some quality $\bar{\theta}$ is $G\left(\bar{\theta}\right)$. The farmer is considered a price-taker in both activities and must decide on the threshold value of land quality to allocate to crop production $\bar{\theta}$. Figure 3.3 shows an example of cumulative distribution and a chosen threshold. It is assumed that all the land is used between these two uses. Therefore the area to the left of the threshold in black will be converted to non-cropland and the remaining under the distribution area will be in crop production. The farmer's production in one activity is assumed not to impact the other activity, that is, production is non-joint between the two activities.

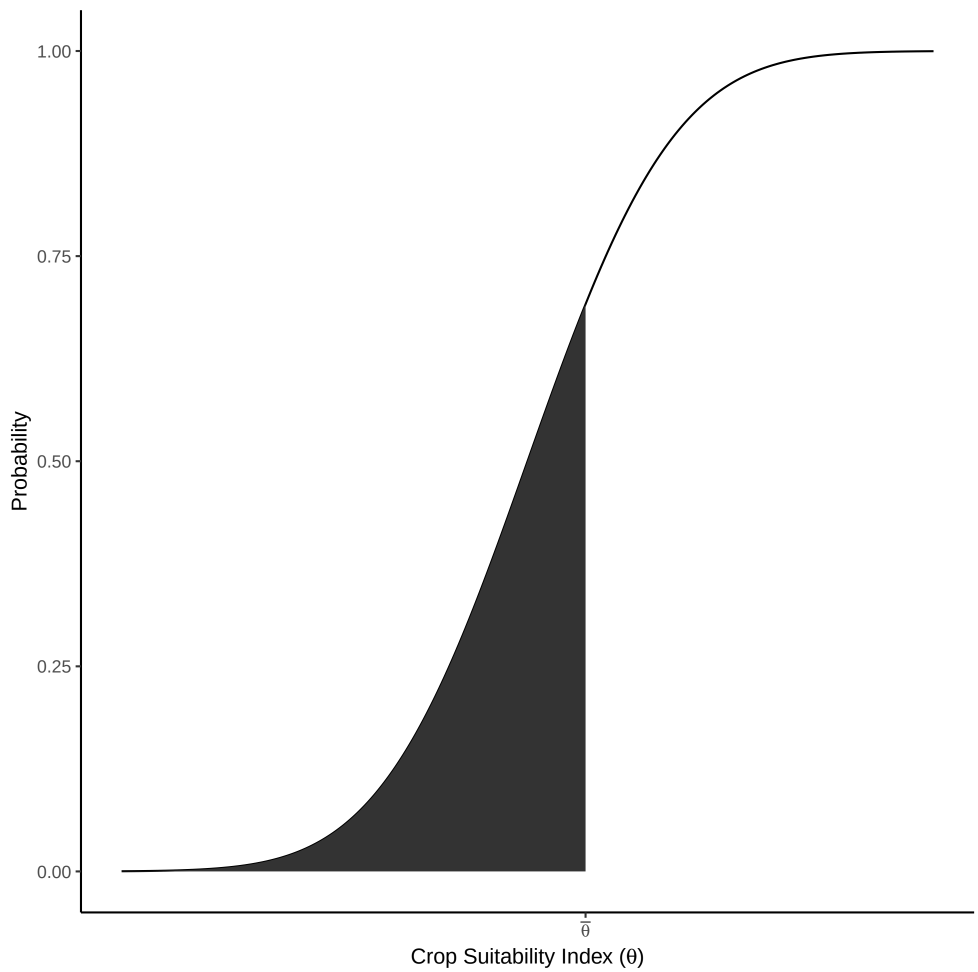

Figure 3.4: *Cumulative Distribution of Cropland Suitability and Crop and Non-Cropland Threshold*

Profits between these crop and non-crop activities differ by the quality of land that is used in the production. In this way the total profit for the farmer can be described as equation 3.2. Where f_C and f_O are the per-acre total product of crop and non-crop activities as a function of the cropland suitability of the land. Without loss of generality, the unit price from the non-crop activity is set to unity and the price of the cropped product is set to p_C. The total profits are determined by the quality within the chosen section of the distribution and by the amount of land area allocated to each activity. Profits are therefore a convex combination of the returns to each respective activity. It is assumed that $\frac{\partial f_C}{\partial \theta}$, $\frac{\partial f_O}{\partial \theta}$ are both positive, $\frac{\partial^2 f_C}{\partial \theta^2}$ and $\frac{\partial^2 f_O}{\partial \theta^2}$ are both negative. Since θ is a crop suitability index, it is also assumed that that the marginal effect of an increase in the land quality is higher for the cropping relative to the non-cropping activity. That is, $\frac{\partial f_C}{\partial \theta} > \frac{\partial f_O}{\partial \theta}$ for all levels of θ. The price-taking profit-maximizing farmer will then select the optimal land-quality threshold by setting the first order condition equal to zero, shown in equation 3.3.[3] The first two terms represent the change in profits due to area effects. Increasing $\bar{\theta}$ reduces the amount of land allocated to cropland and adds it to non-crop uses. The next two terms represent the impact that threshold increases have on the average productivity of land for each use. Increasing $\bar{\theta}$ increases the average productivity across *both* practices.

$$\Pi(\theta) = \max_{\theta} \left\{ G(\theta) f_O(G(\theta)) + p_C (1 - G(\theta)) f_C(G(\theta)) \right\} \tag{3.2}$$

[3] I simplify notation for readability.

$$\frac{\partial \Pi\left(\bar{\theta}\right)}{\partial \theta} = g f_O - g p_C f_C + G f_O' g + (1 - G)\, p_C f_C' g = 0 \tag{3.3}$$

With the first order condition for allocation of land I solve for the change in the land quality threshold using implicit differentiation in equation 3.4. The denominator term is a second derivative of a profit function with respect to a decision variable and therefore should be negative meaning the sign is dictated by the sign of the numerator. The final form shows that the relationship can be represented as a function of the elasticity of marginal product to quality. Conceptually, as long as the increase in land quality dedicated to cropland does not lead to a substantial proportional increase in crop output, the threshold value of cropland will decrease with prices. The effect will be especially high when less land is devoted to cropland (high values of $G\left(\bar{\theta}\right)$) since a small change in land allocation contributes a proportionally large increase in crop output. Since $1 - G$ is less than one, the crop output elasticity will need to be especially elastic for this term to be positive. Since much of the area under crop production already have higher yields, it is unlikely that a reallocation would have this kind of effect and therefore in practice, cropland quality is likely to fall as crop prices increase. Using the same steps, equation 3.5 shows the model also predicts that when the elasticity of crop output is sufficiently small, crop yields will also tend to decrease as crop prices increase.

$$\frac{\partial \bar{\theta}}{\partial p_C} = -\frac{\frac{\partial^2 \Pi}{\partial \theta \partial p_C}}{\frac{\partial^2 \Pi}{\partial \theta^2}} = \frac{g\left[f_C - (1 - G)\, f_C'\right]}{\frac{\partial^2 \Pi}{\partial \bar{\theta}^2}} = \frac{g f_C\left[1 - (1 - G)\,\frac{f_C'}{f_C}\right]}{\frac{\partial^2 \Pi}{\partial \bar{\theta}^2}} \tag{3.4}$$

$$\frac{\partial f_C}{\partial p_C}\mid_{\theta=\bar{\theta}}=\frac{gf_C\left[1-(1-G)\frac{f_C'}{f_C}\right]}{-p_C g}=\frac{f_C\left[1-(1-G)\frac{f_C'}{f_C}\right]}{-p_C} \tag{3.5}$$

3.4 Methodology

The goal of this paper is to estimate the impact that ethanol plant construction has on local land use decisions. Isolating the impact of changing local ethanol plant capacity is difficult because the RFS contributed to both increased ethanol production capacity and elevated crop prices. In addition, where ethanol plants choose to locate is likely non-random. The start-up costs for an ethanol plant can be considerable ranging between just over \$1.00 to \$3.00 per gallon of capacity (Shapouri and Gallagher, 2005). This corresponds to an approximate start-up cost of \$140 million for the average plant in the sample. Problems arise when isolating the causal treatment effect of the ethanol plant since this means that a field's proximity to plants, and therefore the treatment assignment is non-random. Due to this considerable start-up costs, it is reasonable to assume that ethanol plants choose to locate in areas where local producers are willing and able to sufficiently supply the plant. It is likely that the same factors that encourage a field to supply an ethanol plant are also related to the relative profitability of crop production such as yield potential and crop suitability. Models that fail to account for non-random treatment assignment would be biased since they would attribute influence from variables that influence the treatment assignment to the treatment itself.

The Rubin-Neyman causal model (RNCM) provides a theoretical foundation for computing causal estimates. This model in the simplest from has a single treatment year, two states, treated and untreated, and two assignment groups the treatment group and the untreated (control) group. The goal of the RNCM is to estimate the effect that the treatment has on an outcome variable of interest (Y). With this there are three observed outcome values across the sample, for each individual, $Y_{i,t_0,0}$, for individuals in the treatment group before treatment period, $Y_{i,t_1,0}$ for individuals in the control group after treatment period, and $Y_{i,t_1,1}$ for individuals in the treatment group after treatment period.

If individuals are assigned to either a treatment group or a control group ($D = 1$ and $D = 0$ respectfully), and the outcome variables take on values $Y_{i,t_1,1}$ after the individual (i) is exposed to a treatment and $Y_{i,t_1,0}$ if untreated after the treatment period, the goal of the model is said to estimate the average treatment effect on the treated (ATT) shown in equation 3.6. Here X are a set of individual controls, D is the treatment assignment, and Y is the outcome variable. The ATT is of particular interest when treatment assignment is non-random since the estimated effect of the treatment on untreated individuals is generally not relevant outside of prediction applications. Researchers get to observe the treated state of the outcome variable for individuals in the treatment group, and untreated outcome variables for individuals in the control group. They do not get to observe the counterfactual outcomes, the untreated outcome variable for the treatment group and the treated outcomes for the control groups. Therefore the challenge that underlies producing a causal estimate is that while the first term in equation

3.6 is observed, the second term is generally unobserved.

$$ATT = \mathbb{E}\left[Y_{i,t_1,1} \mid X, D = 1\right] - \mathbb{E}\left[Y_{i,t_1,0} \mid X, D = 1\right]. \tag{3.6}$$

While the treatment effect cannot be observed directly, making certain assumptions allows researchers to estimate it. In particular, if researchers assume that treatment was assigned at random and the outcome variable was identically distributed across the treatment assignments, a researcher could simply use the control group's observed average post-treatment outcome as an unbiased estimator of counterfactual outcome ($\mathbb{E}\left[Y_{i,t_1,0} \mid X, D = 0\right]$). In many cases however, the treatment assignment is non-random, making this invalid.

The difference-in-differences (DID) method is one way estimating the treatment effect. The DID is a panel approach in which the pre- and post-treatment period values are used to estimate the treatment effect. The first "D" in the DID approach comes from the fact that we are computing the before and after treatment outcome variables for both the treated and control groups. The second "D" is when these differences are differenced between the treatment and control groups. By the second differencing step, the DID approach has the benefit of controlling for time varying variables that are common to both the treatment and control groups. This is useful in this study since the RFS policy impacted the world prices for major commodities as well as plant construction. This means that areas that were outside of ethanol-producing regions also likely experienced elevated prices.

Smith and Todd (2005) provide a good overview of the DID matching methodology. They represent the treatment effect estimation procedures using two distinct equations for treated individuals and untreated individuals in equations 3.7 and 3.8. Where the U terms have a mean of zero. The outcome variable can be written as a linear combination of the two left hand side variables such that $Y_{it} = D_i Y_{1it} + (1 - D_i) Y_{0it}$. Therefore, the outcome variable between the two groups can be represented in a the single equation shown in 3.9. Here $\alpha(X_{it}) = \phi_1(X_{it}) - \phi_0(X_{it}) + U_{1it} - U_{0it}$ is a function of the controls as are the ϕ terms. The *alpha* term is the treatment effect and, represented this way can vary according to the control variables. The goal is therefore to find the estimate of α which in case of this analysis is the conditional treatment effect of changes in local ethanol plant capacity conditioned on cropland transition variables.

$$Y_{1it} = \phi_1(X_{it}) + U_{1it} \tag{3.7}$$

$$Y_{0it} = \phi_0(X_{it}) + U_{0it} \tag{3.8}$$

$$Y_{it} = \phi_0(X_{it}) + D_i \alpha(X_{it}) + U_{0it} \tag{3.9}$$

The DID approach uses the pre-and post-treatment values from the treated and control groups to estimate the treatment effect (α). Here t_1 is the post-treatment time period and t_0 is the pre-treatment time period. The DID approach requires several assumptions to be valid. The first is that the differences between the treated and untreated error terms are mean zero ($\mathbb{E}[U_{1it} - U_{0it}] = 0$) and that the error difference and therefore the difference in the conditional outcome variables are independent ($\mathbb{E}\left[\left(U_{1it} - U_{0it}\right) D_i\right] = 0$). This means that the treatment assignment is not correlated with the conditional estimate of the outcome variables which precludes endogeneity in treatment assignment. Furthermore, the conventional DID approach also assumes for a given time period t, $\left(\phi_1\left(X_{it}\right) - \phi_0\left(X_{it}\right)\right)$ is a constant. This is called the parallel trend assumption.

Under these assumptions, the standard DID estimator can be written as equation 3.10. A look at equation 3.10 shows that the DID treatment effect estimate $\left(\alpha^{DID}\right)$ amounts to a break from the trend in the output value pre and post-treatment that is attributed to the individuals in the treatment group. This is illustrated in figure 3.5, where the treatment effect is the difference between the observed post-treatment outcome and the predicted counterfactual of the treatment group in the post treatment period. It also graphically shows the parallel trend assumption at work as the treatment effect is estimated as the difference between the realized and assumed parallel counterfactual trend. While the counterfactual trends in the outcome variables are assumed constant across the treatment and the control groups, the value of the output variable can differ across the groups. In the case depicted in figure 3.5, the treatment group groups outcome variable is twice as large as the control group. This is useful since RFS policy is responsible for an increase in ethanol plant construction as well as an estimated 30% increase in corn prices (Roberts and Schlenker, 2013). By differencing, the DID estimator removes the influence of the common increase in crop prices that both the treated and untreated fields experienced.

$$Y_{it_1} - Y_{it_0} = \phi\left(X_{it_1}\right) - \phi\left(X_{it_0}\right) + D_i \alpha^{DID} + \left(U_{it_1} - U_{it_0}\right). \qquad (3.10)$$

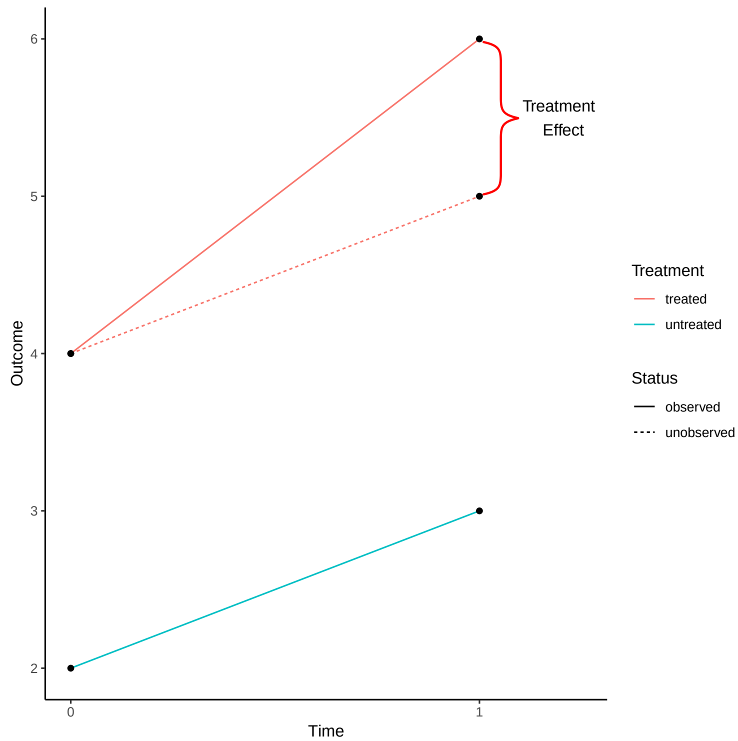

Figure 3.5: *Treatment Effect Illustration*

With its focus on the change in outcome variables across time and groups, the DID approach is useful in many settings. However, the parallel trend assumption may not be valid. With respect to ethanol plants, it is reasonable that ethanol producers select where to place plants based off of profit motivations. In chapter 2 I show that there is a high degree of price response heterogeneity across the country. Plant investors may therefore choose to construct plants in areas where farmers are more sensitive to price changes.

Areas where long-term extensive transitions are not as sensitive to changing crop prices limit the nearby crop product that could be sold to the plant when prices rise.

Matching estimators are one way to address failures in the parallel trend assumption. The standard DID approach assumes that all individuals in the treatment group would have the same counterfactual trend as every individual in the control group. Matching relaxes this assumption by conditioning it on the values of observables (X) where (X) is similarly defined in equation 3.9. Matching estimators rely on the assumption that individuals inside and outside of the treatment groups are functionally identical provided that the features that impact the likelihood of being in the treatment group are controlled for. Formally this means that $Y_1, Y_0 \perp\!\!\!\perp D \mid X$ meaning that the values of the outcome variables are independent of the treatment assignment given independent variables. Without the DID approach the matching estimator would be:

$$\alpha^M = \frac{1}{n_1} \sum_{i \in I_1 \cap S_P} \left[Y_{1i} - \hat{\mathbb{E}} \left[Y_{0i} \mid D_i = 1, P_i \right] \right], \tag{3.11}$$

where P_i is the i^{th} individual's probability of being in the treatment group, I_1 is the set of individuals in the treatment group, S_P is the common support, n_1 is the number of treated observations, and $\hat{\mathbb{E}} \left[Y_{0i} \mid D_i = 1, P_i \right] = \sum_{j \in I_0} W\left(i, j\right) Y_{0j}$. Here $W\left(i, j\right)$ is the weight between the observation (i) in the treatment group and the observation (j) in the control group. $W\left(i, j\right)$ is larger when the probability of treatment is more similar between i and j.

While I've stated how matching works generally, in practical applications, observations are matched using many independent variables. As the dimension of the independent variables grows, it becomes less likely that matches will be found due to the curse of dimensionality (James et al., 2013). To control for this, often *propensity score* matching is used where individuals close in some propensity score $\phi(X)$ are assumed to be close matches. Oftentimes the propensity score is an estimated probability of being in the treatment group given the independent variables. A slight extension to the conditional independence is needed to accommodate propensity score matching $Y_1, Y_0 \perp\!\!\!\perp D \mid \phi(X)$. Conceptually this means that if individuals in the treatment groups and control groups that are close in the propensity score are sufficiently similar to one another, then the treatment assignment is the only feature that can explain average difference between the treatment and control outcomes.

There are many ways to assign weights (and define the support S_P) in the matching procedure. Like many statistical procedures, the choice of the matching procedure manages a bias-variance trade-off. A single-nearest neighbor match tends to have low bias since only the closest individual in the group control is matched to each treated individual. This procedure has high variance since the counterfactual value is influenced by a single observation within the sample. Since the dataset is large, I use a ten-to-one oversampled nearest neighbor match where every treated field is matched with the ten nearest control fields. A drawback to oversampling is that controls become less comparable to the treated field. To regulate this, I restrict relevant control fields to be within a quarter of a propensity score standard deviation. To assign the weights I set $W(i, j)$ equal to one if treated individual i is matched with the control observation j and zero otherwise.

$$\alpha^{DIDM} = \frac{1}{n_1} \sum_{i}^{n_1} \left[Y_{1t_1 i} - Y_{0t_0 i} - \sum_{j \in I_0 \cap S_P} W(i,j) \left(Y_{0t_1 j} - Y_{0t_0 j} \right) \right] \tag{3.12}$$

The primary benefit for using propensity score matching in this study is that matching helps control for heterogeneity in observable characteristics that can either impact the treatment assignment or the treatment effect. Without matching, non-random treatment assignment can bias the ATT estimate. While the standard DID approach helps address potential design problems related to non-random treatment assignment, it relies on the parallel trend assumption. In the face of heterogeneous treatment imply that the parallel trend assumptions are valid for some subsamples of the treatment and control groups but across the groups. Figure 3.5 illustrates that while the outcome variable was higher in for the treated group, the DID approach can still properly estimate the treatment effect. In the context of this study this means that the adjustments that treated and untreated fields to their output values are expected to be the same the pre- and post-treatment periods absent the influence from the treatment. Demonstrated in the earlier chapter, supply response to prices was highly heterogeneous across the US. If ethanol plants were aware of this, they may choose to locate in areas with supply that was more or less price sensitive.

3.4.1 Expected DID Results with Conditional Transitions

In the next section I present the results of the difference in difference models. Before I move on to the results I will explain what the expectations on the signs are and how to interpret the tables. The dependent variables are conditional changes in extensive uses, cropland or non-cropland. My initial expectations are that in the presence of ethanol plant capacity, a field that is currently in non-cropland would be more likely to switch to cropland after the initial influence of an ethanol plant. If on the other hand the field were currently used to produce crops, I would expect the ethanol plant exposure to help retain area in cropland.

Suppose that new ethanol plant enters a region in 2011 and I have two identical fields inside and outside of this region. Suppose in 2008 both fields begin in cropland and remain in cropland until 2011. I expect that the influx of new demand to entice the treated field to convert from non-cropland to cropland. Since I do not allow more than one switch, this transition is considered permanent. Tables 3.2 and 3.3 show the scenarios for the treated and control fields respectively. A single year is lost when computing the conditional transition, and an additional year is lost when computing the differences in transitions across time. Because the treated field transitioned to cropland in 2011, they had a non-cropland to cropland conditional transition in that same year. Differencing the non-cropland to cropland conditional transition in the prior year, the difference will be a positive 1 in 2011. While the treated field does not transition from non-cropland to cropland in the subsequent year, the conditional transition will *not* be (-1) in 2012. The reason is that it is a conditional transition and therefore will be undefined when the previous land use was not in non-cropland. Differencing the two differences, table 3.4 shows that expected DID estimate to be positive.

Table 3.2: *Expected Outcome For Treated when Treatment in 2011 and Initially Non-Crop*

Year	Crop (C)	Non-Crop (N)	$(C \mid N)_1$	$\Delta\,(C \mid N)_1$
2008	0	1	–	–
2009	0	1	0	–
2010	0	1	0	0
2011*	1	0	1	1
2012	1	0	0	-1
2013	1	0	0	0

Table 3.3: *Expected Outcome For Control when Treatment in 2011 and Initially Non-Crop*

Year	Crop (C)	Non-Crop (N)	$(C \mid N)_0$	$\Delta\,(C \mid N)_0$
2008	0	1	–	–
2009	0	1	0	–
2010	0	1	0	0
2011*	0	1	0	0
2012	0	1	0	0
2013	0	1	0	0

Table 3.4: *Expected Cropland Transition DID*

Year	$\Delta\,(C \mid N)_1$	$\Delta\,(C \mid N)_0$	DID
2010	0	0	0
2011*	1	0	1

Tables 3.5 and 3.6 show the same scenario except each field is initially planted to cropland. In this case, I would expect that after 2011, fields that

are exposed to local ethanol demand would be more likely to retain land in cropland. If this is the case then the control group, without the influence of the plant, may transition its land to non-cropland. Since the statistic of interest is crop retainment or the conditional transition to cropland given cropland was initially planted, the control group would therefore have a negative change in cropland retainment. As before, since these are conditional transitions, they are only comparable if in the previous year, both fields were in cropland. Therefore, the expected cropland retainment DID is still positive, as shown in table 3.7.

Table 3.5: *Expected Outcome For Treated when Treatment in 2011 and Initially Crop*

Year	Crop (C)	Non-Crop (N)	$(C \mid C)_1$	$\Delta\,(C \mid C)_1$
2008	1	0	–	–
2009	1	0	1	–
2010	1	0	1	0
2011*	1	0	1	0
2012	1	0	1	0
2013	1	0	1	0

Table 3.6: *Expected Outcome For Control when Treatment in 2011 and Initially Crop*

Year	Crop (C)	Non-Crop (N)	$(C \mid C)_0$	$\Delta\,(C \mid C)_0$
2008	1	0	–	–
2009	1	0	1	–
2010	1	0	1	0
2011*	0	1	0	-1
2012	0	1	0	0
2013	0	1	0	0

Table 3.7: *Expected Cropland Retainment DID*

Year	$\Delta\left(C \mid N\right)_1$	$\Delta\left(C \mid N\right)_0$	DID
2010	0	0	0
2011*	0	-1	1

3.4.2 Ethanol Expansion Areas by Year

The use of a DID matching procedure was motivated by its ability to purge bias resulting from distributional differences of relevant observables. The standard DID approach is useful on its own when the parallel trend assumption is valid across the treatment and control groups. To illustrate the general relationship between ethanol capacity and cropland transitions, I first show the results of four event studies. Event studies are a simple extension of the DID approach which allows the impact of the treatment to vary at different points in time relative to the treatment time period. Modeling as an event study is useful since ethanol plants take around 2 years to complete on average meaning that the land use decisions could be impacted by the anticipation of incoming ethanol demand during the construction phase. The event study considers the distinct treatment effect differences as far back as 4 years before the treatment and up to 6 years after the initial introduction of ethanol capacity.[4]

The matching version of the DID estimator is the most complicated model I ran in this paper. Not every plant was constructed in the same year and fields that were previously treated are not suitable controls in the matching

[4]Lags greater or equal to four are combined to ensure adequate variation in the variable. Therefore, the "minus 4" term will be 1 for a field treated in 2015 will be 1 when year is 2009, 2010, and 2011.

process. I divided the sample in several ways to construct the datasets that I eventually performed the matching over. Firstly, since the outcome variables of interest were conditional transitions from cropland and non-cropland, I only considered fields that had cropland planted in the previous period when estimating the effect of ethanol plant on the probability of cropland retention. Likewise, for cropland conversion, I considered only fields that were not in cropland in the prior period. Second, because previously treated observations are not suitable controls for the contemporaneously treated individuals and ethanol plants entered markets in different years I subset the samples further by year where for a given year, only fields that were either exposed to ethanol plant capacity in the given year, never exposed to ethanol plant capacity, or were exposed at least one year after the current year. I made this last restriction because ethanol plant construction takes one to two years and it is reasonable that farmers in the area would anticipate new capacity entering the area in the year leading up to its opening.

Fields were considered treated if they were exposed newly to ethanol plant capacity. Because of this, the treated groups in a given year tended to decline over time. Figures 3.6 and 3.7 show the ethanol expansion areas when I assume either a 25% or 50% area corn-ethanol conversion rate. The radial assumptions change the shape of the expansion areas where under smaller radii there are expansion areas within more dense ethanol coverage and under larger radii there are larger expansion areas near the fringe of corn-growing regions. While the literature gives some attention to the radius size, to my knowledge no one has discussed that smaller radii tend to designate more treated fields in areas with denser ethanol plant markets. The figures also show that ethanol expansion was more frequent in the earlier years. Much of the expansion in 2009 was from new plant construction while later years expansion was from additional capacity being added to existing markets. The 2013 crop year saw the smallest increase in ethanol capacity with caused by modest expansions to existing capacity. In all but 2013, for every year of the study, at least one ethanol plant came online in a new market.

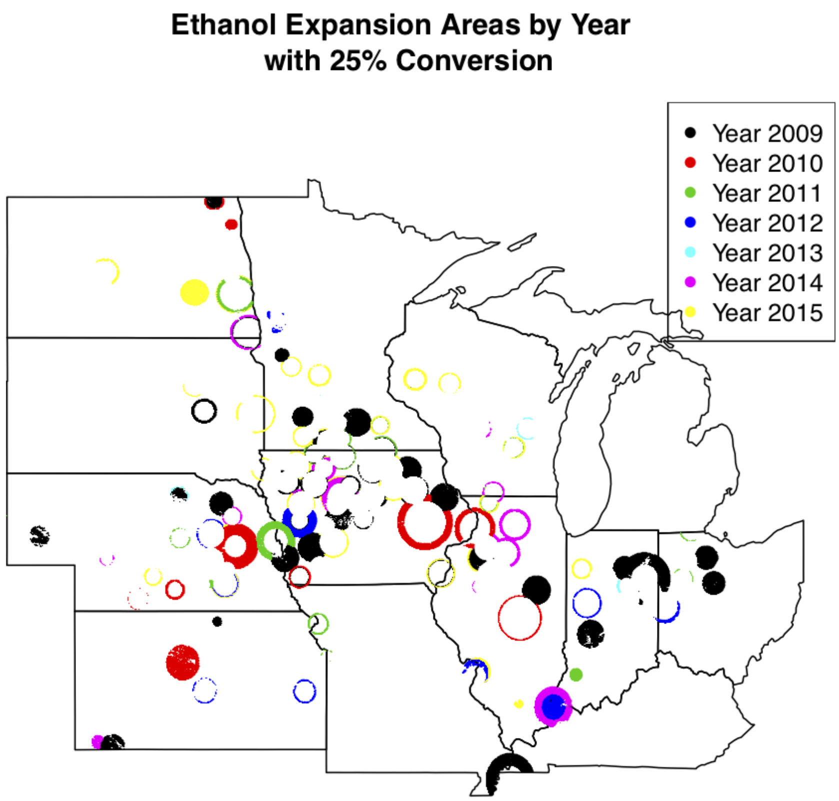

Figure 3.6: *Areas of Ethanol Expansion by Year with 25% Conversion Assumption*

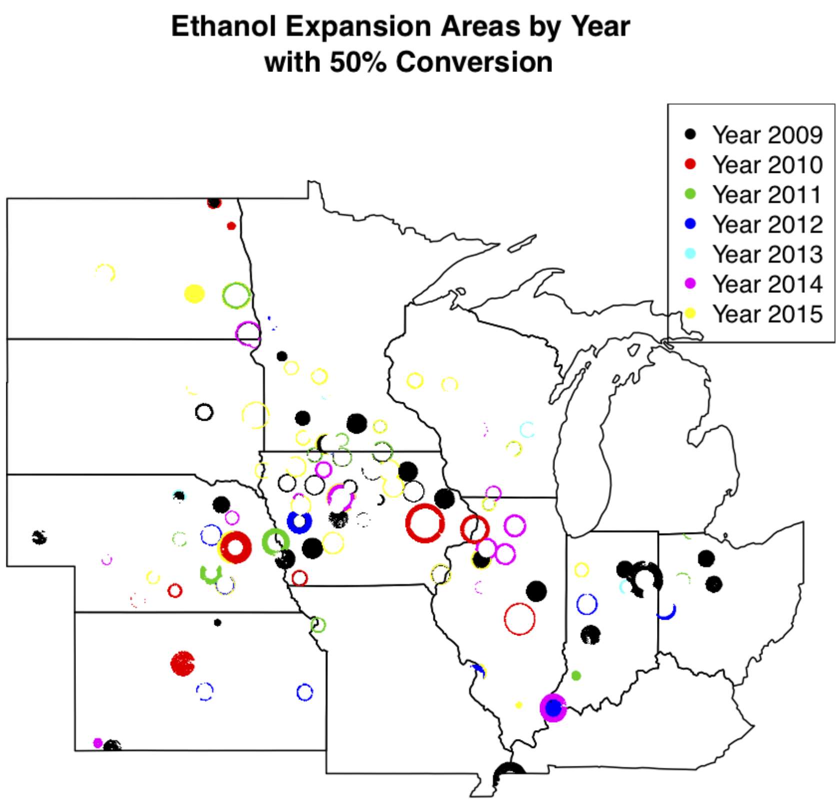

Figure 3.7: *Areas of Ethanol Expansion by Year with 50% Conversion Assumption*

3.5 Results

3.5.1 Transitional Summary

I start the results section by summarizing the starting and ending status of land in the dataset. Table 3.8 provides a summary of fields in the dataset starting in either cropland or non-cropland in 2008 and their respective ending status by 2016. Before sub-sampling, the total dataset consisted of a little over 5.6 million fields with approximately equal proportions in non-cropland and cropland.

Table 3.8: *Fields by Starting and Ending Status from 2008 to 2016*

		Starting Land Status	
		Non-Cropland	Cropland
Ending Land Status	Non-Cropland	95.1% (2,477,004)	0.7% (20,072)
	Cropland	4.9% (127,530)	99.3% (3,017,893)
	Total Count	2,604,534	3,037,965

Overall, fields tended to retain their original status. Cropland conversion was far more common than cropland abandonment with over 95% of fields starting in non-cropland remained in non-cropland over the course of the study, and over 99% of fields starting in cropland remained in cropland. In the subsections to come these numbers are good to keep in mind. The lack of general movement in cropland status by the majority of lands restricts the size of the coefficients that follow. For instance, plants could not, on average, reduce the probability of cropland conversion by over 5%. Many of the results to follow have small coefficient values. In context however, they can be fairly influential. For instance a 0.1% increase in cropland retention means the probability of cropland abandonment falls from 0.7% to 0.6%, a proportional decrease of 14% in overall cropland abandonment. A 0.1% increase in cropland conversion increases the probability of cropland conversion from 4.9% to 5.0% which is proportional increase of 0.02 in overall cropland conversion.

3.5.2 Difference-In-Differences Results

Here I present the results of the standard DID procedure across every radius assumption over two subsamples. The first subsample is a standard subsample of across the entire national dataset. I call this the "unrestricted" sample. For the second subsample I use only those fields that were considered a part of the common support during the propensity score matching procedure. Separate matching is performed over each year so I draw a random sample over fields in at least one of the common supports for one of the years of the analysis. Due to computational complexity I took a random sample from these 5 million fields. Both subsamples contain 60,000 fields with approximately 8 observations between 2009 and 2016. Tables 3.9 and 3.10 show the results of the standard DID approach estimated using a year and field-fixed linear model estimated over unrestricted samples. Table 3.9 shows that being exposed to ethanol plant capacity makes fields more likely to retain land in cropland.

Exposure to new ethanol plants increased cropland retainment by around a tenth of a percent. This corresponds to approximately a 14% decrease in the amount of abandoned cropland. While insignificant at the 100% and

50% radii, it was significant at the 95% and 90% confidence level for the 25% and 10% radii assumptions respectively. There are two potential reasons for the differences in significance, firstly because they are larger, the 25% and 10% radii generally consider more fields as treated this leads to a higher proportion of treated observations in a given dataset. This in turn reduces the estimated variance on the treated coefficient. A second reason could be due to the fact that the larger radii assumptions (e.g. the 10% and 25%) tend to designate more fields outside of dense markets as "treated" while smaller radii tend to "fill in" within dense markets as capacity expands. Fields outside the traditional corn markets may be less likely to inherently retain cropland which could lead to more dramatic changes once ethanol plant enter their area. Table 3.10 shows the DID results for cropland conversions. When considering conversion, the treatment effect was not consistent in sign across the different radii and was insignificant.

Table 3.9: *DID Results for Cropland Retention Over Unrestricted Sample*

	Dependent variable:			
	Crop-to-Crop Transitions			
	100 Pct. Radius	50 Pct. Radius	25 Pct. Radius	10 Pct. Radius
	(1)	(2)	(3)	(4)
Ethanol Plant Neighborhood	0.001	0.0001	0.001^{**}	0.001^{*}
	(0.001)	(0.0005)	(0.0004)	(0.0003)
Fixed Effects	Yes	Yes	Yes	Yes
Observations	470,056	470,247	469,785	470,136
F Statistic	83.883^{***} (df = 8; 410048)	76.543^{***} (df = 8; 410239)	86.714^{***} (df = 8; 409777)	82.628^{***} (df = 8; 410128)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 3.10: *DID Results for Cropland Conversion Over Unrestricted Sample*

	Dependent variable:			
	Non-Crop-to-Crop Transitions			
	100 Pct. Radius	50 Pct. Radius	25 Pct. Radius	10 Pct. Radius
	(1)	(2)	(3)	(4)
Ethanol Plant Neighborhood	0.0002	−0.001	−0.0003	−0.001
	(0.003)	(0.002)	(0.002)	(0.001)
Fixed Effects	Yes	Yes	Yes	Yes
Observations	466,733	466,551	466,255	466,378
F Statistic	326.064*** (df = 8; 406725)	331.570*** (df = 8; 406543)	326.623*** (df = 8; 406247)	330.381*** (df = 8; 406370)
Note:				*p<0.1; **p<0.05; ***p<0.01

The common support represents the fields with a propensity score above the lowest propensity score for treated observations and below the highest propensity scores for the untreated observations. Conceptually the common support is a set of observations where the treated and control groups have a comparable conditional probability of receiving treatment at some point in time. Using observations that are only in the common support ensures that observations between the treatment and control group are somewhat comparable in observations that could influence treatment effects. The common support is plotted in figure 3.8 where non-common support observations are in red and common support observations in black. Although there was considerable geographical overlap, there were areas with fewer fields within the common support. The circular areas inside were generally areas that had long-established ethanol plants. Since these plants were long established, they were removed altogether from the common support. The common support also removed areas with marginal cropland in the western Dakotas, southern Missouri, northern Wisconsin and Minnesota and eastern Ohio. Tables 3.11 and 3.12 show the results of the DID models over different radii assumptions for cropland retainment and cropland conversions over the common support of within the matching procedure.

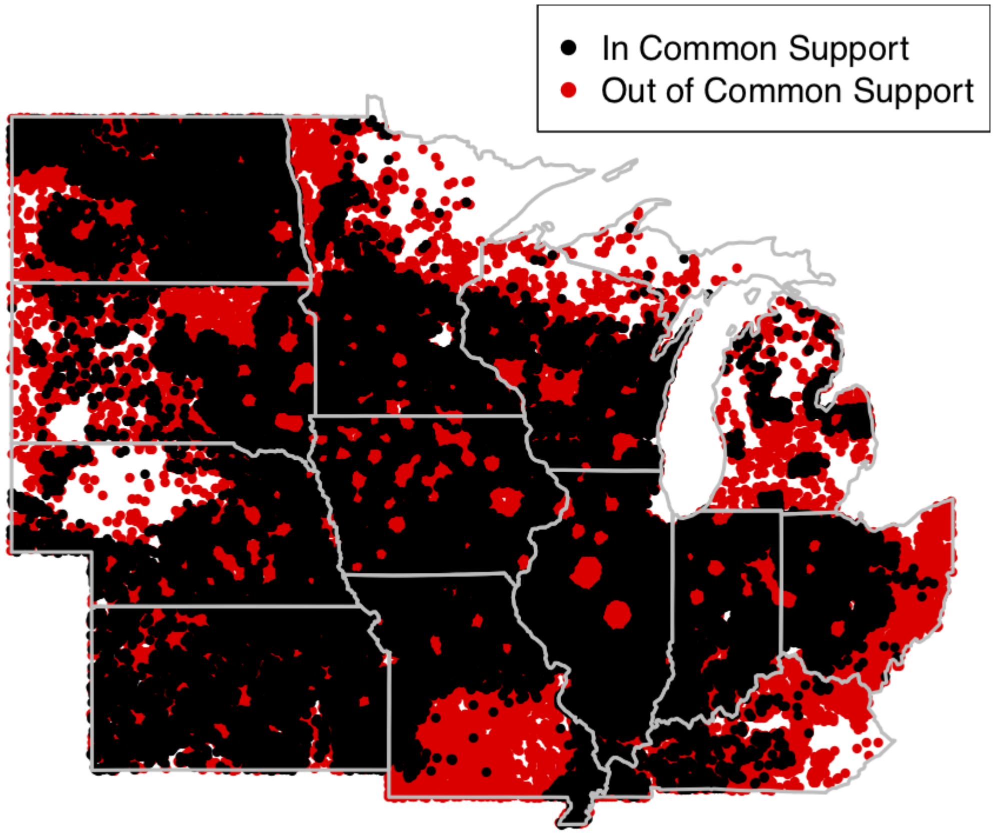

Figure 3.8: *Common Support Area with the 50% Radius Assumption*

Tables 3.11 and 3.12 show the DID results for observations in the common support. In general, the cropland retention results were very similar across the radii. At the 100% radius, the treatment effect was not statistically significant. and the coefficients had smaller standard errors under larger radii. While most of the coefficients were positive, they lack statistical significance. The more interesting results were on the non-cropland to cropland transitions. While the model suggests there is a negative treatment effect between ethanol plants and cropland conversion and the effect is even significant under the 50% treatment radius, the sign flips positive and is highly significant at the 10% radius indicating a 0.3% increase in cropland conversion under the 10% radius. This corresponds to a 6% increase in the lands converted to cropland. Since larger radii "treat" more of the sample outside of dense ethanol markets, this could suggest that the treatment effect is larger in areas where the ethanol industry is not as established. This is consistent with the recent findings of Ifft et al. (2018) who note that CRP policy changes increased the rate of CRP acres retained in CRP and was particularly influential in areas with greater ethanol production.

Table 3.11: *DID Results for Cropland Retention Over Common Support*

	Dependent variable:			
	Crop-to-Crop Transitions			
	100 Pct. Radius	50 Pct. Radius	25 Pct. Radius	10 Pct. Radius
	(1)	(2)	(3)	(4)
Ethanol Plant Neighborhood	−0.00002	0.001	0.001	0.001
	(0.001)	(0.0005)	(0.0004)	(0.0003)
Fixed Effects	Yes	Yes	Yes	Yes
Observations	471,058	470,575	468,462	465,424
F Statistic	67.281*** (df = 8; 411050)	61.144*** (df = 8; 410567)	84.240*** (df = 8; 408454)	98.028*** (df = 8; 405416)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 3.12: *DID Results for Cropland Conversion Over Common Support*

	Dependent variable:			
	Non-Crop-to-Crop Transitions			
	100 Pct. Radius	50 Pct. Radius	25 Pct. Radius	10 Pct. Radius
	(1)	(2)	(3)	(4)
Ethanol Plant Neighborhood	−0.003	−0.004**	−0.0004	0.003***
	(0.003)	(0.002)	(0.002)	(0.001)
Fixed Effects	Yes	Yes	Yes	Yes
Observations	463,487	464,033	465,386	466,712
F Statistic	396.613*** (df = 8; 403479)	383.124*** (df = 8; 404025)	357.353*** (df = 8; 405378)	341.349*** (df = 8; 406704)
Note:				*p<0.1; **p<0.05; ***p<0.01

3.5.3 Event-Study Results

I repeat the DID analysis by considering crop retention and cropland conversion models but now consider the event study models. Event studies are nearly identical to the DID approach. Instead of a single treatment effect variable that remains one for every post-treatment year, event studies separately consider the effect of being inside of a treatment group before and after treatment. In this event study, I include 11 different year-treatment combinations ranging from being in a treatment group 4 years before treatment to being in a treatment group 6 periods after treatment. Figures 3.11 and 3.12 show whisker plots of each of the year-treatment coefficients under unrestricted samples across all four treatment radii. Across these figures the coefficient variance tends to decrease as the treatment-year lags decline. This is because the majority of treated individuals were treated in earlier sample years (e.g. 2008 and 2009). The relative lack of variation in the pre-treatment variables led to higher variance. Likewise, the variance on coefficients related to the larger 10 and 25% radii were smaller as well. This is because larger radii treat more observations leading to higher variability in the treatment-year variables. Interestingly, 3.11 shows that being in an ethanol expansion area four years before the expansion takes place were less likely to retain cropland. This persisted until two years after the expansion had taken place after which, there was no statistically significant effect. It is unlikely that an ethanol plant would impact planting decision 4 years prior to its construction. It is more likely that ethanol plants locate in areas where there is less cropland retention and then after the expansion takes place, the plant has a negligible effect on cropland retention. A low rate of cropland retention may be a consequence of lower crop prices which would correspond to lower input costs to the plant. The results of the cropland transition event study in figure 3.12 show that, while often not significant, there is a small increase in the probability of transitioning to cropland following an ethanol plant expansion. Figures 3.11 and 3.12 are the event study coefficients for cropland retention and cropland conversion when the sample is restricted to the common support. These graphs have the same general trends as their unrestricted counterparts but many of the coefficients become insignificant at the 95% confidence level.

3.5.4 Matching DID Results

The results of the event study suggest that the expansion of ethanol plants were not random. The non-zero pre-treatment effects from the event study were detected up to four periods before the plant came into the area. This suggests that the treatment assignment is significant correlated with cropland choices. This and the the attenuation of the post-treatment effects under the common support seems to suggest that the DID approach can be improved through matching. Tables 3.13 and 3.14 show the average treatment effect on the treated (ATT), for each year, their respective p-values, and a 95% confidence intervals under the 25% and 50% radii assumptions for cropland retention and transitions respectively. Matching requires pairing treated fields with control fields that were either never treated or treated at least one period after the given year. To compute the p-values and confidence

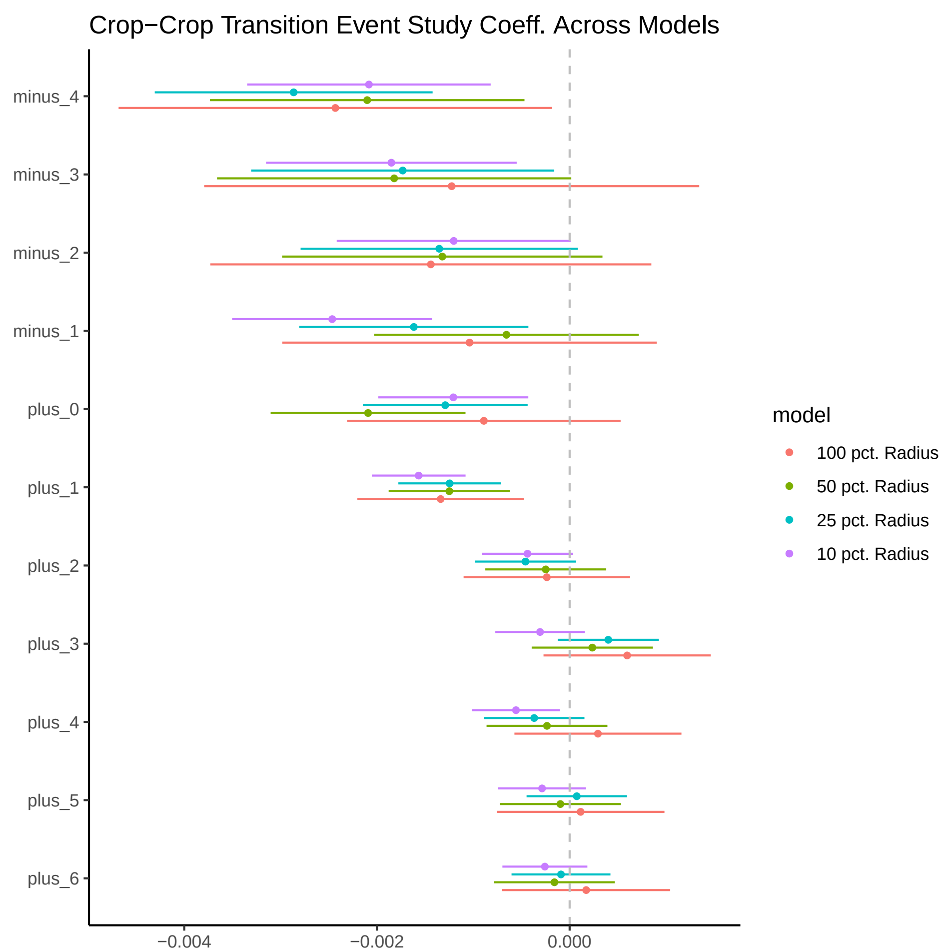

Figure 3.9: *Cropland Retention Event Study Treatment Coefficients Over Unrestricted Sample*

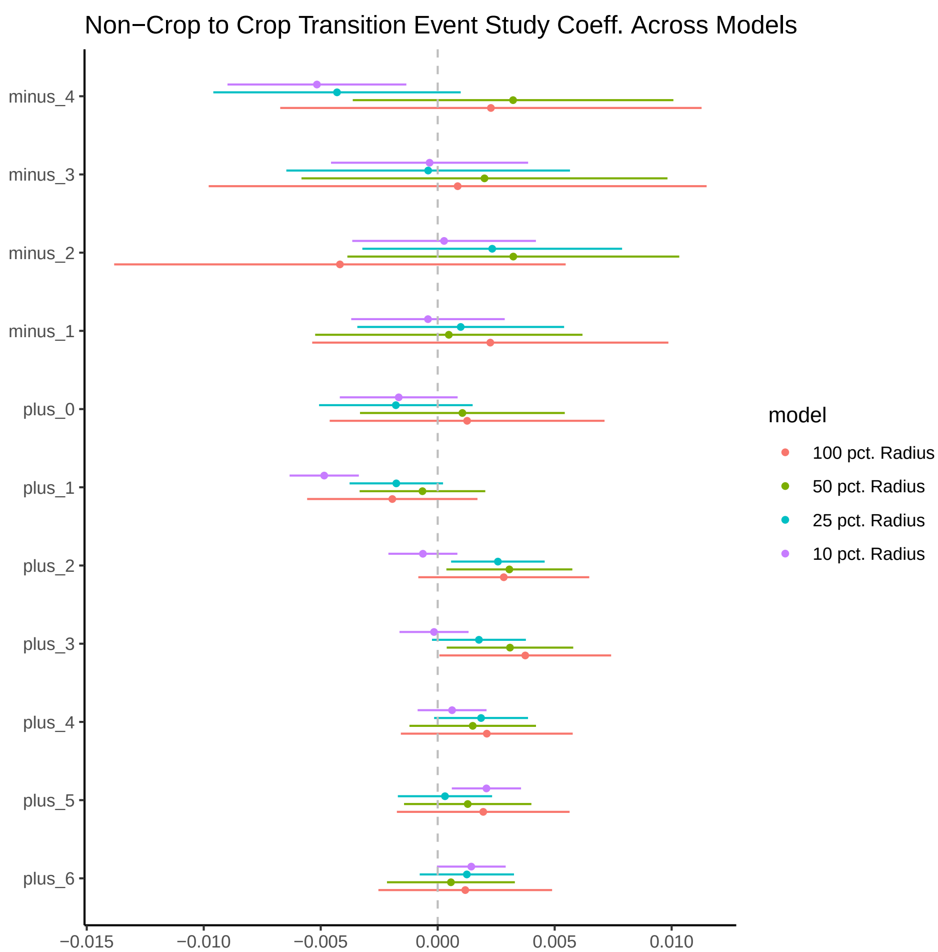

Figure 3.10: *Cropland Conversion Event Study Treatment Coefficients Over Unrestricted Sample*

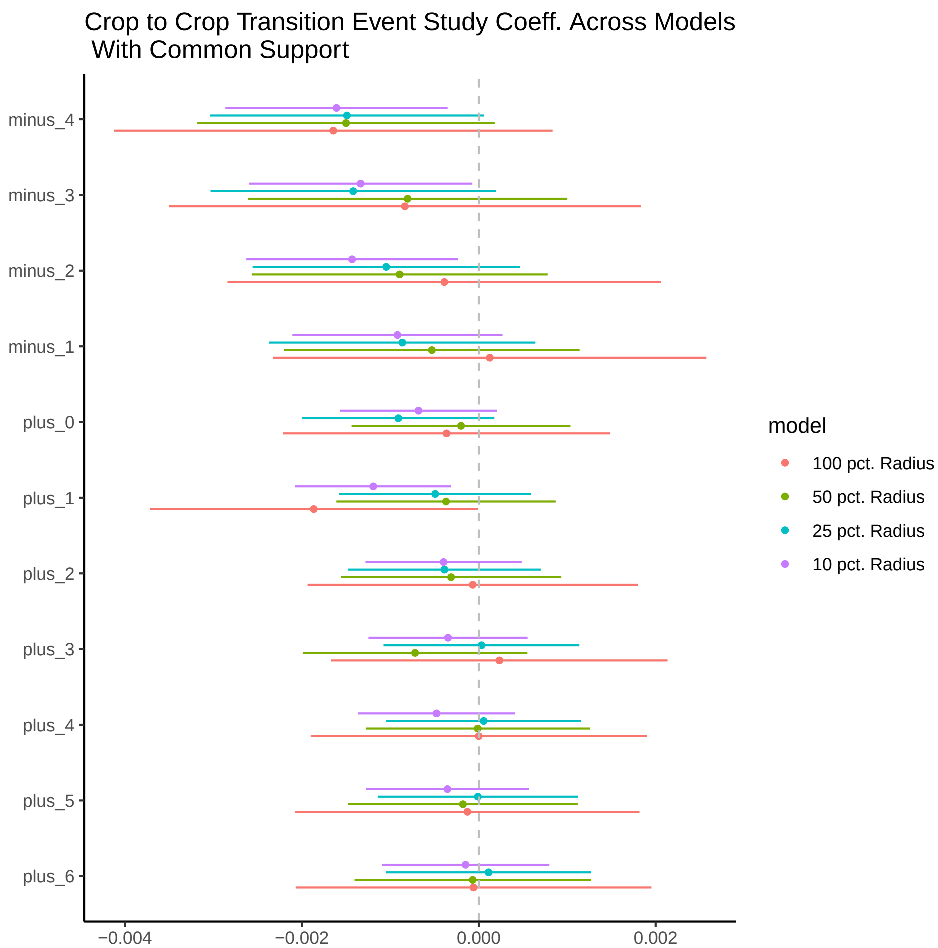

Figure 3.11: *Cropland Retention Event Study Treatment Coefficients Over Common Support*

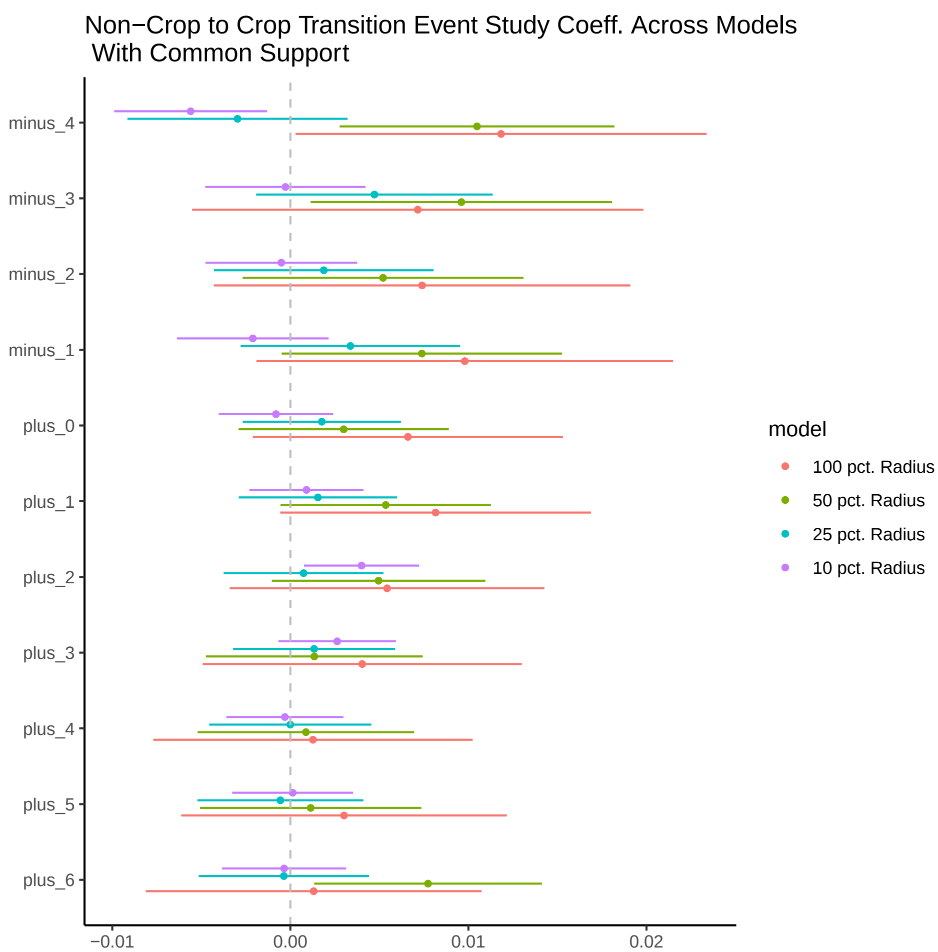

Figure 3.12: *Cropland Conversion Event Study Treatment Coefficients Over Common Support*

intervals, I used a Welch two-sample t-test. In 2013 the government first adjusted the mandate in response to economic and environmental constraints of reaching the increasing blending requirements. This lead to a temporary stagnation in ethanol construction due to regulatory uncertainty. While construction resumed after 2013, it was considerably slower than the initial RFS policy years of 2008 and 2009. Since, under the matching procedure, for each year, newly treated observations needed to be matched with control observations and there was a lack of expansion in 2013, these observations were removed from all matched models. Further years were removed from the 50% radius matched DID study there was a lack of variability in the outcome variable for the treated groups. The matched DID procedure indicates shows that in most years, there were significant treatment effects. In particular cropland retention was significantly higher in areas with expanding ethanol production. The probability of land transitioning to cropland dropped when ethanol capacity entered the market. The confidence interval indicates that new ethanol production lead to up to a 1% reduction in cropland.

While they may seem contradictory, taken together, these results agree with the recent work of Ifft et al. (2018). They found that changes to the Conservation Reserve Program (CRP), a voluntary cropland retirement program, that occurred around the same time as the 2007 ethanol mandate may have contributed to higher re-enrollment rates in CRP. In particular, the continuous CRP (CCRP) and the temporary extension re-enrollment (REX) contracts which both make re-enrolling in CRP easier and less risky were used extensively in the late 2000s. They found that in areas with ethanol production, CRP re-enrollment was especially high. While these changes could have impacted CRP retention, they likely had little to no impact on enrollment. To enroll land in the CRP program, producers are often placed on waiting lists unless the land is especially sensitive or small. Fields under 15 acres were omitted. These CRP changes therefore would not impact cropland retention. Ifft et al. (2018) also note that the impact of these changes are likely to disproportionately impact areas with high ethanol production. Findings with the standard DID at the 10% radius under a common support are consistent with this. Under the 10% radius which showed a positive effect on cropland conversion from ethanol plant expansion. Since radii under a 10% conversion assumption are larger, land expansions will be further away from established ethanol production areas.

Table 3.13: *Cropland Retention Transition Matching Difference-in-Differences Estimates*

	25% Radius				50% Radius			
Year	ATT Est.	ATT P-Value	95% Conf. Int.		ATT Est.	ATT P-Value	95% Conf. Int.	
2010	0.0020	0.0000	0.0013	0.0027	0.0021	0.0000	0.0012	0.0031
2011	0.0038	0.0000	0.0022	0.0054	0.0032	0.0002	0.0015	0.0049
2012	0.0018	0.0002	0.0009	0.0027	0.0018	0.0047	0.0006	0.0031
2013	–	–	–	–	–	–	–	–
2014	0.0005	0.0455	0.0000	0.0009	–	–	–	–
2015	0.0005	0.0455	0.0000	0.0009	0.0002	0.3174	-0.0002	0.0006

Table 3.14: *Cropland Conversion Matching Difference-in-Differences Estimates*

	25% Radius				50% Radius			
Year	ATT Est.	ATT P-Value	95% Conf. Int.		ATT Est.	ATT P-Value	95% Conf. Int.	
2010	0.0029	0.7741	-0.0169	0.0226	-0.0070	0.0000	-0.0095	-0.0045
2011	-0.0094	0.0000	-0.0133	-0.0056	-0.0086	0.0015	-0.0139	-0.0033
2012	-0.0077	0.0000	-0.0109	-0.0045	–	–	–	–
2013	–	–	–	–	–	–	–	–
2014	-0.0055	0.0005	-0.0086	-0.0024	–	–	–	–
2015	-0.0043	0.0000	-0.0063	-0.0023	-0.0041	0.0027	-0.0068	-0.0014

The ATT values overall were fairly small in magnitude but not inconsequential. Land-use transitions do not occur frequently and therefore even small changes can have long-term implications. Generally around 98% of land in cropland remained cropland in the next year. Therefore, the most retainment could increase by was 2% making a 0.3% change influential to the status quo. Similarly approximately 95% of non-cropland remained in non-cropland in the subsequent year. At the lower bound, the model suggests that the introduction of an ethanol plant lead to a 1.3% reduction in cropland conversion. Caution should be taken with these estimates however.

While the models produced a mixture of expected and unexpected results, the balancing statistics of the matching procedure indicate that the matching did not perform well. Recall that in matching, the goal is to pair treatment observations with control observations that are close in the control variables. In this way, the difference in the outcome variable between the treatment and control group can be attributable to the treatment. If matching is performing well, then the distributions of the matching variables should be similar across the treatment and controls. Tables 3.15 and 3.16 show the years where the matching variables were identical across the groups at the 95% confidence level for cropland retainment and conversion respectively. While nearly all of the matching variables had at least one year where they were statistically identical across groups, few were identical across all years. This indicates that the matching process could be improved if more variables were added. For 2010, matching was considerably poor, balancing only corn basis in the cropland retainment model with a 25% radius. This suggests that more variables need to be added to the set of matching variables. Obvious variables to include are county-level CRP payment indicators used by Towe and Tra (2012) and (Ifft et al., 2018).

Table 3.15: *Cropland Retention Matching Covariate Balanced Years at 95% Confidence Level*

Variable Name	Balanced Year 25% Radius	Balanced Year 50% Radius
Slope	2015, 2014, 2011	2015, 2011
Sand %	2015, 2014	
Silt %	2015, 2011	2015, 2011
Population 100k	2015, 2011	2011
NCCPI	2015, 2011	2015, 2011
Wind Erosion Index	2015, 2014, 2011	2015, 2011, 2010
Hydric Status	2015, 2012, 2011	2015, 2012, 2011
Land Res. Region F	2011	2011
Land Res. Region G		
Land Res. Region H	2011	2011
Land Res. Region J	2012, 2011	2015, 2012
Land Res. Region K	2015, 2014	2015, 2011
Land Res. Region L	2015	
Land Res. Region M	2015	2015, 2011
Land Res. Region N		2012
Land Res. Region O		
Land Res. Region P		
Corn Basis 2004	2011, 2010	2012, 2011, 2010
GDD Mean	2011	
GDD Std. Dev.	2012	2012, 2011
Extreme DD Mean	2012	2012, 2011
Extreme DD Std. Dev.	2012	2011
Precip. Mean	2011	2015, 2011
Precip. Std. Dev	2014, 2011	2015, 2011

Table 3.16: *Cropland Conversion Matching Covariate Balanced Years at 95% Confidence Level*

Variable Name	Balanced Year	
	25% Radius	50% Radius
Slope	2015, 2014, 2012, 2011, 2010	2015, 2011, 2010
Sand %	2014	2015, 2011, 2010
Silt %	2015, 2014	2015, 2011
Population 100k	2015, 2014, 2012, 2010	2010
NCCPI	2015, 2014, 2011	2010
Wind Erosion Index	2014, 2011, 2010	2015, 2010
Hydric Status	2015, 2012, 2011, 2010	2015, 2011, 2010
Land Res. Region F	2015	2015
Land Res. Region G		
Land Res. Region H	2015, 2014, 2012	2011
Land Res. Region J		2015, 2011
Land Res. Region K	2015	
Land Res. Region L	2015	
Land Res. Region M	2015, 2014, 2012	2015, 2011, 2010
Land Res. Region N	2012	
Land Res. Region O	2015	2015, 2011, 2010
Land Res. Region P		
Corn Basis 2004	2014, 2011, 2010	2015, 2011
GDD Mean	2014, 2011	2015, 2011
GDD Std. Dev.	2014, 2012, 2011	2015, 2011
Extreme DD Mean	2014, 2012, 2011	2015, 2011
Extreme DD Std. Dev.	2014, 2012, 2011	2015, 2011
Precip. Mean	2015, 2014, 2011	2015, 2011, 2010
Precip. Std. Dev	2015, 2014, 2012, 2011	2015, 2011, 2010

3.6 Conclusions

Land transitions at the extensive margin have widespread implications for environmental and efficiency outcomes in agriculture. In this chapter, I use broad land use decisions across nearly 5 million fields in 13 states and how they were impacted by the introduction and expansion of ethanol plants. Using the difference-in-differences (DID) approach, event studies, and a matching DID, I modeled cropland retainment and conversions over four different treatment assignments. The results suggest that fields in areas with ethanol plant expansion were more likely to retain land in cropland after the expansion. However, there was also evidence that ethanol plant expansions lead to a reduction in cropland conversion. Changes to the CRP program that encouraged re-enrollment around the same time the RFS ethanol mandates came into effect may explain these contradictory results. In addition, the balance statistics for the matching variables across treatment and control groups indicate that the matching could be improved by adding more variables. If the impact of CRP program changes is the driver of the results, this suggests that county-level CRP payment information could be added as parameters in future iterations of this study.

There are several extensions that could be made to improve future analysis. Firstly other variables can be introduced into the model. Two feasible variables are the field's distance from railway lines and the field's distance from major metropolitan areas. The distance from railway lines has been identified as a potential control for distance from an ethanol plant. Motamed et al. (2016) used this variable as a instrumental variable arguing that the field's distance from an ethanol plant has little bearing on contemporary crop decisions and is positively correlated with the proximity of ethanol plants. Towe and Tra (2012) used the distance from railway lines as a matching parameter for similar reasons. Towe and Tra (2012) also use distance to metropolitan areas as another control. This control correlates with distance to markets and therefore transportation costs. In this way the distance to metropolitan areas can be viewed as both a control and for ethanol plant location. Since transportation costs are not equal across all land uses. For example heavy logging from forestry use has unique transportation need that conventional crop production does not (Lubowski et al., 2006).

Other extensions are methodological. The economics community is becoming more interested in utilizing machine learning in performing analysis over "big" datasets (Varian, 2014). Since inference is the goal of most of economic analysis, the use of machine learning is complicated in economic research. However, there is a growing interest in using machine learning in intermediate steps of analysis where the prediction and classification capabilities of these methods can be exploited without compromising theoretical grounding of the primary model. Matching procedures is one area where machine learning, particularly classification trees and their variations hold promise (Athey and Imbens, 2016; Lee et al., 2010). While expansive, much of the data went unused due to computational issues. While computational time can still be problematic under large datasets, classification trees are often less computationally burdensome than models utilizing maximum likelihood and come with the added benefit of being able to incorporate highly non-linear control relationships.

Bibliography

Gaurav Arora, Peter T Wolter, Hongli Feng, and David A Hennessy. Role of ethanol plants in dakotas land use change: Incorporating flexible trends in the difference-in-difference framework with remotely-sensed data. *Center for Agriculture and Rural Development Working Paper*, 2016.

Susan Athey and Guido Imbens. The state of applied econometrics-causality and policy evaluation. *arXiv preprint arXiv:1607.00699*, 2016.

Susan Athey, Julie Tibshirani, and Stefan Wager. Generalized random forests. *arXiv preprint arXiv:1610.01271*, 2016.

Raleigh Barlowe. Land resource economic: The economics of real property. *Auflage, New Jersey (Englewood Cliffs)*, 1972.

Ellis C. Benham, Robert J. Ahrens, and W. Dennis Nettleton. Clarification of soil texture class boundaries. MO5 Soil Technical Note 16, National Soil Survey Center, USDA-NRCS, March 2009.

Richard Blundell and Thomas M Stoker. Heterogeneity and aggregation. *Journal of Economic Literature*, 43(2):347–391, 2005.

J Christopher Brown, Eric Hanley, Jason Bergtold, Marcelus Caldas, Vijay Barve, Dana Peterson, Ryan Callihan, Jane Gibson, Benjamin Gray, Nathan Hendricks, et al. Ethanol plant location and intensification vs.

extensification of corn cropping in kansas. *Applied Geography*, 53:141–148, 2014.

Jesslyn F Brown and Md Shahriar Pervez. Merging remote sensing data and national agricultural statistics to model change in irrigated agriculture. *Agricultural Systems*, 127:28–40, 2014.

A Colin Cameron and Douglas L Miller. A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2):317–372, 2015.

A Colin Cameron, Jonah B Gelbach, and Douglas L Miller. Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3):414–427, 2008.

Fernando Carriazo, Roger Claassen, Joe Cooper, Daniel Hellerstein, Kohei Ueda, et al. Grassland to cropland conversion in the northern plains: The role of markets and policy. In *2010 Annual Meeting, July 25-27, 2010, Denver, Colorado*, number 61625. Agricultural and Applied Economics Association, 2010.

Colin A Carter, Gordon C Rausser, and Aaron Smith. Commodity storage and the market effects of biofuel policies. *American Journal of Agricultural Economics*, page aaw010, 2016.

Jean-Paul Chavas and Matthew T Holt. Acreage decisions under risk: The case of corn and soybeans. *American Journal of Agricultural Economics*, 72(3):529–538, 1990.

Jean-Paul Chavas, Matthew T Holt, et al. Economic behavior under uncertainty: A joint analysis of risk preferences and technology. *Review of economics and Statistics*, 78(2):329–335, 1996.

Roger Claassen, Christian Langpap, and JunJie Wu. Impacts of federal crop insurance on land use and environmental quality. *American Journal of Agricultural Economics*, 99(3):592–613, 2017.

R Dobos, H Sinclair, and K Hipple. User guide national commodity crop productivity index (nccpi) version 1.0. *US Department of Agriculture, Natural Resources Conservation Service*, 2008.

Dave Donaldson and Adam Storeygard. The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives*, 30(4): 171–98, 2016.

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

US EPA. Regulation of fuels and fuel addiatives: Changes to renewable fuel standard program; final rule. *Fed Regist*, 75(58):14682, 2010.

Dale Eddy Farnham. Corn planting guide. 2001.

Bruce L Gardner. Futures prices in supply analysis. *American Journal of Agricultural Economics*, 58(1):81–84, 1976.

Mekbib G Haile, Matthias Kalkuhl, and Joachim Braun. Inter-and intra-seasonal crop acreage response to international food prices and implications of volatility. *Agricultural Economics*, 45(6):693–710, 2014.

Mekbib G Haile, Matthias Kalkuhl, and Joachim von Braun. Worldwide acreage and yield response to international price change and volatility: A dynamic panel data analysis for wheat, rice, corn, and soybeans. In *Food Price Volatility and Its Implications for Food Security and Policy*, pages 139–165. Springer, 2016.

James J Heckman, Hidehiko Ichimura, and Petra E Todd. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies*, 64(4):605–654, 1997.

Nathan P Hendricks, Joseph P Janzen, and Aaron Smith. Futures prices in supply analysis: Are instrumental variables necessary? *American Journal of Agricultural Economics*, 97(1):22–39, 2014a.

Nathan P Hendricks, Sumathy Sinnathamby, Kyle Douglas-Mankin, Aaron Smith, Daniel A Sumner, and Dietrich H Earnhart. The environmental effects of crop price increases: Nitrogen losses in the us corn belt. *Journal of Environmental Economics and Management*, 68(3):507–526, 2014b.

Nathan P Hendricks, Aaron Smith, and Daniel A Sumner. Crop supply dynamics and the illusion of partial adjustment. *American Journal of Agricultural Economics*, page aau024, 2014c.

David A Hennessy. On monoculture and the structure of crop rotations. *American Journal of Agricultural Economics*, 88(4):900–914, 2006.

Matthew T Holt. A linear approximate acreage allocation model. *Journal of Agricultural and Resource Economics*, pages 383–397, 1999.

Haixiao Huang, Madhu Khanna, et al. An econometric analysis of us crop yield and cropland acreage: implications for the impact of climate change. In *AAEA annual meeting, Denver, Colorado*, pages 25–27, 2010.

Jennifer Ifft, Deepak Rajagopal, and Ryan Weldzuis. Ethanol plant location and land use: A case study of crp and the ethanol mandate. *Applied Economic Perspectives and Policy*, 2018.

Samuel Jackson. Ethanol: A primer. *University of Tennessee Ag Research Extension*, SP700(B), 2018.

Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning*, volume 112. Springer, 2013.

Hyunseok Kim and GianCarlo Moschini. The dynamics of supply: Us corn and soybeans in the biofuel era. *Land Economics*, 84(4):593–613, Nov 2018.

Barrett E Kirwan and Michael J Roberts. Who really benefits from agricultural subsidies? evidence from field-level data. *American Journal of Agricultural Economics*, 98(4):1095–1113, 2016.

Patrick Kline and Andres Santos. A score based approach to wild bootstrap inference. *Journal of Econometric Methods*, 1(1):23–41, 2012.

Obafèmi Philippe Koutchadé, Alain Carpentier, and Fabienne Femenia. Modeling heterogeneous farm responses to european union biofuel support with a random parameter multicrop model. *American Journal of Agricultural Economics*, 2018.

Anne Lacroix and Alban Thomas. Estimating the environmental impact of land and production decisions with multivariate selection rules and panel data. *American Journal of Agricultural Economics*, 93(3):784–802, 2011.

Gabriel E Lade, CY Cynthia Lin Lawell, and Aaron Smith. Designing climate policy: Lessons from the renewable fuel standard and the blend wall. *American Journal of Agricultural Economics*, 100(2):585–599, 2018.

Christian Langpap and JunJie Wu. Potential environmental impacts of increased reliance on corn-based bioenergy. *Environmental and Resource Economics*, 49(2):147–171, 2011.

Tyler J Lark, J Meghan Salmon, and Holly K Gibbs. Cropland expansion outpaces agricultural and biofuel policies in the united states. *Environmental Research Letters*, 10(4):044003, 2015.

Tyler J Lark, Richard M Mueller, David M Johnson, and Holly K Gibbs. Measuring land-use and land-cover change using the us department of agriculture's cropland data layer: Cautions and recommendations. *International Journal of Applied Earth Observation and Geoinformation*, 62: 224–235, 2017.

Brian K Lee, Justin Lessler, and Elizabeth A Stuart. Improving propensity score weighting using machine learning. *Statistics in Medicine*, 29(3):337–346, 2010.

David R Lee and Peter G Helmberger. Estimating supply response in the presence of farm programs. *American Journal of Agricultural Economics*, 67(2):193–203, 1985.

William Lin and Robert Dismukes. Supply response under risk: Implications for counter-cyclical payments' production impact. *Review of Agricultural Economics*, 29(1):64–86, 2007.

Ruben N Lubowski, Shawn Bucholtz, Roger Claassen, Michael J Roberts, Joseph C Cooper, Anna Gueorguieva, and Robert Johansson. Environmental effects of agricultural land-use change. *Economic Research Report*, 25:1–75, 2006.

Ruben N Lubowski, Andrew J Plantinga, and Robert N Stavins. What drives land-use change in the united states? a national analysis of landowner decisions. *Land Economics*, 84(4):529–550, 2008.

Bruce MacKellar and Eric Anderson. Planting corn in wet conditions – is it worth it?, May 2016. URL http://msue.anr.msu.edu/news/planting_corn_in_wet_conditions_is_it_worth_it.

Enno Mammen. Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics*, pages 255–285, 1993.

Kevin McNew and Duane Griffith. Measuring the impact of ethanol plants on local grain prices. *Review of Agricultural Economics*, 27(2):164–180, 2005.

Ruiqing Miao, Madhu Khanna, and Haixiao Huang. Responsiveness of crop yield and acreage to prices and climate. *American Journal of Agricultural Economics*, 98(1):191–211, 2016.

Mesbah Motamed, Lihong McPhail, and Ryan Williams. Corn area response to local ethanol markets in the united states: A grid cell level analysis. *American Journal of Agricultural Economics*, page aav095, 2016.

NASS, USDA. Field crops: Usual planting and harvesting dates. *USDA National Agricultural Statistics Service, Agriculural Handbook*, 628, 2010.

Natural Resources Conservation Service. National resources inventory: Background, January 2001.

Marc Nerlove. Estimates of the elasticities of supply of selected agricultural commodities. *Journal of Farm Economics*, 38(2):496–509, 1956.

NRCS. Hydric soils - overview, October 2018. URL https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/use/hydric/?cid=nrcs142p2_053985.

USDA NRCS. Land resource regions and major land resource areas of the united states, the caribbean, and the pacific basin. *US Department of Agriculture Handbook*, 296, 2006.

Navin Ramankutty and Jonathan A Foley. Estimating historical changes in global land cover: Croplands from 1700 to 1992. *Global Biogeochemical Cycles*, 13(4):997–1027, 1999.

David Ricardo. *Principles of Political Economy and Taxation*. G. Bell, 1891.

Michael J Roberts and Wolfram Schlenker. Identifying supply and demand elasticities of agricultural commodities: Implications for the us ethanol mandate. *The American Economic Review*, 103(6):2265–2295, 2013.

Timothy Searchinger, Ralph Heimlich, Richard A Houghton, Fengxia Dong, Amani Elobeid, Jacinto Fabiosa, Simla Tokgoz, Dermot Hayes, and Tun-Hsiang Yu. Use of us croplands for biofuels increases greenhouse gases through emissions from land-use change. *Science*, 319(5867):1238–1240, 2008.

Hosein Shapouri and Paul Gallagher. Usda's 2002 ethanol cost-of-production survey. 2005.

Jeffrey A Smith and Petra E Todd. Does matching overcome lalonde's critique of nonexperimental estimators? *Journal of Econometrics*, 125(1-2): 305–353, 2005.

Soil Survey Staff. *Keys to Soil Taxonomy*. USDA-Natural Resources Conservation Service, Washington, DC., 12th edition, 2014.

Loyd R Stone and Alan J Schlegel. Crop water use in limited-irrigation environments. In *Proceedings of the 2006 Central Plains Irrigation Conference, Colby, KS*, pages 21–23. Citeseer, 2006.

Charles Towe and Constant I Tra. Vegetable spirits and energy policy. *American Journal of Agricultural Economics*, 95(1):1–16, 2012.

Hal R Varian. Big data: New tricks for econometrics. *The Journal of Economic Perspectives*, 28(2):3–27, 2014.

Ekaterina Vorotnikova, Serhat Asci, and James L Seale Jr. Effect of relative price changes of top principle crops on farm land allocation in post-soviet russia: Do prices matter. *Agricultural and Applied Economics Association*, 2014.

Sun Ling Wang, Paul Heisey, David Schimmelpfennig, and V Eldon Ball. Agricultural productivity growth in the united states: Measurement, trends, and drivers. *Economic Research Service, Paper No. Err-189*, 2015.

Tong Wang, Moses Luri, Larry Janssen, David A Hennessy, Hongli Feng, Michael C Wimberly, and Gaurav Arora. Determinants of motives for land use decisions at the margins of the corn belt. *Ecological Economics*, 134:227–237, 2017.

Matthew D Webb. Reworking wild bootstrap based inference for clustered errors. Technical report, Queen's Economics Department Working Paper, 2013.

James K Whittaker and Robert L Bancroft. Corn acreage response-function estimation with pooled time-series and cross-sectional data. *American Journal of Agricultural Economics*, 61(3):551–553, 1979.

Joshua Woodard. Big data and ag-analytics: An open source, open data platform for agricultural & environmental finance, insurance, and risk. *Agricultural Finance Review*, 76(1):15–26, 2016a.

Joshua D Woodard. Data science and management for large scale empirical applications in agricultural and applied economics research. *Applied Economic Perspectives and Policy*, 38(3):373–388, 2016b.

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

JunJie Wu, Richard M Adams, Catherine L Kling, and Katsuya Tanaka. From Microlevel Decisions to Landscape Changes: An Assessment of Agricultural Conservation Policies. *American Journal of Agricultural Economics*, 86(1):26–41, 2004a. doi: 10.1111/j.0092-5853.2004.00560.x. URL http://ajae.oxfordjournals.org/content/86/1/26.abstract.

JunJie Wu, Richard M Adams, Catherine L Kling, and Katsuya Tanaka. From microlevel decisions to landscape changes: An assessment of agricultural conservation policies. *American Journal of Agricultural Economics*, 86(1):26–41, 2004b.

L Yan and DP Roy. Conterminous united states crop field size quantification from multi-temporal landsat data. *Remote Sensing of Environment*, 172: 67–86, 2016.