Title
Essays on the Impacts of Climatic Events on Farmers' Decisions, Farm Profits, and Agricultural Production

Author
Lee, Seunghyun

Publication Date
2022

Essays on the Impacts of Climatic Events on Farmers' Decisions, Farm Profits, and Agricultural Production

By

SEUNGHYUN LEE
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Agricultural and Resource Economics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

____

Aaron Smith (Chair)

____

Julian Alston

____

Frances Moore

Committee in Charge

2022

Abstract

Agricultural production around the world has undergone tremendous changes over the past century. However, two features remain salient in modern agricultural production. One is that global crop production depends heavily on a few major crop-producing regions, and the other is that climatic conditions still play a vital role in determining crop production in these regions. Not surprisingly, anthropogenic global warming has been considered one of the greatest threats to feeding the world's growing populations. This dissertation studies the impacts of climatic events on agriculture in three aspects: farmers' decisions in the first chapter, farm profits in the second chapter, and agricultural production in the third chapter.

In the first chapter, I ask the following question: can we find statistical evidence that farmers in the US Corn Belt show *recency effects* associated with local yield shocks? As climate change is projected to increase extreme heat events in frequency and severity, farmers are expected to see an increase in crop yield variability. Using field-level crop-choice data in the major US Corn Belt states (Iowa, Illinois, and Indiana), this chapter uncovers statistical evidence that crop-choice decisions in the region feature *recency effects* associated with local yield shocks largely driven by plausibly random weather. In the region, short-run acreage adjustment occurs mainly along the intensive margin (transition between corn and soybeans). I show that farmers are less likely to plant corn—which is more susceptible to heat/water stress than soybeans—after a hotter or drier than average year, irrespective of the within-season timing of the heat. This means that low yields of corn or soybeans in one year predict less corn being planted in the subsequent year. Interestingly, relative yield does not predict the probability of growing corn next year. Based on the insights from my conceptual model, these empirical results suggest that farmers respond sensitively to total exposure to extreme heat during the previous growing season but do not respond to when it was concentrated (e.g., 1st half or 2nd half of the growing season) or which crop it affected most.

In the second chapter, I ask the following question: what are the economic costs of large-scale droughts to crop producers? Extreme weather events, such as heatwaves and droughts, in a major crop-producing area decrease crop yields but tend to increase crop prices. Such a negative correlation makes it difficult to quantify net crop revenue impacts. This paper proposes a panel approach to estimating the impacts of extreme weather events in major crop-producing regions on crop revenues accounting for the correlation between crop price and yield in the context of US crop production. I first show that, under some conditions, weather-induced changes in crop revenues are identical to those of crop profits, for which data are scant. To

estimate weather-induced crop revenue effects, I use a national-level yield shock as an explanatory variable in addition to local weather variables. This variable bears two appealing features. First, it can account for heterogeneous crop yield responses to weather across regions. Second, it permits coefficients that can be interpreted as the conditional price flexibilities of demand. I show that, when estimating crop revenue impacts of weather, it is important to additionally account for spatially varying degrees of the correlation between local and aggregate yield shocks, temporally varying price flexibilities, and spatially heterogeneous yield response to weather. I apply this approach to the 1988 and 2012 US Midwest droughts to quantify the impacts of the droughts on crop revenues across US counties. I estimate that crop revenue was impacted by (-)11% for corn and (+)1% for soybeans in 1988, and (+)11% for corn and 0% in 2012. I also document that, in the two years, regional inequality of crop revenues substantially deteriorated.

In the third chapter, I work with John Abatzoglou to answer the following question: how does planting-season weather affect agricultural production? Record-high prevented planting of staple crops such as corn and soybeans in the US Corn Belt due to heavy rainfall in recent years has spurred the concern over food security, as growing evidence suggests winter and spring precipitation extremes will occur more frequently in the upper US Corn Belt in the coming decades. We examine within-season time-varying effects of planting-season water balance—precipitation minus reference evapotranspiration—on prevented planting of corn and soybeans in the US Corn Belt. Our results show significant impacts of excess moisture on preventing planting and suggest a 58-176% increase in prevented planting during the months of April-June per standard deviation increase in water balance. This framework is additionally used alongside downscaled climate change projections to estimate future changes in county-level prevented planting during the mid-century (2036–2065) under the moderate emission scenario (RCP 4.5). We find that prevented planting will increase in parts of Iowa, Minnesota, and Wisconsin by 0-30% and generally decrease in the other parts of the US Corn Belt. This work highlights the value of incorporating water balance data in assessing prevented-planting impacts and is the first known study to examine changing risk of prevented planting under future climate scenarios that may help inform adaptation efforts to avoid losses.

Acknowledgments

I would like to express my deepest gratitude to my advisor, Professor Aaron Smith, for his guidance on this dissertation. Aaron has been not only academically influential and helpful but also supportive and patient in many ways. I was fortunate to meet him and work with him on many interesting projects. I am grateful to him for giving me all the opportunities to develop my interest and expertise. My Ph.D. journey wouldn't have even started as I did, and have not been shaped this way without his unconditional trust and support to me. I also thank Professor Julian Alston and Professor Frances Moore for giving me thoughtful and careful feedback on my dissertation. I would also like to thank Professor Steve Boucher for his guidance on my dissertation proposal. I would also like to give a special thanks to Professor John Abazoglou at UC Merced for collaborating with me on the project on the effects of planting-season weather on prevented planting, which is included in the third chapter of this dissertation. Including an interdisciplinary research project in my dissertation has been one of my long-held wishes and he made it possible. I wish to continue to work with him on other projects that address important issues in agriculture challenged by climate change.

I owe a lot to many professors who taught me at UC Davis during my Ph.D. program. I also thank many friends in our department for making my Ph.D. journey full of fun, excitement, encouragement, motivation, and relaxation. I am a little saddened that Covid-19 did not allow us to spend much time in person as we did in the first half of my Ph.D. program. I have found our department very helpful and inspiring. I have constantly felt that I am very lucky to be surrounded by people working hard on many issues that I think are important and contributing to advancing our society. I thank you to all people in our department for making such a great environment.

I thank my family in South Korea for their love and support. Lastly, I thank my beautiful wife, Goeun, and my lovely daughter, Hannah, for making my life full of joy and love. Goeun, it's now your turn!

Contents

List of Figures

List of Tables

CHAPTER 1

Recency Effects of Yield Shocks on Crop Choice in the US Corn Belt

1.1. Introduction

Numerous studies in the behavioral science identified a variety of biases and heuristics in people's probabilistic decision-making processes (Thaler 2016; Tversky and Kahneman 1973). Such biases and heuristics appear even when professionals make decisions relevant to their expertise (Amir and Ganzach 1998; Nofsinger and Varma 2013; Singh 2021). One of the cognitive biases is *recency bias*. This cognitive bias arises when agents make decisions by weighing recent information as more relevant than older information. Such bias is prevalent in part because people tend to think that the next outcome will be similar to those preceding it (i.e., strong belief in serial correlation of events).

This paper seeks to answer the following question: can we find statistical evidence that farmers in the US Corn Belt show *recency effects* associated with plausibly random yield shocks when they make crop choice decisions? Understanding how weather-driven shocks affect farmers' crop choice is important particularly in a changing climate. As extreme heat events are projected to increase in frequency and severity in the coming decades, farmers are expected to experience more variable crop yields (Collins et al. 2013). Because farmers' crop choice has direct implications for food security and water quality from land-use change, *recency effects* on crop choice mean that past yield shocks could have lingering impacts on food production and local environment in subsequent years (Hendricks et al. 2014a; Lark et al. 2022; Metaxoglou and Smith 2020; Metaxogolou and Smith 2022). This would also mean that policy concerning cropland management should be more carefully designed in a changing climate.

To answer my question, I use field-level crop choice data from 2004 to 2020 in Iowa, Illinois, and Indiana along with the conditional grouped coefficient estimator proposed by Hendricks et al. (2014b). In the study region, corn and soybeans are the two major crops, and acreage adjustment occurs mainly through transition between corn and soybeans (i.e., intensive margin). My results suggest that farmers' previous-year experiences of weather and yield shocks do affect their crop choice decisions in the following year.

This behavioral pattern is surprising given that, in the study region, growing-season weather is serially uncorrelated and does not appear to provide useful information about weather for the subsequent growing season. These findings are consistent with previous studies in the field of behavioral science suggesting that recent experiences play greater roles in people's decisions than objective probabilities (e.g., Durand et al. 2021; Fredrickson and Kahneman 1993; Gallagher 2014; Hertwig et al. 2004; Singh 2021). To be more specific, I show that farmers are less likely to plant corn——which is more susceptible to heat/water stress than soybeans——after a hotter-or-drier-than-average year, irrespective of the timing of extreme heat. This means that low yields of either corn or soybeans predicts less corn planted next year, but relative yield does not predict the probability of growing corn next year. I show that these results are not explained by the conventional "market" expected crop revenues, or perfect-foresight crop yield or revenue shocks.

The US Corn Belt provides an ideal setting for testing *recency effects* on crop choice. First, farming practices have been highly homogeneous across fields in the study area in the sense that most fields grow corn and soybeans in rotation. Second, because of limited irrigation, crop yields depend heavily on weather conditions in the study area. Third, growing-season weather in the US Corn Belt has been highly random from year to year. Not surprisingly, past weather has little predictive power of future crop yields, which I statistically show in the paper.

Besides inter-annual randomness of weather, *recency effects* on crop choice in the US Corn Belt are somewhat surprising considering the fact that farmers in the region typically make highly calculated planting decisions from checking potential rotation benefits and local crop prices to soil conditions. Crop choice is arguably the most important decision to make in a crop year, among others including seed density and fertilization. Mainly because of lack of irrigation in most parts of the area, crop yield is largely at the mercy of weather leaving not much room for farmers to affect yields once the crop has been planted.

Focusing on the intensive margin, I use a binary crop choice (corn or soybeans) as the outcome of interest. The main explanatory variables of interest are two types of (lagged) county-level yield shocks: 1) relative yield shock, measured by yield shock for corn relative to that for soybeans and 2) average yield shock (weighted by acreage across corn and soybeans). I also consider (lagged) crop-specific yield shocks and (lagged) total exposure to extreme heat. I control for contemporaneous-year crop-specific i) planting-time market expectations of crop revenues and ii) harvest-time realized crop revenue shocks.

In the conceptual framework, I develop testable hypotheses that allow me to use coefficients on various forms of (lagged) yield shocks to infer to which weather signal farmers sensitively respond: *total exposure* to extreme heat or *timing* of extreme heat (i.e., during which growing-season phase it is concentrated). To do so, I conceptualize crop yield response to weather focusing on extreme heat, which the literature identified as the strongest weather predictor of crop yields in the US Corn Belt (Ortiz-Bobea et al. 2019; Schlenker and Roberts 2009). I utilize two stylized facts: i) corn and soybeans are different in terms of when extreme heat is particularly damaging and ii) extreme heat tends to peak around the phase when extreme heat is particularly damaging to corn. The conceptual model suggests that average (and crop-specific) yield shocks better predict the probability of planting corn than relative yield shocks do if farmers sensitively respond to *total exposure* to extreme heat but little to *timing* of extreme heat, and vice versa.

Econometrically, I use the conditional grouped coefficients approach proposed by Hendricks et al. (2014b). This approach allows me to account for the dynamic crop choice incentives coming from crop rotation effects and to mitigate potential bias arising from parameter heterogeneity in the short dynamic panel. Following Hendricks et al. (2014b), I estimate separate econometric models for regions with different soil and climate characteristics to alleviate the potential bias induced by coefficient heterogeneity in the dynamic panel with small T. The major difference of this work from Hendricks et al. (2014b) is the explanatory variables of interest. While Hendricks et al. (2014b) focused on estimating the acreage response to expected prices, I focus on estimating a potential *recency effect* of local yield shocks on crop choice.

I find that, conditional on corn having been planted in the prior year, a 10 percentage points (pp.) increase in the average yield shock in the previous year predicts a 1.5 pp. increase in the probability of corn being planted. The magnitude is much smaller (0.5 pp.) for fields that were planted to soybeans in the prior year due to asymmetric yield penalties associated with continuous cropping. Interestingly, relative yield does not affect farmers' crop choice decisions. Based on the insights from the conceptual model, my empirical findings that crop-specific and average yield shocks better predict the probability of corn than relative yield shocks suggest that farmers sensitively respond to the total exposure to extreme heat during the previous growing season but reacts little to when it was concentrated (e.g., 1st half or 2nd half of the growing season).

This paper builds on and integrates multiple strands of literature. A body of studies in a variety of disciplines have documented that a decision-maker's recent experiences play an important role in their decisions and beliefs (e.g., Beltrán et al. 2018; Gallagher 2014; Nofsinger and Varma 2013). Using a panel dataset of large regional floods and flood insurance policies in the US, Gallagher (2014) shows that insurance take-up substantially increases the year after a flood and then steadily drops off. Leveraging electronic health records spanning 86,000 deliveries, Singh (2021) finds that, when physicians choose a delivery decision between vaginal and cesarean delivery, they tend to switch to the other delivery mode for the subsequent patient—regardless of patient indications—if the prior patient had complications in one delivery mode.

Using a controlled experimental setting where participants make crop choice decisions after observing a sequence of drought conditions, Demnitz and Joslyn (2020) find that people tend to make overly cautious crop choices when droughts occurred in the recent sequences. They also find that providing correct information prior to decision-making helps reduce over-cautious crop choice. Moore et al. (2019) find that people appear to form a belief about normal weather conditions based on weather experienced between 2 to 8 years ago, which is a much shorter time window than scientists often use to define climate or normal weather (e.g., 30 years). There is a nascent literature on recency effects in the field of agricultural economics, but these studies have been focused on crop insurance (Che et al. 2020; Fezzi et al. 2021). My paper is the first known work to examine how farmers' recent experiences of weather-induced yield shocks affect their acreage allocation decisions at the intensive margin in the US Corn Belt.

My work also broadly speaks to the literature on the impacts of climate change on agriculture. Many studies have used crop choice in their conceptual framework as a textbook adaptation strategy to climate change (Burke and Emerick 2016; Deschênes and Greenstone 2007; Mendelsohn et al. 1994). However, few studies used crop choice as their main outcome variable of interest in their empirical analysis. Some exceptions are Cui (2020) and Ramsey et al. (2021). Cui (2020) uses US county data on planted acres aggregating corn and soybeans to study how acreage responds to long-run average temperatures. As his focus is on acreage response to gradual temperature changes, his analysis is on the extensive margin whereas I focus on the intensive margin. Ramsey et al. (2021) studies how short-run past local weather trend affects farmers' crop choice in eleven counties in Kansas, using a dynamic multinomial probit model and field-level data. As mentioned in their paper, biophysical environment in the US Corn Belt is quite different from that in the western Kansas. Importantly, given that the US Corn Belt is a major producer and exporter of corn and soybeans, crop production in my study region has more significant implications for global food security.

On top of that, my conceptual framework contributes to the land-use literature by showing that an appropriate consideration of agronomic and climatic contexts can be essential to understanding farmers' decisions. My model illustrates that farmers' acreage allocation could be highly nuanced and even counter-intuitive without considering local agroclimatic contexts. This is an important point given that, despite a wide availability of fine-resolution geospatial data, it is still costly to obtain individual-level data that would allow us to investigate detailed mechanisms of agents' belief formation.

1.2. Background

In this section, I discuss salient agronomic and climatic features that I make use of when developing my context-based conceptual model in the next section. In the study region, the most active planting season is May. Although planting seasons for the two crops often overlap, corn tends to be planted a few weeks earlier than soybeans. Different active planting time windows are driven by the fact that two crops are different in terms of ideal planting conditions and potential yield penalties from missing them.

Fact.1 *Extreme heat in July (silking period for corn) is particularly damaging to corn, whereas extreme heat in August (podding period for soybeans) is particularly damaging to soybeans.*
Figure 1.1 shows within-season time-varying crop yield effects of extreme heat, measured by degree days above $30°C$ from May to September.[1] (See appendix for details of estimation.) Extreme heat is particularly damaging to corn around July (silking period) and to soybeans around August (podding period). The time-varying effects of extreme heat on crop yields in the figure are in line with previous studies (Berry et al. 2014; Ortiz-Bobea et al. 2019; Zipper et al. 2016).

Fact.2 *Extreme heat tends to peak when it is particularly damaging to corn.*
The bottom plot in figure 1.1 indicates that July and August comprise the growing-season period that most differentiates corn and soybean yields. Red bins under each subplot show the distribution of extreme heat over the typical growing season in the study area. It tends to peak in early to mid July when extreme heat can have detrimental effects on corn but not as much on soybeans.

[1] April to September is the growing season commonly used in the literature on statistical crop yield modelling. When extreme heat is the main weather variable of interest, ignoring the active planting season would not be a problem because of negligible exposure to extreme heat.

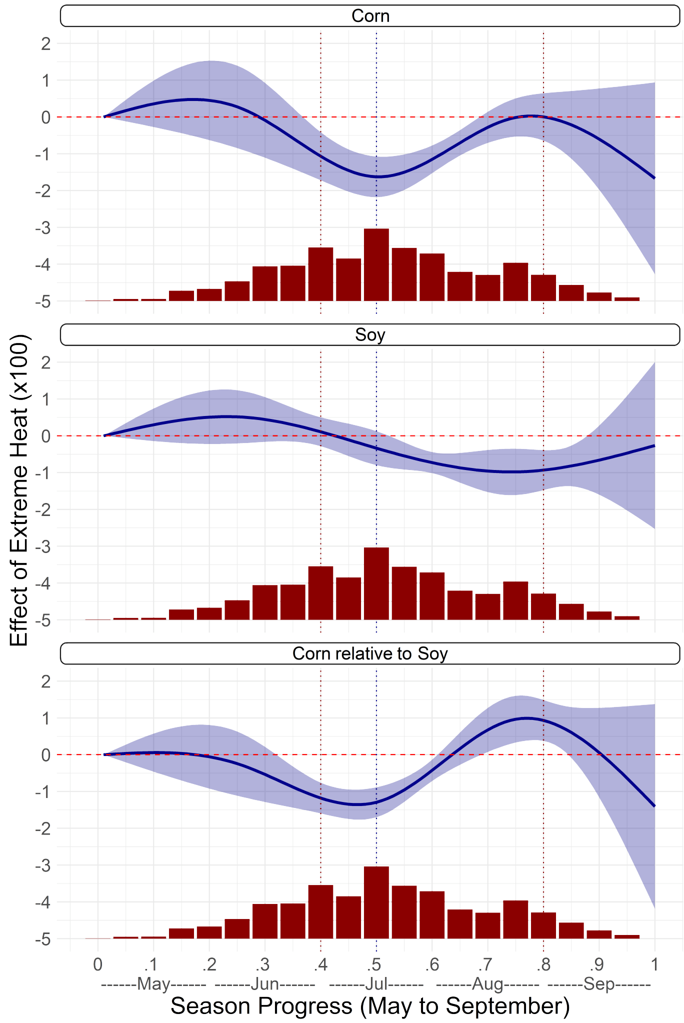


FIGURE 1.1. Time-Varying Effects of Exposure to Temperature above 30°C over the Growing Season

Note: The blue lines represent within-season marginal effects of exposure to temperature above 30°C on crop yield shocks in percentage points. For *Corn* and *Soy*, the dependent variables were constructed by the ratio of realized yield to linear trend yields. For *Corn relative to Soy*, I divided the ratio for *Corn* by that for *Soy*. The shaded areas show the 95% confidence intervals around the marginal effects. Standard errors were clustered by year. Red bins represent the within-season distribution of exposure to extreme heat during the typical growing season in the study area.

Growing-season weather in 2003 and 2012 in Iowa illustrates the importance of such heterogeneous time-varying extreme-heat effects between the two crops. In the two years, Iowa counties saw a hotter-than-usual summer with a different timing of when extreme heat was concentrated. Extreme heat was concentrated in August for 2003 but it was in July for 2012 as seen in figure 1.2. The dashed lines represent

10th, 50th, and 90th percentiles of the historical distribution of monthly degree days above $30°C$. The different concentration phases led to different patterns of relative yield shocks in the two years. As can be seen in figure 1.3, corn yield was much higher than soybean yield in 2003, but the opposite was the case in 2012.

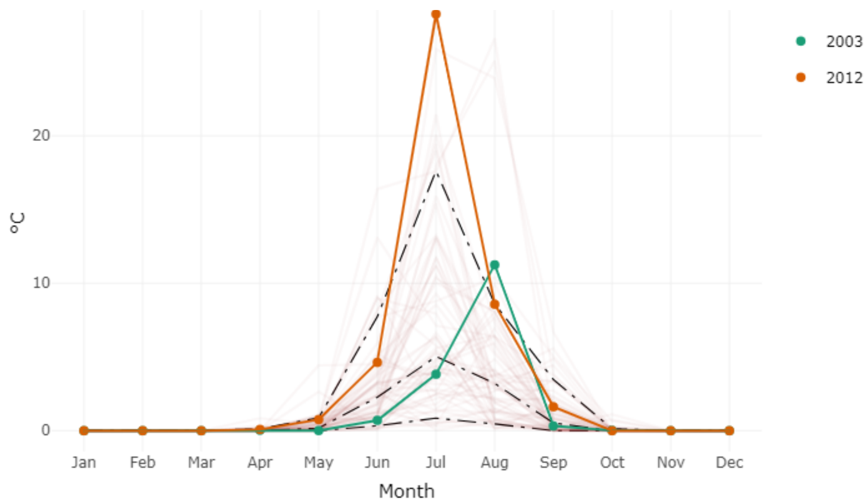


FIGURE 1.2. Monthly Exposure to Temperature above $30°C$ in Iowa
Source: www.asmith.ucdavis.edu/data/weather

Note: Thick green and orange lines show monthly total exposure to temperature above $30°C$ in an average county of Iowa in 2003 and 2012. Black dashed lines represent the 10th, 50th, and 90th percentiles of the historical distribution of monthly total exposure to temperature above $30°C$ in an average county of Iowa.

Fact.3 *Soybeans can adapt to water stress better than corn.*

Putting aside the heterogeneous time-varying effects of extreme heat, soybeans can better adapt in times of water stress due to its internal self-protective mechanism against water stress (Rippey 2015). During droughts, soybeans can continue to grow while adjusting its growth pace and compensate for water deficit once water becomes available. This means that soybeans can endure water stress over a longer period and range of growth stages. Back in 2012, a substantial amount of corn in Illinois died when extreme heat spiked in July, but soybeans slowed down its growth by shutting down their pods to conserve water. When rains returned in mid-August, soybean plants rebounded by filling out remaining pods that survived. As a result, soybean yields were not too far below the normal level.

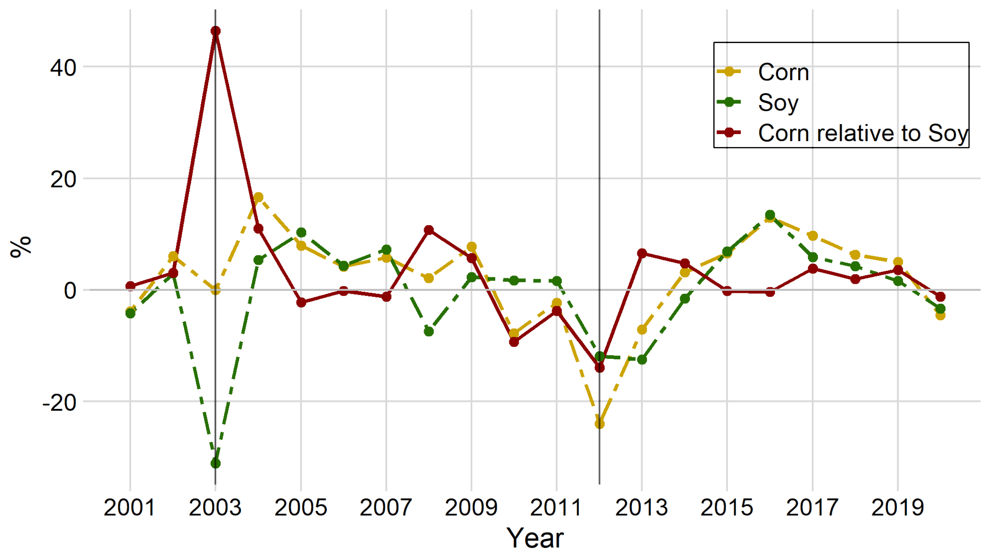


FIGURE 1.3. Percentage Yield Shocks in Iowa

Fact.4 *Total growing-season exposure to extreme heat is 80% more damaging to corn than soybeans in the study region.*

In table 1.1, I report the results from the regressions of log yields on growing-season weather variables using the county-level data from 1961 to 2019 based on the results from Schlenker and Roberts (2009) for corn and soybeans:

$$log(y_{cit}) = \beta_{c1}prec_{it} + \beta_{c2}prec_{it}^2 + \beta_{c3}mdd_{it} + \beta_{c4}hdd_{it} + \alpha_i + f_i(t) + \epsilon_{cit}, \tag{1.1}$$

where y_{cit} denotes the yield for crop c in county i in year t. $prec$ denotes total precipitation, mdd represents total exposure to beneficial heat (Moderate Degree Days: degree days between $10°C$ and $29°C$ or $30°C$), and hdd represents total exposure to harmful heat (Heating Degree Days: degree days above $29°C$ or $30°C$) during the growing season (April to September). α_i represents the county fixed effects, which capture county-specific time invariant unobservables, such as soil quality. $f_i(t)$ is the county-specific linear time trend, which captures variation of log yield that is not explained by weather variables, such as technological improvements.

TABLE 1.1. OLS Results: Log Yield on Weather Variables

	Corn				Soy			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
prec (Apr-Sep)	0.151***	0.154***	0.143***	0.147***	0.111***	0.107***	0.107***	0.103***
	(0.041)	(0.041)	(0.042)	(0.043)	(0.022)	(0.021)	(0.022)	(0.020)
precsq (Apr-Sep)	-0.013***	-0.013***	-0.012***	-0.013***	-0.009***	-0.008***	-0.009***	-0.008***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)
mdd (Apr-Sep)	0.040***	0.042***	0.029**	0.030**	0.049***	0.046***	0.042***	0.041***
	(0.014)	(0.015)	(0.013)	(0.013)	(0.009)	(0.008)	(0.008)	(0.008)
hdd (Apr-Sep)	-0.798***	-0.898***	-1.04***	-1.20***	-0.447***	-0.315***	-0.587***	-0.425***
	(0.097)	(0.169)	(0.128)	(0.213)	(0.038)	(0.070)	(0.050)	(0.090)
hdd (Aug)		0.277		0.445		-0.365**		-0.457**
		(0.271)		(0.345)		(0.168)		(0.210)
Temp. Threshold	$29°C$	$29°C$	$30°C$	$30°C$	$29°C$	$29°C$	$30°C$	$30°C$
County FE	✓	✓	✓	✓	✓	✓	✓	✓
County Linear Trend	✓	✓	✓	✓	✓	✓	✓	✓
R^2	0.824	0.826	0.827	0.829	0.826	0.830	0.827	0.831
Observations	17,090	17,090	17,090	17,090	17,089	17,089	17,089	17,089

Note: The table shows the regressions of log yield on weather variables using US county data in Iowa, Illinois, and Indiana from 1961 to 2019. Cluster robust standard errors are in parentheses. Standard errors were clustered by year. *prec* denotes total precipitation, *precsq* denotes squared total precipitation, *mdd* denotes total degree days between $10°C$ and a temperature threshold ($29°C$ for corn or $30°C$ for soy). *hdd* denotes total degree days above the temperature threshold. Asterisks indicate the following: ***=1% significance level, **=5% significance level, and *=10% significance level.

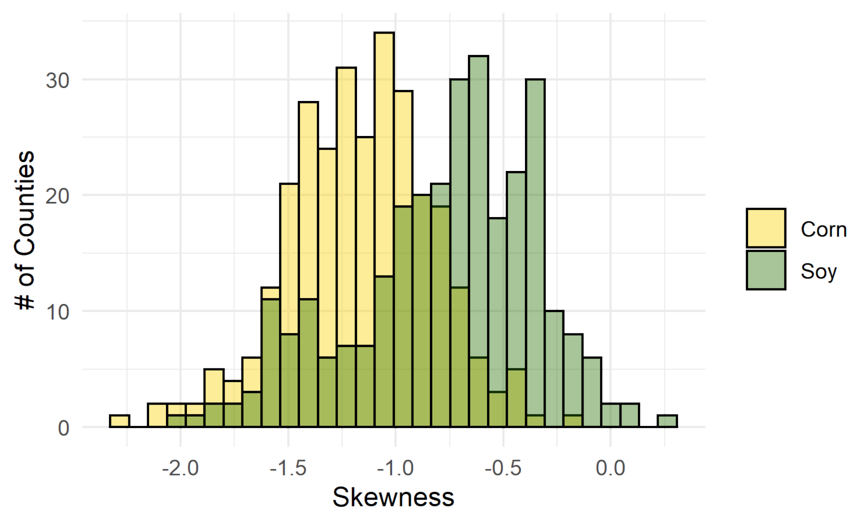


FIGURE 1.4. Histogram of (county-level) Skewness of Yield Shocks

Note: The figure uses county-level yield shocks measured by deviations from a linear time trend (1961–2020).

Columns (1) and (5) show that aggregate exposure to extreme heat over the growing season is 80% more damaging to corn than to soybeans in the region. This could be the result of two compounding effects. One is that corn is generally more susceptible to water stress and the other is extreme heat tends to peak around the phase when it is more damaging to corn. Not surprisingly, corn yield shocks tend to be more negatively skewed in the study region as figure 1.4 shows. Comparison between columns (2) and (6) confirms the heterogeneous time-varying effects on crop yields between the two crops.

1.3. Conceptual Framework

In what follows, I describe how the degree of recency effects associated with two weather signals—total exposure to or timing of extreme heat—can affect crop choice decisions at the intensive margin. In doing this, I make use of details on the unique agroclimatic contexts in the study region. I then relate the insights from the model to my empirical setting. That is, I develop testable hypotheses that allow me to use coefficients on different types of yield shocks to infer to which weather signal farmers respond sensitively. Another important goal of the conceptual model is to show that recency effects of weather-induced yield shocks on intensive-margin crop choice decisions can be very subtle and even counter-intuitive if one does not carefully consider the local unique agroclimatic features. Specifically, it is possible that a farmer experiencing local weather that favored one crop over the other this year may form a belief that the unfavored crop would have a higher yield in the subsequent year.

Because of limitations of the observational data used in this study, I do not attempt to assess the causes or rationality of a potential recency effect. I use the term recency effect to simply refer to how much a farmer weighs recent information more heavily than older information in her decision-making. In practice, recency effects could arise for various reasons: limited memory, time-varying states, high level of ambiguity about the distribution of growing-season weather, high level of ambiguity aversion, and high reliance on heuristics due to difficulty of processing information (Hogarth and Einhorn 1990; Kala 2015). For example, Kala (2015) shows that high ambiguity aversion could lead to a recency bias even when agents believe that states do not vary over time.

I use a simple model that reflects climatic conditions and heterogeneous crop yield responses to weather between corn and soybeans with an emphasis on extreme heat in the study area.[2] Consider a farmer who

[2] The literature has documented that extreme heat is by far the most influential weather variable that explains crop yield variability for the two crops in the region (Schlenker and Roberts 2009).

decides to plant either corn or soybeans on a nonirrigated field.[3] Assume that yields are exogenously determined by local weather. For simplicity, I model crop yields as a function of extreme heat during the growing season. One may consider extreme heat in the model as exposure to heat and water stress combined. This is because heat and water stress are highly intertwined. In the statistical framework of crop yields, extreme heat partially picks up water stress during crop growth stages, when water stress is not explicitly controlled for in a regression model (Ortiz-Bobea et al. 2019).

For simplicity, I assume that extreme heat occurs only once each growing season and is independently and identically distributed across seasons. Let s_t denote the within-season timing of extreme heat in year t and let i_t denote its intensity. This simplified representation is far from how weather is realized in practice because jointly probabilistic climatic conditions are drawn over the entire growing season. However, reflecting more realistic climatic settings complicates the model without improving the intuition. I further assume that fields are relatively homogeneous in terms of climate and soil characteristics so that the farmer can learn what her yield would have been even for the crop she did not plant simply by observing yields on her neighboring fields.

Let $y^c(s_t, i_t)$ denote realized yield for crop c in year t with s_t and i_t and, similarly, let $y^c(s_t, i_t = 0)$ denote the potential yield achievable when there is no extreme heat ($i_t = 0$). For comparability of crop yields between corn and soybeans, I write crop yields as realized yield normalized by potential yield: $y_t^c \equiv \frac{y^c(s_t,i_t)}{y^c(s_t,i_t=0)}$. Thus, corn-relative-to-soybean yield shock (hereinafter, relative yield shock) can be written as $Y_t(s_t, i_t) \equiv \frac{y_t^{corn}}{y_t^{soy}} - 1$.

Holding all other things constant, such as expected crop prices and rotational benefits, the farmer would plant the crop that she expects to have a higher yield. Since extreme heat is the only component of crop yield in the model, a belief about extreme heat is sufficient for a belief about relative yield shock. Suppose the farmer forms a belief about the timing of extreme heat this year based on observations available up to and including last year as follows:

$$E_{t-1}[s_t] = (1 - \rho_s)\bar{s}_{t-2} + \rho_s s_{t-1} \tag{1.2}$$

$$\approx (1 - \rho_s)s^a + \rho_s s_{t-1}, \tag{1.3}$$

[3] Acreage adjustment along the extensive margin (transition between corn or soybeans and the other land uses including idle) has been limited in the study area.

where $\bar{s}_{t-2} = \frac{1}{t-2}\sum_{\tau=1}^{t-2} s_\tau \approx s^a$ and s^a is the population mean. Similarly for intensity,

$$E_{t-1}[i_t] = (1-\rho_i)\bar{i}_{t-2} + \rho_i i_{t-1} \tag{1.4}$$

$$\approx (1-\rho_i)i^a + \rho_i i_{t-1}. \tag{1.5}$$

$\rho = (\rho_s, \rho_i)$ is a vector of *recency effect* parameters. These parameters describe how much the farmer weighs last year's observation than the older ones in the dimension of the timing and intensity of extreme heat. Without more information about farmers' belief formation, there is little reason to assume $\rho_s = \rho_i$ or $\rho_s, \rho_i \in [0,1]$. For example, a farmer might respond sensitively to a higher-than-normal heat intensity last year, if the farmer considers it as a signal of a larger heat wave occurring.

Suppose the farmer forms a belief about extreme heat in the coming year based on the historical distribution of extreme heat around the phase that critically differentiates corn and soybean yields. Normalize the length of the differentiating time window to 1 (e.g., mid of June to end of August). For tractability of the model, I adopt a simple functional form that maps extreme heat (s,i) to relative yield shock (Y) as follows: $Y(s,i) = i(s-0.5)$. This functional form reflects the stylized facts that corn (soybeans) becomes less (more) susceptible to extreme heat as the growing season progresses (See figure 1.1). Around $s = 0.5$, relative yield shock becomes zero.

Suppose, under climate (i.e., the typical growing season) denoted as a, $s^a = 0.25$ and $i^a = 1$. The assumption $s^a = 0.25$ reflects the fact that extreme heat occurs when corn is more susceptible to the heat than soybeans under climate. Let a vector $a \equiv (s^a = 0.25, i^a = 1, Y(s^a, i^a))$ represent the timing and intensity of extreme heat under climate and its corresponding relative yield shock. Figure 1.5 depicts such extreme heat and relative yield shock in the (s, Y) space while i is represented via slope of a line $Y(s,i)$. Given the functional form and a, expectation of relative yield shock for this year can be expressed as

$$E_{t-1}[Y_t | Y_{t-1} = Y(s_{t-1}, i_{t-1}), \rho_s, \rho_i] = ((1-\rho_i) + \rho_i i_{t-1})((1-\rho_s)0.25 + \rho_s s_{t-1} - 0.5)$$

Suppose that, last year, the farmer experienced extreme heat and yield at $b \equiv (s^b = 0.8, i^b = 2, Y(s^b, i^b))$. The intensity was twice higher than normal and the timing was toward the phase when soybeans are relatively more susceptible to extreme heat. This weather experience is similar to growing-season weather in 2003 in Iowa when weather favored corn over soybeans. If the farmer is free of *recency effects*, the weather shock b

would have a limited effect on her belief formation about relative yield shock. Formally,

$$
\begin{aligned}
&E_{t-1}[Y_t|Y_{t-1}, \rho_s = 0, \rho_i = 0] \\
=&E_{t-1}[Y_t|\rho_s = 0, \rho_i = 0] \\
=&Y(s^a, i^a) \\
\equiv&Y^a
\end{aligned}
$$

The first equality implies that, if a farmer is free of *recency effects*, then her belief for this year is independent of last year's shock. The second equality means that her expected relative yield shock is identical to relative yield shock under climate.

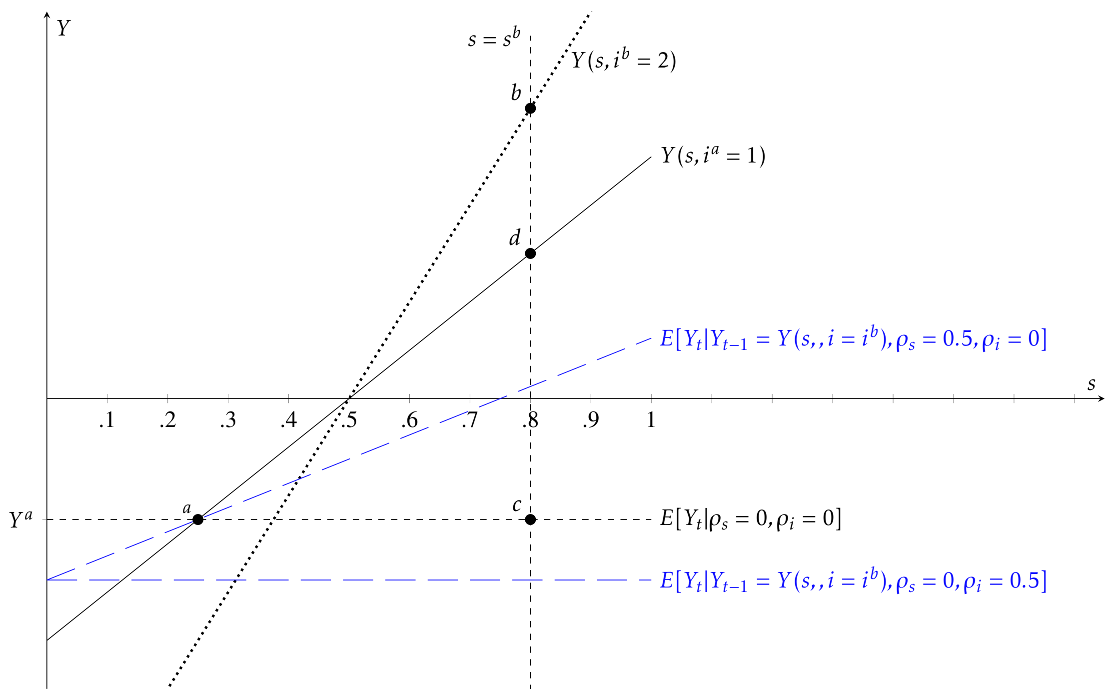


FIGURE 1.5. (*Conceptual Framework*) Canonical Cases

To build some intuition behind how *recency* parameters affect her belief formation, consider two canonical cases depicted in figure 1.5: i) $\rho_s > 0$ and $\rho_i = 0$, and ii) $\rho_s = 0$ and $\rho_i > 0$. Suppose $\rho_s > 0$ and $\rho_i = 0$. In this case, a higher-than-normal relative yield shock last year (i.e., $Y_{t-1} > Y^a$) leads the farmer to form a belief of a higher-than-normal relative yield shock this year (i.e., $E_{t-1}[Y_t|Y_{t-1}, \rho_s > 0, \rho_i = 0] > Y^a$). Graphically with b, $E_{t-1}[Y_t|Y_{t-1} = Y(s^b, i^b), \rho_s > 0, \rho_i = 0]$ would be positioned somewhere between point

c and d as long as $\rho_s < 1$. This is because $\rho_i = 0$ requires the line $E_{t-1}[Y_t|Y_{t-1}, \rho_s > 0, \rho_i = 0]$ pivot around (s^a, Y^a).[4] Suppose $\rho_s = 0$ and $\rho_i > 0$. In this case, a higher-than-normal relative yield shock last year does not necessarily lead the farmer to form an expectation of a higher-than-normal yield shock this year. Instead, regardless of last year's relative yield shock or the timing of extreme heat, greater-than-normal heat intensity leads the farmer to expect that weather would be less favorable-than-normal to corn. Graphically with b, $E_{t-1}[Y_t|Y_{t-1} = Y(s^b, i^b), \rho_s = 0, \rho_i > 0]$ would be positioned below $Y(s^a, i^a)$. This is because $\rho_s = 0$ requires the slope of the line $E_{t-1}[Y_t|Y_{t-1} = Y(s^b, i^b), \rho_s = 0, \rho_i > 0]$ to be zero and the expectation of higher-than-normal intensity lowers the expected relative yield shock. In this case, even if weather was more favorable-than-normal to corn than soybeans last year, the farmer believes that corn yield would be lower than soybeans this year relative to normal. This is because the farmer updates her belief about extreme heat only in the intensity dimension and, under the normal timing ($s = s^a$), corn is more vulnerable to extreme heat than soybeans.[5]

The insights from the model allow me to develop some conjectures for my empirical analysis. In the empirical analysis, I regress a binary crop choice (1 for corn and 0 for soybeans) on (lagged) relative yield shock, crop-specific yield shocks, and average yield shock of corn and soybeans. If farmers tend to react sensitively to timing (i.e., during which growing season phase extreme heat was concentrated) but little to intensity (i.e., total growing-season exposure to extreme heat), *ceteris paribus* farmers are more likely to plant the crop with a higher yield last year. In this case, we expect that a higher relative yield shock would predict a higher probability of planting corn. On the other hand, if farmers tend to react sensitively to intensity but little to timing, they are more likely to plant soybeans if either crop is damaged more than normal. In this case, relative yield shock would not predict the probability of planting corn but higher crop-specific and average yield shocks would predict a higher probability of planting corn. Finally, if farmers are

[4]More generally,

$$\begin{aligned} sign\{E_{t-1}[Y_t|Y_{t-1}, \rho_s > 0, \rho_i = 0] - Y^*\} &= sign\{Y_{t-1} - Y^a\} \\ &= sign\{s_{t-1} - s^a\}. \end{aligned}$$

[5]More generally,

$$sign\{E_{t-1}[Y_t|Y_{t-1}, \rho_s = 0, \rho_i > 0] - Y^a\} = -sign\{i_{t-1} - i^a\}$$

Put differently,

$$sign\{E_{t-1}[Y_t|Y_{t-1}, \rho_s = 0, \rho_i] - Y^a\} = \begin{cases} sign\{Y_{t-1} - Y(i_{t-1} = 1, s_{t-1})\}, & \text{if } s_{t-1} < .5 \\ -sign\{Y_{t-1} - Y(i_{t-1} = 1, s_{t-1})\}, & \text{otherwise} \end{cases}$$

sensitive to both intensity and timing, it is not straightforward to form a hypothesis although the insights from the canonical cases remain still valid. This is because it is less clear how relative magnitudes of two *recency* parameters would interplay. In figure S1.1 and S1.2, I include non canonical cases where both *recency* parameters are positive.

1.4. Data

The dataset for my empirical analysis covers Iowa, Illinois and Indiana from 2004 to 2020. The unit of analysis is field-by-year. Although crop classification data date back to earlier than 2004 for the three states, local crop price data at hand are available from 2004. The dependent variable of interest is field-level binary crop choice (corn or soybeans). The potential independent variables of interest are county-level yield shocks and exposure to extreme heat over the growing season. The control variables include contemporaneous-year crop-specific ”market” expectation of crop revenues, realized revenue shocks, and planting-time soil moisture.

Field-level Crop Choice

To construct annual field-level crop choice data, I use two types of geospatial data following Hendricks et al. (2014b), Hendricks et al. (2014a) and Pates and Hendricks (2021). One is pixel-level crop classification data. The other is agricultural field boundaries. I obtain pixel-level crop classification data from the U.S. Department of Agriculture (USDA)'s Cropland Data Layer (CDL). The CDL is a 30m-resolution satellite-derived crop map produced annually to assess crops and cropland area across the contiguous United States. Any satellite-derived classification data contain classification errors. Such classification errors can vary across classifications, regions, and time. Fortunately, corn and soybeans are among the most accurate crops in the CDLs with accuracy well above 90% in the sample period for the study area. (See Lark et al. (2021) for more details.)

For field boundaries, I use the Common Land Unit (CLU) from the USDA Farm Service Agency (FSA). The CLU delineates the smallest unit of land that has a common land cover and land management in agricultural land associated with USDA farm programs. I use the 2008 version of the CLU because it

CDL

Field-level Mode after Cleaning

FIGURE 1.6. CDL with Field Boundaries from CLU

is the only version available at hand.[6] To construct annual field-level crop data, I take the mode of crop classifications (i.e., dominant crop) for each field in a given year and use the mode as field-level crop choice.

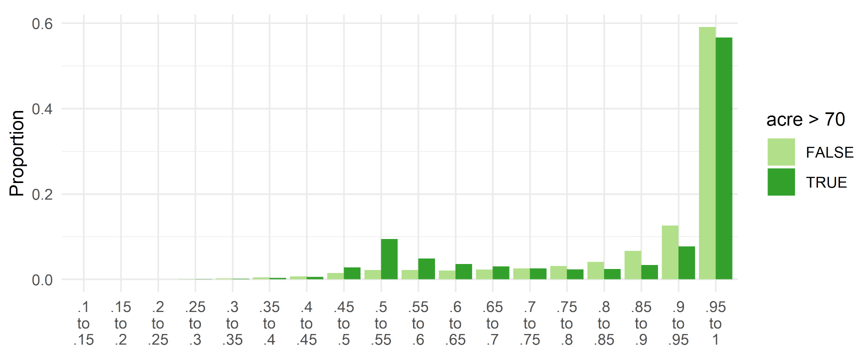


FIGURE 1.7. Distribution of the Share of Field-Level CDL Mode

One potential issue with use of time-invariant field boundaries is that they may have evolved over time, for example, by someone renting or buying/selling a portion of a field. Figure 1.6 shows some examples of crop classifications from the CDL overlaid with field boundaries from the CLU. As can be seen from the figure, some instances can be observed with significant portions of multiple crops in a field boundary. Without further information, it is not possible to distinguish whether these observations are a result of a farmer planting multiple crops in one field or multiple farmers making different crop choice decisions.

[6]Following Hendricks et al. (2014b), I drop polygons smaller than 15 acres because these polygons are likely to be non-agricultural boundaries, such as gullies, waterways, or farmsteads.

Figure 1.7 shows such a tendency is more pronounced for large fields. To mitigate such concerns and classification errors from the CDL, I drop observations whose proportion of field-level mode is smaller than 0.8. This data-cleaning process leaves me 11 million observations (left panel in figure 1.9). Among these observations, I drop fields that always rotate or never rotate. A quarter of fields in the sample belong to the former category and 4% belongs to the latter category.

Geographic Boundaries

Following Hendricks et al. (2014b) and Pates and Hendricks (2021), I run separate econometric models over a set of geographic boundaries known as Major Land Resource Areas (MLRAs) from the Natural Resources Conservation Service (NRCS). Each MLRA shares similar biophysical and environmental characteristics in terms of soil, water, climate, vegetation, land use, and type of farming. A total of 33 MLRAs intersect with or are included in the study area. Some MLRAs contain much smaller numbers of fields than others. I merge small MLRAs into relatively large neighboring MLRAs to finally have 20 geographic groups. Figure 1.8 shows these boundaries together with the CDL. Figure 1.10 shows how similar soil characteristics are within MLRAs and different across MLRAs in terms of clay, sand, silt and ph in 0-5cm depth (Walkinshaw et al. 2020).

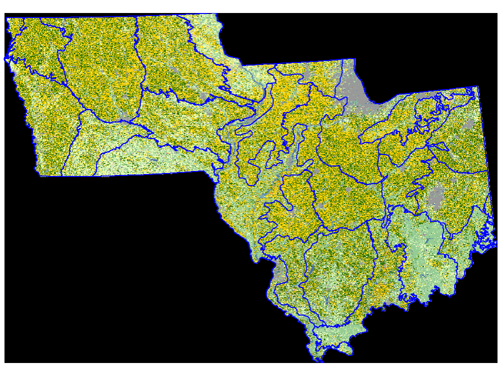


FIGURE 1.8. CDL with MLRA

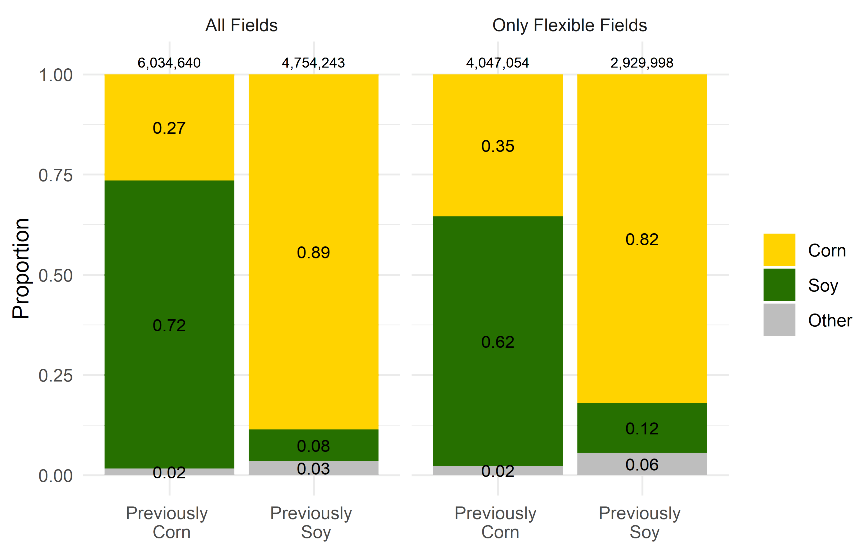


FIGURE 1.9. Share of Crop Choice by Previous Crop

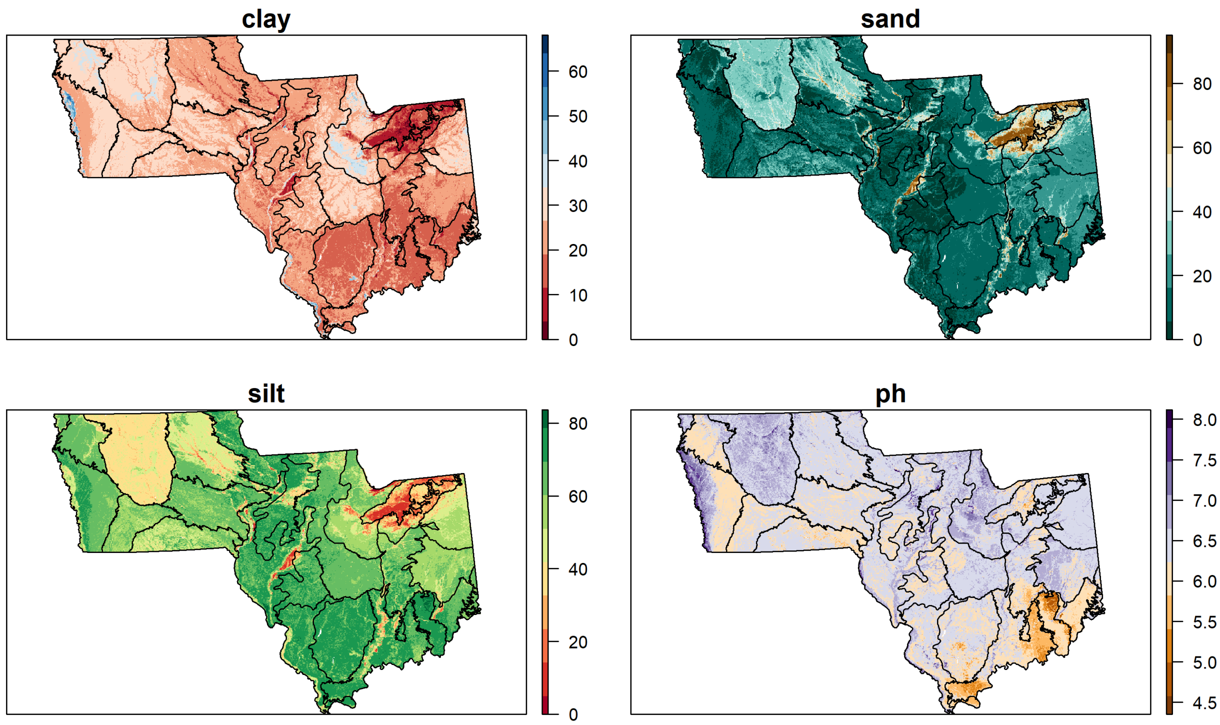


FIGURE 1.10. Soil Properties (0-5cm)

Note: Values in clay, sand, and silt are percent by weight.

Yield and Acreage

I obtained the county-level data on planted acres and yields for corn and soybeans from the USDA National Agricultural Statistics Service (NASS).

Crop Prices

The local crop price data are sourced from Data Transmission Network (DTN) and Cash Grain Bids (CGB). These price data are point-level mostly at grain elevators.[7] I interpolate these point-level price data to a high resolution grid using the inverse-distance weighting method. I take the simple average of interpolated gridded price data within each field to construct field-level price data.

For post-harvest price, I take the simple average of the daily spot price in December for corn and in November for soybeans. For planting-time ”market” expectation of post-harvest crop price, I use a futures price (national component) adjusted by expected local crop basis (local component), which is measured

[7] I thank Matthieu Stigler for providing geolocation information about grain elevators.

by the difference between local spot price and nearby futures contract price. For futures prices, I use futures contracts traded on the Chicago Board of Trade (CBOT). For corn (soybeans), I average December (November) contract futures prices traded in the months of January and February. Although January and February are months ahead of planting seasons for many regions in the US Corn Belt, farmers tend to start to prepare inputs, including seeds and fertilizers in the months. For expected basis, I average local grain basis in March. Figure 1.11 shows interpolated planting-time expected prices and post harvest prices for corn and soybeans in 2012 when there was a historic drought in the region. The figure suggests that there exists some variation in prices over space but it is limited.

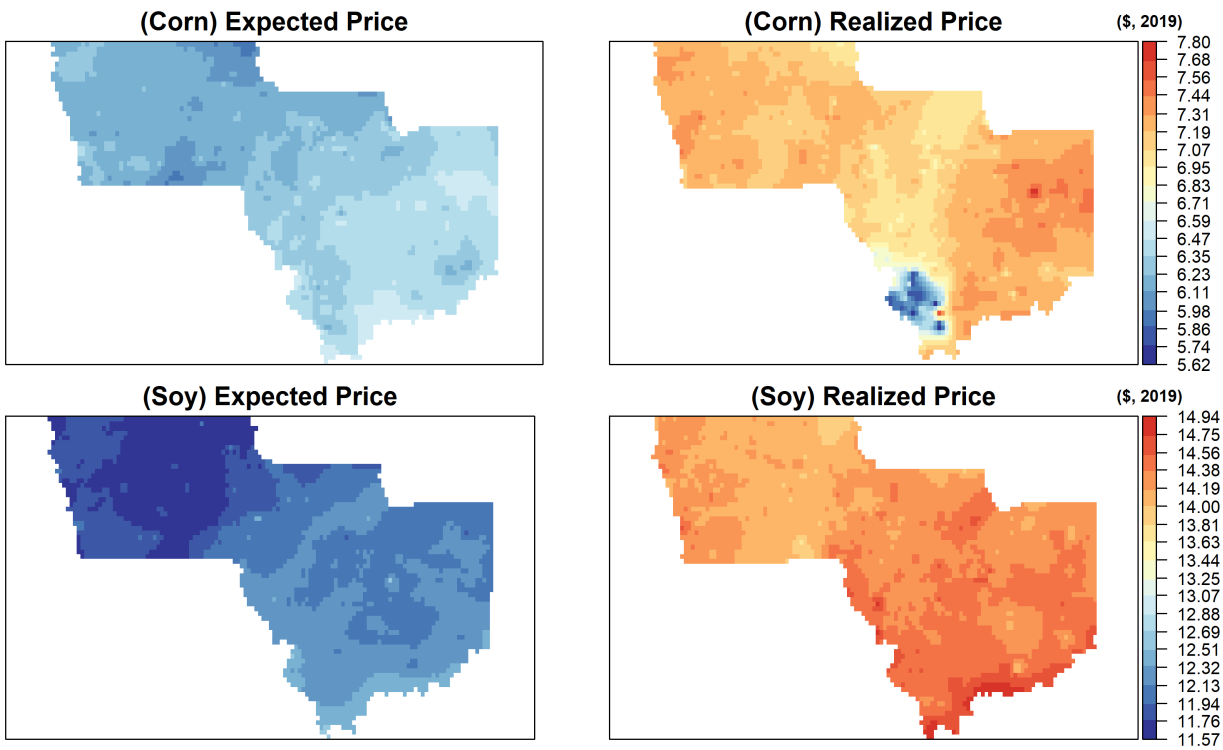


FIGURE 1.11. Interpolated Corn and Soybean Prices in 2012

Climate

To be consistent with yield data, I use county-level weather data for my empirical analysis. I first construct daily county-level weather variables on precipitation and (minimum and maximum) temperatures, I aggregate gridded data (4km resolution) from the PRISM Climate Group to the county level using cropland areas as weights. For farmland weights, I use the National Land Cover Database 2019, which identifies land use at a 30m resolution across the US. Using daily temperature data, I calculate degree days above temperature thresholds (e.g., $10°C$) over the growing season (April-September). I conduct auxiliary analysis on weather and crop yields using either county-level panel data or county-specific time-series data for 1961–2019. Because the PRISM gridded data are available only after 1980, for this analysis, I use county-level temperature and precipitation data made available by Wolfram Schlenker. This dataset was derived from the balanced panel of weather stations.[8]

Soil Moisture

In my analysis, I control for planting-time soil moisture level in 2cm depth at the field level, because sometimes excessive wet soil moisture prevents farmers from planting corn making them plant soybeans instead. I obtain gridded surface soil moisture data from the National Aeronautics and Space Administration (NASA)'s Gravity Recovery and Climate Experiment (GRACE; 2002–2017) and GRACE Follow On (GRACE-FO; 2018-present) satellites (Kornfeld et al. 2019; Tapley et al. 2004). The data report wetness percentiles relative to the period 1948–2012 at a 14km resolution on a weekly basis. I take the simple average of surface moisture level in April. Table 1.2 shows descriptive statistics of some important variables. (Construction of yield shocks will be explained in the next section.)

TABLE 1.2. Descriptive Statistics

	min	25th	median	75th	max	SD
expected price for corn ($/bushel, 2019)	2.46	3.80	4.31	5.58	7.52	1.07
expected price for soy ($/bushel, 2019)	5.27	9.12	10.24	12.96	16.58	2.22
yield for corn (bushel/acre)	19.00	158.00	176.60	191.20	246.70	29.98
yield for soy (bushel/acre)	19.00	47.72	52.40	57.00	80.40	7.42
expected revenue for corn ($/acre, 2019)	303.52	682.73	754.23	892.56	1327.77	187.90
expected revenue for soy ($/acre, 2019)	247.10	479.79	540.11	643.96	898.78	120.86
average yield shock	-0.66	-0.05	0.02	0.07	0.34	0.11
relative yield shock	-0.79	-0.05	0.00	0.06	1.28	0.12
surface soil moisture (percentile)	0.78	41.39	62.04	80.83	100.00	23.68

1.5. Empirical Strategy

Many fields in the US Corn Belt rotate their crops from one year to the next year, most switching between corn and soybeans. Crop rotation provides several benefits. For example, crop rotation naturally

[8] I used data and code available at Wolfram Schlenker's webpage (`http://www.columbia.edu/~ws2162/links.html`).

replenishes soil nutrients and reduces pest populations. These benefits often lead to yield-boosting and/or input-saving effect (Hennessy, 2006). Because of these agronomic benefits of crop rotation, the incentive to plant a crop this year largely depends on what was planted in the previous year. As the crop choice is the main outcome of interest, such dependence gives rise to a dynamic panel setting.

I adopt the econometric framework of Hendricks et al. (2014b). Hendricks et al. (2014b) uses a first-order Markov transition model to characterize crop choice dynamics. Following Hendricks et al. (2014b) I estimate the model using ordinary least squares (OLS) along with the linear probability model. I run a separate econometric model for each MLRA group to alleviate the potential bias induced by coefficient heterogeneity. Hendricks et al. (2014b) show that failure to account for such coefficient heterogeneity can lead to an illusion that acreage response to expected crop prices is larger in the long run than in the short run. This illusion contradicts the agronomic intuition that predicts higher short-run acreage response in the presence of crop rotation benefits.

I focus on the intensive margin (transition between corn and soybeans) because this is the main margin of acreage adjustment particularly in the short run. I characterize the set of linear Markov transition probabilities as follows:

$$\phi_{it}^{cc} = Pr(c_{it} = 1 | c_{i,t-1} = 1) = \beta_{1i}^{l} \mathrm{S}_{j(i),t-1}^{l} + \theta_{1i}' X_{it} + f_{1i}(t) + \alpha_{1i}, \tag{1.6}$$

$$\phi_{it}^{sc} = Pr(c_{it} = 1 | c_{i,t-1} = 0) = \beta_{0i}^{l} \mathrm{S}_{j(i),t-1}^{l} + \theta_{0i}' X_{it} + f_{0i}(t) + \alpha_{0i}, \tag{1.7}$$

where ϕ_{it}^{cc} (ϕ_{it}^{sc}) denotes the probability of planting corn on field i in year t conditional on corn (soybeans) having been planted in the previous year. The main parameters of interest are βs. $S_{j(i)t}^{rel}$ denotes the relative yield shock in county j for field i in year t. Similarly, $S_{j(i)t}^{avg}$ denotes the average yield shock (weighted by acreage across corn and soybeans). X_{it} includes a vector of control variables: crop-specific expected revenues and realized revenue shocks in the current year, and planting-time surface soil moisture. For time trend $f_i(t)$, I use a county-specific quadratic time trend. α_i denotes field fixed effects, which will pick up time-invariant field-level characteristics, such as soil quality. Following Hendricks et al. (2014b), I use subscript i to indicate that coefficients are heterogeneous across fields. The probability of transition from corn to soybeans is simply $\phi^{cs} = 1 - \phi^{cc}$, and the probability of transition from soybeans to soybeans is $\phi^{ss} = 1 - \phi^{sc}$.

I construct relative yield shock, $S^{rel.}_{j(i)t}$, and average yield shock, $S^{avg.}_{j(i)t}$ as below.

$$S^{rel.}_{j(i)t} = \frac{y^{corn}_{j(i)t}/\hat{y}^{corn}_{j(i)t}}{y^{soy}_{j(i)t}/\hat{y}^{soy}_{j(i)t}} - 1, \tag{1.8}$$

$$S^{avg.}_{j(i)t} = \frac{\sum_{c\in\{corn,soy\}} a^{c}_{j(i)t} y^{c}_{j(i)t}/\hat{y}^{c}_{j(i)t}}{\sum_{c\in\{corn,soy\}} a^{c}_{j(i)t}} - 1, \tag{1.9}$$

where $y^{c}_{j(i)t}$ is the realized yield, $\hat{y}^{c}_{j(i)t}$ is the predicted yield, and $a^{c}_{j(i)t}$ denotes planted acre. In the baseline model, for predicted yields $\hat{y}^{c}_{j(i)t}$, I use a linear time trend to fit (county-level) yield data spanning from 1991 to 2020 for $\hat{y}^{c}_{j(i)t}$. I measure these shocks at the county level, which will effectively capture yield shocks that fields in a county experience on average. Although farming practices, and soil and climate characteristics are highly similar within counties in the study region, some counties would have more heterogeneous fields than others, potentially leading to less precise parameter estimates. One might raise some concerns over use of realized yields when constructing yield shocks because not all yield shocks are driven by weather fluctuations. For example, previous studies have found that severity of air pollution affects crop yields and cleaner air has contributed to yield gains for corn and soybeans over the past two decades in the US (Lobell and Burney 2021; Metaxoglou and Smith 2020). In the results section, I also report results from a regression model that uses yield shock variables constructed with predicted yields based on weather.

I consider only one-year lag of local yield shocks because including more than one-year lag complicates interpretation of my results by introducing direct and indirect channels. Suppose we include both two-year lag and one-year lag of yield shocks in the baseline econometric model. Conditional on last year's shock, the shock two years ago affects the farmer's crop choice this year in two channels. Through the (direct) belief-changing channel, the shock two years ago might have affected what the farmer expects her yields to be this year if *recency effects* last more than one year. Through the (indirect) rotation-incentive-changing channel, the farmer might have planted a different crop last year than what she would have planted without the shock two years ago, and the potentially different crop choice last year might alter a rotation incentive this year.

I include as control variables contemporaneous-year crop-specific a) "market" expectation of crop revenues and b) realized crop revenue shocks. I approximate expected revenues and revenue shocks are calculated as follows:

$$\widetilde{rev}^c_{it} = log(\tilde{p}^c_{it}\hat{y}^c_{j(i)t}) \tag{1.10}$$

$$rev^c_{it} = log(p^c_{it}y^c_{j(i)t}/\tilde{p}^c_{it}\hat{y}^c_{j(i)t}) \tag{1.11}$$

where $\tilde{p}^c_{it}$ is planting-time expected post-harvest price and $\hat{y}^c_{j(i)t}$ is the trend yield. p^c_{it} is the realized post-harvest price and $y^c_{j(i)t}$ is the realized yield. Although I included expected crop revenues $\widetilde{rev}^c_{it}$, most of the within-field variation comes from temporal variation in price due to limited year-to-year fluctuations in trend yields. For this reason, using expected prices rather than expected revenues as control variables does not change the results presented in this paper. I also control for average soil moisture level in April. Because the crop choice response to planting-time soil moisture could be nonlinear and the appropriate functional form is not apparent, I fit a natural spine with 3 knots. The number and location of knots do not change the estimated parameters of interest.

Following Hendricks et al. (2014b), I run a separate model for each MLRA. When reporting aggregate estimated parameters, I average them weighted by crop acres of MLRAs. For standard errors, I use a wild cluster bootstrap method with 300 replications. I cluster standard errors by year. Clustering standard errors by year allows for cross-sectional dependence across fields in any given year but assumes independence between years. For each replication, I preserve the regressors but reconstruct the dependent variable by summing the predicted value from OLS and the reconstructed residuals whose sign switches by year with the probability of 0.5 (Cameron et al., 2008).

I exploit within-field variation for identification. Hendricks et al. (2014b) and Pates and Hendricks (2021) exploit both cross-sectional and temporal variation while including field characteristics (e.g., soil characteristics and irrigation status) as controls in their econometric model rather than including field fixed effects. This implicitly assumes fields and the farmers who use them are identical within a MLRA conditional on field characteristics. Exploiting cross-sectional variation could be less preferred in my setting. While the explanatory variable of interest in the previous studies was crop-specific expected prices which vary across fields, my explanatory variables of interest—-yield shocks—-vary only across counties in a given year. Because there are not many counties in each MLRA, parameter estimates could be less stable among different specifications. Admittedly, including individual fixed effects could cause biases in the dynamic panel model (Nickell 1981). In 1.8.2, I show that the magnitude of the potential bias induced by fixed effects would be negligible. In practice, coefficients obtained by exploiting both cross-sectional and temporal variation and those obtained by using only within-field variation tend to be similar, but estimates tend to be more robust to multiple specifications for the latter.

1.6. Results

In what follows, I present empirical results focusing on the fields that were planted to corn in the prior year, because these fields tend to leave more room for crop choice adjustment than fields previously planted to soybeans. Seifert et al. (2017) show that yield drag from continuous corn is 4.3% and that for continuous soybeans is 10.4% although there is some heterogeneity across sub-regions. A larger magnitude of yield drag from continuous soybeans implies that fields that were planted to corn have more room for flexibly changing crop choices. Because of this asymmetric yield drag, continuous corn is much more common than continuous soybeans as shown in figure 1.9. In appendix, I include the results for the fields that were planted to soybeans in the previous year. As expected, the magnitudes of estimated parameters are generally much smaller for fields that were previously planted to soybeans.

Before presenting my main results, I address some potential concerns readers might raise.

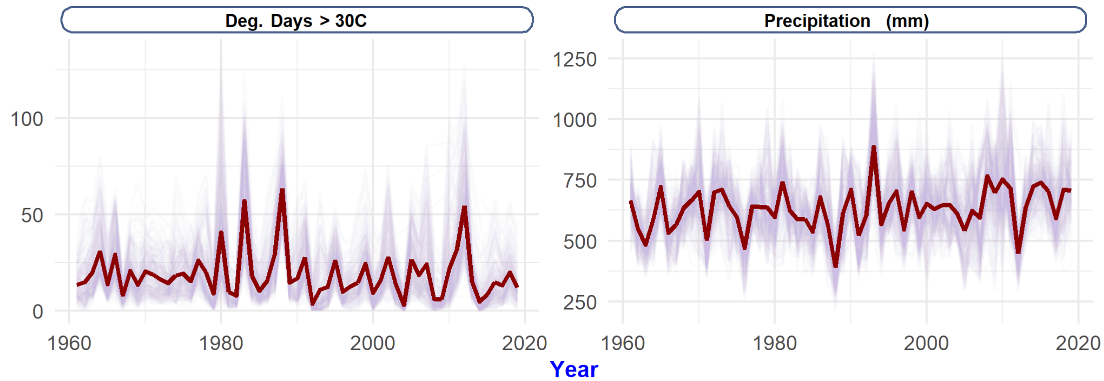
Are growing-season weather and yield shocks random from year to year?

My results suggest yes. Figure 1.12 shows county-level time series of growing-season weather variables and yield shocks from 1961 to 2019 with thick lines representing average values across counties in the study region. Each figure shows little autucorrelation over the period, To formally test stationarity of these variables, I conduct the augmented Dickey Fuller (ADF) test for each county. I consider only one lag because this time lag is most relevant to my empirical setting. Results show that, for all weather and yield shock variables used in the figure, the null hypothesis of having a unit root is rejected at a 5% significance level in more than 99% percent of the counties in the study region.

Does prior-year weather have predictive power of crop yields?

My results suggest no. To see if prior-year weather has predictive power of crop yields, I regress log yields on a set of lagged weather variables using county-level data spanning from 1961 to 2019. As table 1.3 suggests, both for corn and soybeans, the null hypotheses of the joint nullity of the coefficients on lagged

(a) (growing-season) Weather



(b) Yield Shocks

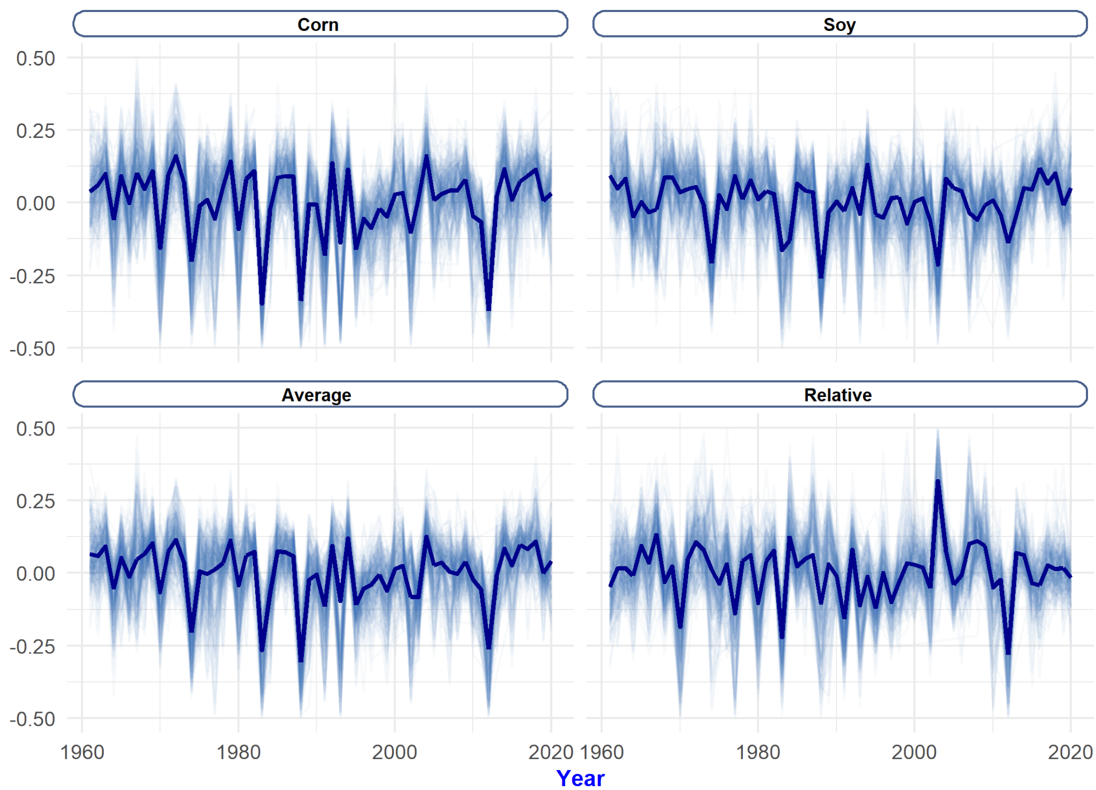


FIGURE 1.12. Time Series of (county-specific) Weather Variables and Yield Shocks

Note: The thin lines show county-specific values, while the thick lines represent the average values across all counties in Iowa, Illinois, and Indiana. For *Corn* and *Soy*, yield shocks were constructed by the ratio of realized yield to the linear trend yields minus 1. For information about construction of average and relative yield shocks, see Eq.(1.8) and Eq.(1.9).

TABLE 1.3. OLS Results: Log Yield on (lagged) Weather Variables

	Corn			Soy		
	(1)	(2)	(3)	(4)	(5)	(6)
(t-1) prec (Apr-Sep)	-0.022	-0.025	-0.027	-0.053*	-0.057*	-0.062**
	(0.040)	(0.041)	(0.041)	(0.030)	(0.030)	(0.028)
(t-1) precsq (Apr-Sep)	0.001	0.001	0.002	0.004*	0.004*	0.004**
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
(t-1) mdd (Apr-Sep)	0.008	0.008	0.008	-0.012	-0.014	-0.016
	(0.020)	(0.022)	(0.022)	(0.011)	(0.012)	(0.012)
(t-1) hdd (Apr-Sep)	-0.039	-0.038	-0.035	-0.120	-0.113	-0.117
	(0.141)	(0.150)	(0.154)	(0.074)	(0.078)	(0.081)
P value (Joint Nullity)	0.9611	0.9430	0.9216	0.1397	0.1166	0.0717
County FE	✓	✓	✓	✓	✓	✓
County time trend	linear	quadratic	spline	linear	quadratic	spline
R^2	0.674	0.677	0.680	0.762	0.769	0.774
Observations	17,082	17,082	17,082	17,081	17,081	17,081

Notes: The table shows the regressions of log yield on (lagged) weather variables using US county data in Iowa, Illinois, and Indiana from 1961 to 2019. Cluster robust standard errors are in parentheses. Standard errors were clustered by year. $prec$ denotes total precipitation, $precsq$ denotes squared total precipitation, mdd denotes total degree days between $10°C$ and a temperature threshold ($29°C$ for corn or $30°C$ for soy). hdd denotes total degree days above the temperature threshold. Asterisks indicate the following: ***=1% significance level, **=5% significance level, and *=10% significance level.

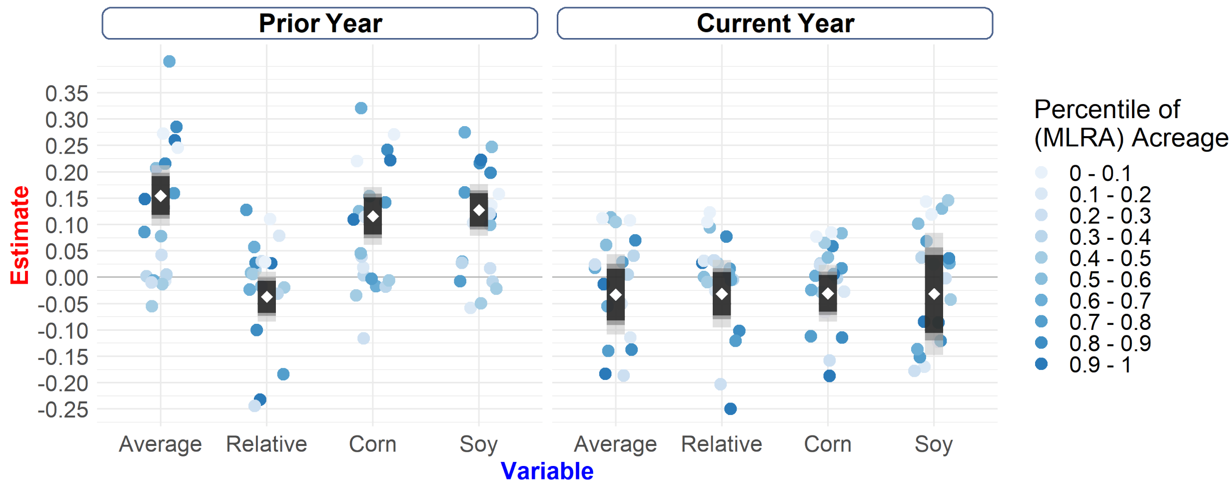
weather variables are rejected at the 5% significance level. This suggests that there is little evidence that prior-year weather predicts current-year crop yields. This result would not be surprising given serially uncorrelated weather variables. Admittedly, for soybeans, (lagged) precipitation and precipitation squared have statistically significant coefficients but the signs are opposite to those observed in the regression results of log yields on current-year weather variables.

1.6.1. Results with various types of lagged shocks.

Do prior-year yield shocks affect crop choice?

My results suggest yes for average and crop-specific yield shocks but no for relative yield shocks. The top left plot in figure 1.13 shows estimation results for various (lagged) shocks. In each column, I regress a binary crop choice (1 for corn and 0 for soybeans) on a different shock variable conditional on the control

(a) Yield Shocks



(b) Degree Days above Extreme Heat

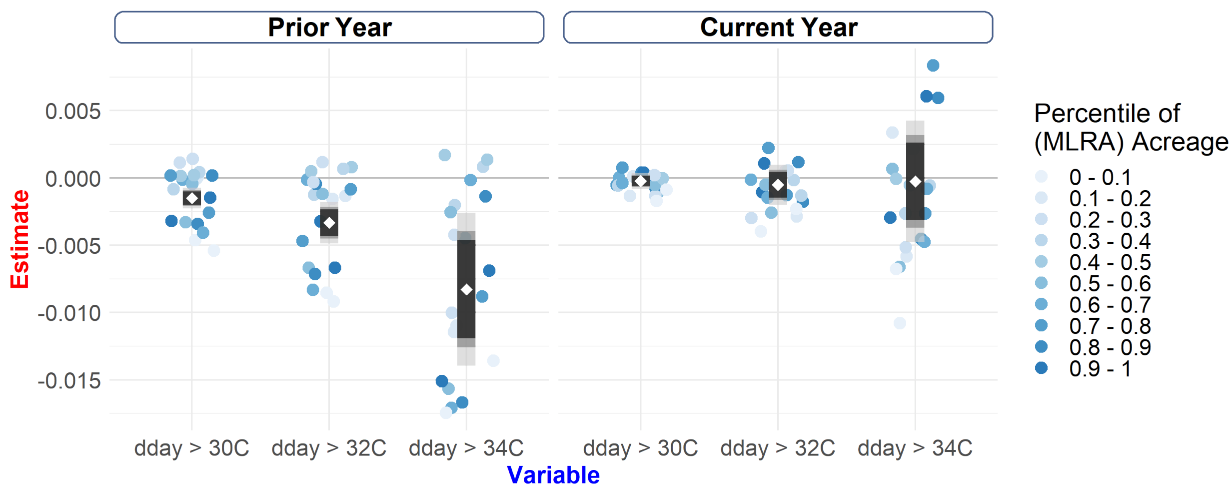


FIGURE 1.13. Effects of (lagged) Yield and Weather Shocks on the Probability of Planting Corn

Note: The figure visualizes OLS results from the linear probability model: probability of planting corn conditional on corn being planted in the previous year. *Average* denotes average yield shocks weighted by acreage across corn and soybean, *Relative* denotes corn-relative-to-soybean yield shocks. *dday* means total degree days from April to September. White dots represent average coefficients weighted by the size of MRLAs and dark lines show their 90%, 95% and, 99% confidence intervals obtained from standard errors from a wild bootstrap whereby data were resampled by year. Colored points show the point estimates for individual MLRAs with color indicating the size of MLRA.

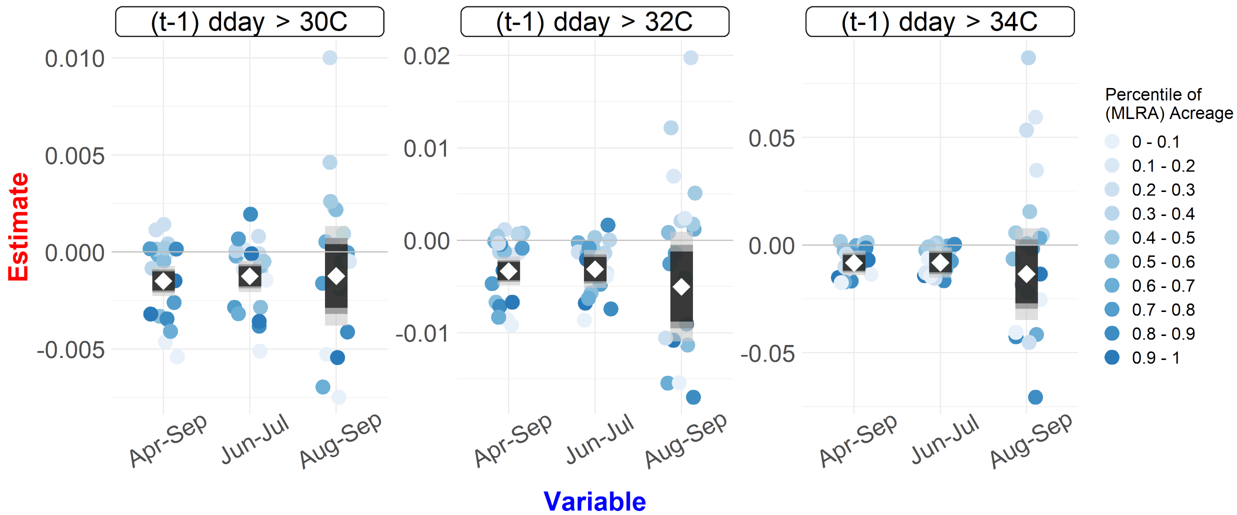


FIGURE 1.14. Coefficients on (lagged) Extreme Heat

*Note:*The figure visualizes OLS results from the linear probability model: probability of planting corn conditional on corn being planted in the previous year. White dots represent average coefficients on weighted by the size of MRLAs and dark lines show their 90%, 95% and, 99% confidence intervals obtained from standard errors from a wild bootstrap whereby data were resampled by year. Colored points show the point estimates for individual MLRAs with color indicating the size of MLRA.

variables: current-year crop-specific expected revenues, realized revenue shocks, and planting-time soil moisture. Colored circle-shaped points show MLRA-specific coefficients and white diamond-shaped points represent average coefficients weighted by the size of MLRAs.

Average yield shock has a statistically significant positive effect on the probability of planting corn. An increase in average yield shock by 10 percentage points (approximately one standard deviation) leads to an increase in the probability of planting corn by 1.5 percentage points in the current year. (In the next subsection, I discuss in more detail the implications of the magnitudes of these estimates.) As can be seen in the figure, there is substantial coefficient heterogeneity across MLRAs, implying that failure to account for such coefficient heterogeneity may lead to a pooling bias (Pesaran and Smith 1995). It is also noticeable that crop-specific yield shocks also significantly and positively predict the probability of planting corn. Interestingly, the average coefficient on relative yield shock is close to zero. Based on the insight from the conceptual model, the results in figure 1.13 are consistent with the canonical case in which a farmer shows a strong *recency effect* in the dimension of intensity but not in that of timing of extreme heat.

Although assessing what drives coefficient heterogeneity is beyond the scope of this paper, coefficients on (lagged) average yield shocks tend to be large—with a few noticeable exceptions—in the regions where temporal variability of yield shocks is low, soil quality is high, and the crop choice response to prices is strong. These heterogeneous effects could arise for many reasons. For example, heterogeneous biophysical environments could make crop choice a more viable and effective strategy for mitigating potential weather shocks than other farming practices, such as soil management in some regions than in others.

Do prior-year weather shocks affect crop choice?

My results suggest yes. The bottom left plot in figure 1.13 shows coefficients on total degree days above $30°C$, $32°C$, and $34°C$ from April to September in the prior year.[9] The results suggest that farmers tend to plant less corn if exposure to extreme heat was larger-than-normal in the previous year. In figure 1.14, I disaggregate the extreme heat variables into two sub-periods: June-July and August-September. The coefficients on extreme heat variables for the months of August and September tend to be noisier, because extreme heat tends to be concentrated in the months of June and July. With this caveat, however, average coefficients for the late growing season tend to be larger than those for the early growing season. Recall that soybeans are particularly susceptible to extreme heat in the later phase of the growing season. This means that farmers are more likely to plant soybeans even if extreme heat was concentrated when extreme heat is particularly damaging to soybeans in the previous year.

Do farmers anticipate crop yields?

To investigate this question, I run the same models used to generate the left plots in figure 1.13 but use contemporaneous-year shocks rather than prior-year shocks. The right plots in figure 1.13 show that there is no systematic evidence that farmers anticipate contemporaneous-year shocks. It is possible that farmers anticipate yield shocks based on planting-season weather regardless of what happened in the previous years. However, planting-season weather has played a limited role in predicting growing-season weather. In one interview in 2013, Joe Glauber, a former chief economist at the USDA, states ”early season moisture levels are “a poor predictor” of how things will play out during the growing season. During the past 60 years in Iowa, when the season began with low subsoil-moisture levels, half had corn yields that were above

[9] I include additional (lagged) weather variables as control variables: precipitation, squared precipitation, and total growing degree days.

average" (Doering et al., 2013). Even in 2012 when a historic drought in the US Corn Belt significantly favored soybeans over corn, there was no noticeable acreage shift toward soybeans.

1.6.2. How did the 2012 US Midwest drought affect farmers' acreage allocation in 2013?

In 2012, the US Corn Belt experienced one of the most disastrous droughts in its history. Figure 1.15 shows the spatial distribution of percentage average yield shocks in 2012. For many counties in the study region, average yield dropped by more than 20%. In what follows, I use my regression results to estimate the impact of the drought on acreage allocation at the local and aggregate levels. I assume, for each county, the counterfactual average yield shock to be 0 in 2012, which could have been achieved if corn and soybean yields did not deviate from the trend yields. Figure 1.16 shows spatially heterogeneous coefficients on (lagged) average yield shock across MLRAs for both types of fields: fields that were planted to corn and those that were planted to soybeans in the prior year.

I calculate field-specific estimated effects of the drought on the probability of planting corn by simply multiplying coefficients on the (lagged) average yield shock by county-level average yield shocks in 2012:

$$\textit{For corn fields in 2012}: \frac{\partial Pr_i(c_{it}=1|c_{i,t-1}=1)}{\partial S^{avg}_{j(i),t-1}} S^{avg}_{j(i),2012}$$

$$\textit{For soy fields in 2012}: \frac{\partial Pr_i(c_{it}=1|c_{i,t-1}=0)}{\partial S^{avg}_{j(i),t-1}} S^{avg}_{j(i),2012}$$

Figure 1.17 shows the spatial distribution of estimated effects of average yield shock on the probability of planting corn. As the figure suggests, there is substantial heterogeneity in the estimated effects across regions and crops planted in 2012 due to heterogeneity in coefficients and average yield shocks in 2012. My estimates suggest that for some fields the probabilities of planting corn dropped by up to 15 percentage points.

By taking acreage-weighted average of the estimated changes in the probability of planting corn, I estimate that, on average, for the fields that were planted to corn (soybeans) in 2012, 3% (1.4%) of the acreage in the study region were planted to soybeans instead of corn in 2013 due to the 2012 US Midwest drought. Recall that I include only flexible fields, which exclude fields that always rotate or never rotate. These flexible fields accounted for 64% of the total corn and soybeans acreage in 2012 and, out of these flexible fields, around 66% were planted to corn. Taken all together, I estimate that around 1.4% of the acreage in the study region was planted to soybeans instead of corn owing to the 2012 drought.

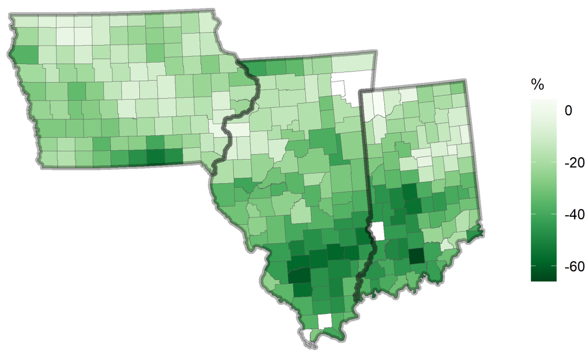


FIGURE 1.15. Average Yield Shock in 2012

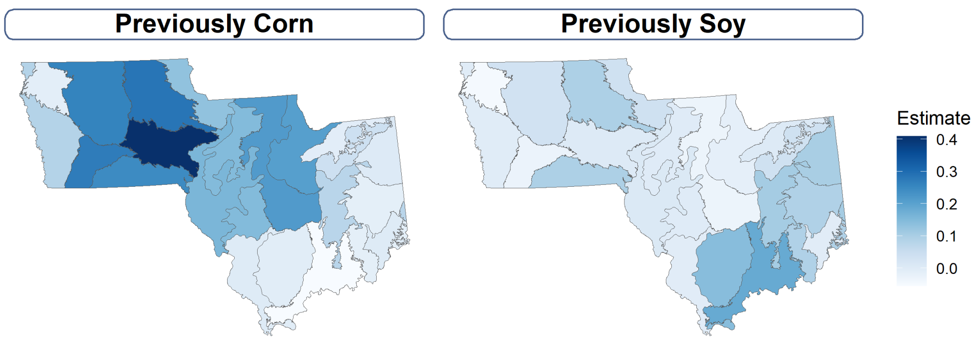


FIGURE 1.16. Coefficients on (lagged) Average Yield Shock on Probability of Planting Corn

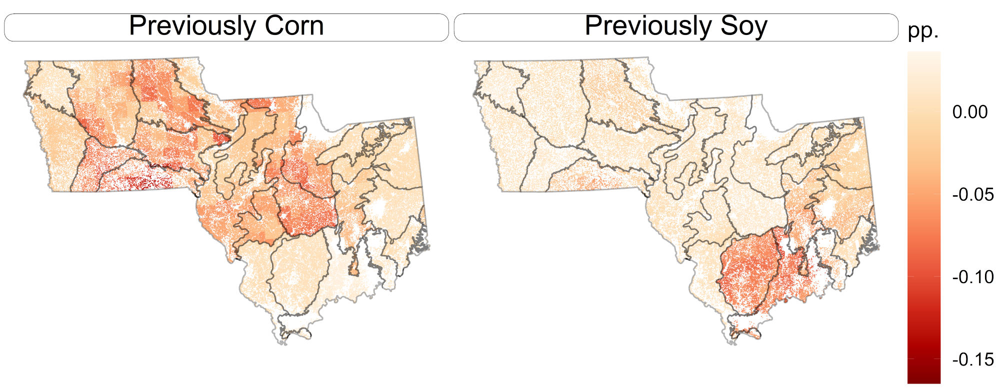


FIGURE 1.17. Effect of Average Yield Shock in 2012 on Probability of Planting Corn in 2013

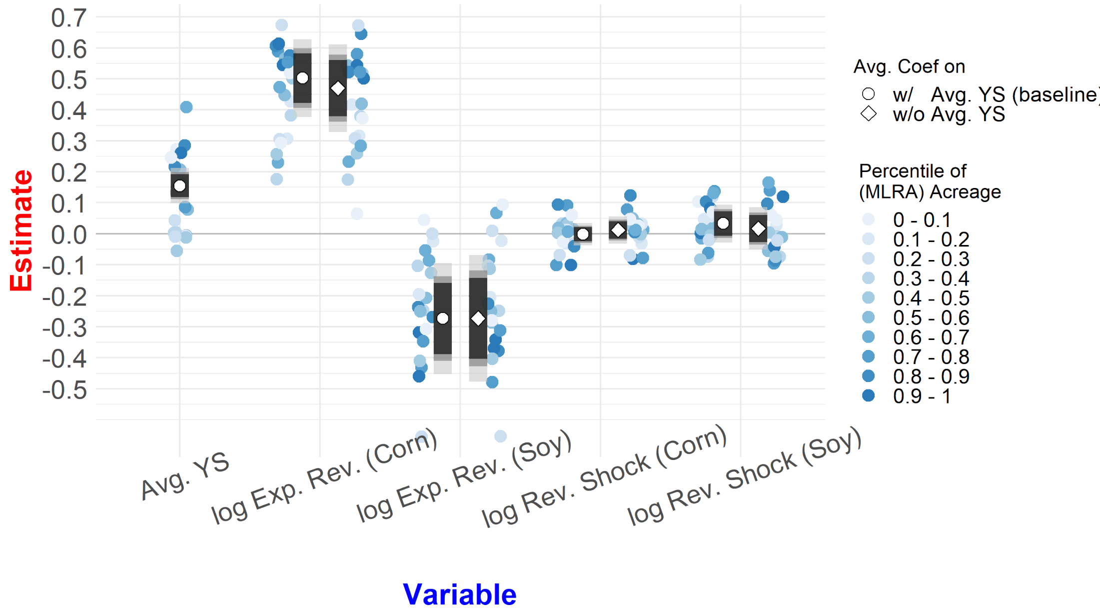


FIGURE 1.18. Coefficients on Control Variables

Note: The figure visualizes OLS results from the linear probability model: probability of planting corn conditional on corn being planted in the previous year. White dots represent average coefficients weighted by the size of MRLAs and dark lines show their 90%, 95% and, 99% confidence intervals obtained from standard errors from a wild bootstrap whereby data were resampled by year. *log Exp. Rev.* means log of expected revenue. *log Rev. shock* means log of realized crop revenue shocks. For more information about variable construction, see the econometric strategy section.

1.6.3. Other results with (lagged) average yield shock.

Hereinafter, I present additional results focusing on (lagged) average yield shocks. One potential concern regarding my specification is that prior-year yield or weather shocks and market expected prices could be highly correlated. This is because weather tends to be spatially correlated and weather in the US Corn Belt tends to influence crop prices. If the high correlation between yield shocks in one year and market expected prices in the planting season of the following year is problematic, then coefficients on market expected revenues—a strong predictor of crop choice—should change substantially depending on inclusion of (lagged) weather or yield shock variables. As figure 1.18 suggests, however, including (lagged) average yield shock barely changes coefficients on control variables. Figure 1.18 also suggests that farmers do not seem to anticipate revenue shocks.

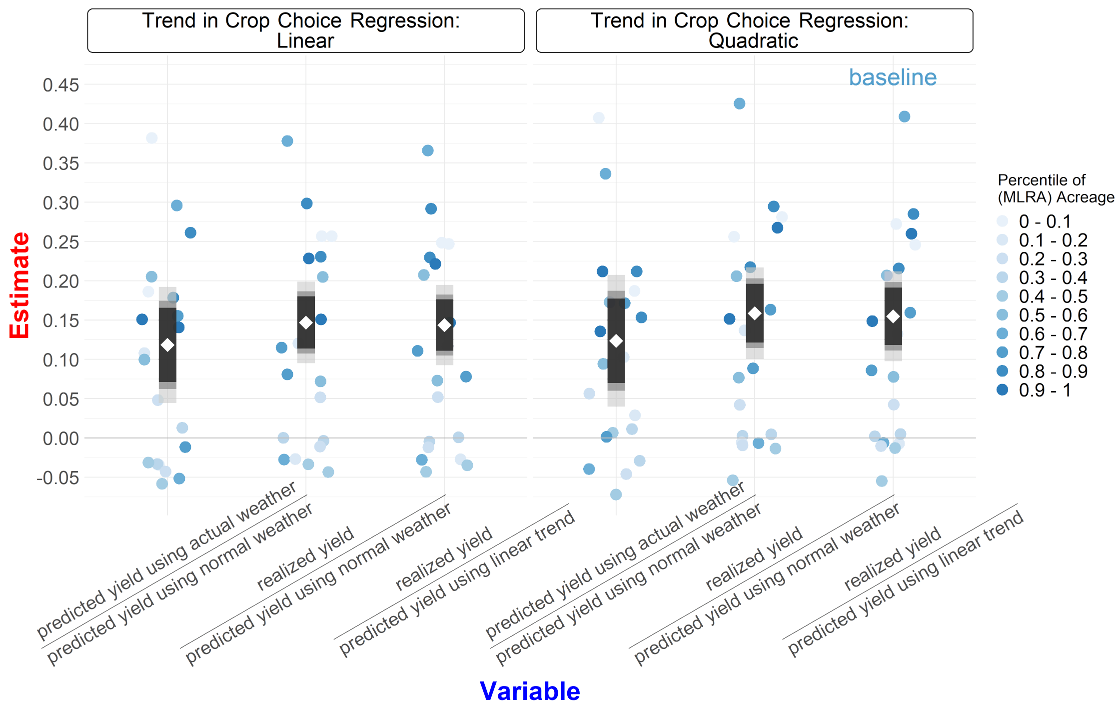


FIGURE 1.19. Coefficients on Multiple Types of (lagged) Average Yield Shocks

Note: The figure visualizes OLS results from the linear probability model: probability of planting corn conditional on corn being planted in the previous year. White dots represent average coefficients weighted by the size of MRLAs and dark lines show their 90%, 95% and, 99% confidence intervals obtained from standard errors from a wild bootstrap whereby data were resampled by year. Colored points show the point estimates for individual MLRAs with color indicating the size of MLRA. *Avg. YS* denotes (lagged) average yield shock.

Does the way of constructing yield shocks affect the main results?

Not substantially. So far I have measured yield shocks as realized yields relative to trend yields. One potential concern would be that non-climatic factors could also influence yields. An alternative way of constructing yield shocks is to use predicted yields from a regression of yield on weather. A major potential limitation of this approach is that it is challenging to precisely quantify the magnitude of weather-driven components at a unit of observation (here county-by-year) using a panel-data econometric approach. One of the widely adopted specifications is the panel model in Eq.(1.1). This statistical approach is simple but assumes constant marginal effect of extreme heat over the growing season. As hinted from figure 1.1, this approach substantially underestimates the crop yield declines of soybeans in 2003 and of corn in 2012.

Following Berry et al. (2014), I parsimoniously extends Eq.(1.1) to allow effects of extreme heat to smoothly vary as the growing season progresses. I focus on extreme heat because the literature has documented that exposure to extreme heat during the growing season is the most influential weather variable for crop yield variation. I model crop yields as:

$$log(y_{it}) = g(hdd_{0it}, ..., hdd_{Dit}) + \beta_2 mdd_{it} + \beta_3 prec_{it} + \beta_4 prec_{it}^2 + \alpha_i + f_i(t) + \epsilon_{it}, \tag{1.12}$$

hdd_{dit} is degree days above a temperature threshold ($29°C$ for corn and $30°C$ for soybeans) in $d \in \{1, ..., D\}$, where D is the number of days from April 1 to September 30. Using a natural cubic spline with 4 knots for $g()$, I allow coefficients on extreme temperature to flexibly change over the entire growing season. For mdd, I use $10°C$ and $29°C$ ($30°C$) for corn (soybeans) as temperature thresholds.

The right panel in figure 1.19 shows that the average coefficient from the baseline model changes from .15 to .125 when average yield shocks are constructed using predicted yields from Eq.(1.12). Comparison between the the left and right panels in figure 1.19 shows that results do not substantially change when a (county-specific) linear trend is used rather than a (county-specific) quadratic trend in Eq.(1.6).

Do farmers respond differently to negative or positive, or small or large shocks?

My results suggest no. To explore potential nonlinear or asymmetric responses, I parsimoniously extend the baseline model by including a dummy variable, $D^k_{j(i)m}$, interacted with (lagged) average yield shock:

$$Pr(c_{it} = 1 | c_{i,t-1} = 1) = \beta^k_{1i} S^{avg.}_{j(i),t-1} + \beta^k_{2i}(S^{avg.}_{j(i),t-1} \times D^k_{j(i)m,t-1}) + \theta'_{1i} X_{it} + \alpha_{1i} + f_{1i}(t). \tag{1.13}$$

I consider four dummy variables. The first two dummy variables aim to investigate farmers' potential asymmetric responses to negative and positive yield shocks:

$$D^{positive}_{j(i)m,t-1} = \begin{cases} 1 & if\, S^{avg}_{j(i)m,t-1} > 0, \\ 0 & otherwise \end{cases}, \qquad D^{negative}_{j(i)m,t-1} = \begin{cases} 1 & if\, S^{avg}_{j(i)m,t-1} \leq 0, \\ 0 & otherwise \end{cases}$$

Previous studies have documented that people tend to react differently to positive and negative events (Ding et al. 2004; Kahneman and Tversky 1979; Krishnamurthi et al. 1992). It is also possible that farmers react more sensitively to either extremely positive or negative yield shocks. To explore such nonlinear responses, I construct dummy variables using 10th and 90th percentiles of the MLRA-specific distribution of average

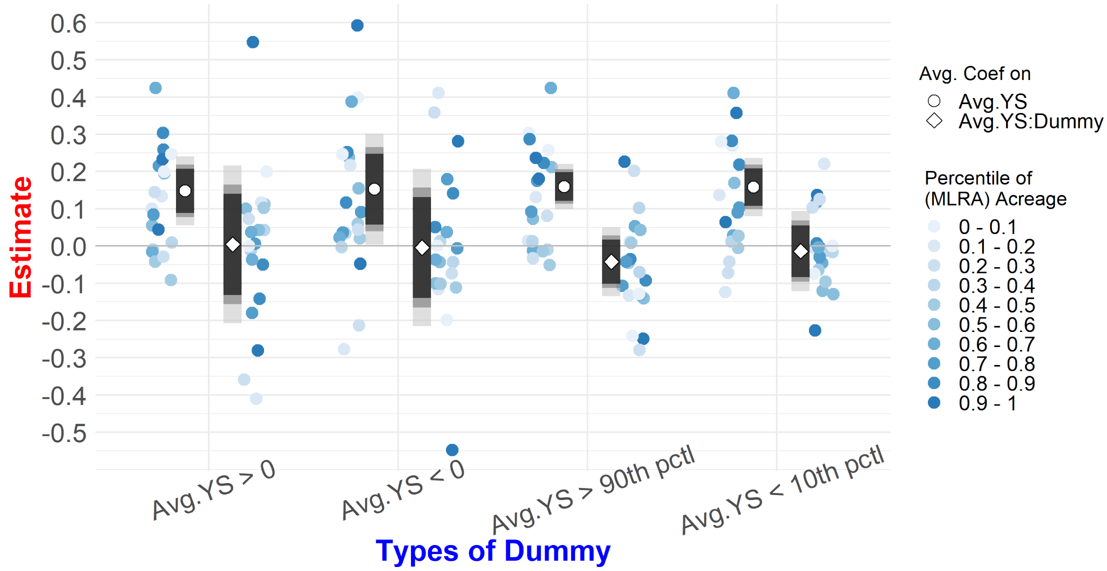


FIGURE 1.20. Coefficients on Dummy Variables Interacted with (lagged) Average Yield Shock

*Note:*The figure visualizes OLS results from the linear probability model: probability of planting corn conditional on corn being planted in the previous year. White dots represent average coefficients weighted by the size of MRLAs and dark lines show their 90%, 95% and, 99% confidence intervals obtained from standard errors from a wild bootstrap whereby data were resampled by year. Colored points show the point estimates for individual MLRAs with color indicating the size of MLRA. *Avg. YS* denotes (lagged) average yield shock.

yield shocks:

$$D^{high}_{j(i)m,t-1} = \begin{cases} 1 & if\ S^{avg}_{j(i)m,t-1} > \psi^m_{90}(S^{avg}_{j(i)m,t-1}), \\ 0 & otherwise \end{cases}, \quad D^{low}_{j(i)m,t-1} = \begin{cases} 1 & if\ S^{avg}_{j(i)m,t-1} \leq \psi^m_{10}(S^{avg}_{j(i)m,t-1}), \\ 0 & otherwise, \end{cases}$$

where ψ^m_k denotes kth percentile of the distribution of average yield shocks for MLRA m. In columns in figure 1.20, I report the regression results of the form in Eq.(1.13) with $D^{positive}_{j(i)m,t-1}, D^{negative}_{j(i)m,t-1}, D^{high}_{j(i)m,t-1}$, and $D^{low}_{j(i)m,t-1}$ (from left to right). In each column of figure 1.20, a white circle-shaped point represents estimated acreage-weighted β^k_{1i}, whereas a white diamond-shaped point represents estimated acreage-weighted β^k_{2i} in Eq.(1.13). Figure 1.20 shows that coefficients on the average yield shocks interacted with dummy variables are not statistically different from zero at the aggregate level, meaning that there is no strong evidence of asymmetric or nonlinear crop choice responses to (lagged) average yield shocks.

1.7. Conclusion

Using field-level data, this paper studied *recency effects* of weather-induced yield shocks on crop choice responses along the rotational margin (transition between corn and soybeans) in the US Corn Belt while accounting for rotation incentives and coefficient heterogeneity. Accounting for rotation incentives is essential given that farmers' crop choice decisions depends strongly on what they planted in the previous year, particularly in the short run.

In the conceptual model, I showed that short-run crop choice responses to past weather could be highly subtle and even counter-intuitive, if we do not account carefully for agroclimatic contexts. My empirical results reveal that farmers tend to plant less corn, when they experienced a hotter or drier summer than average in the prior year. Interestingly, corn-relative-to-soybean yield shocks do not predict the probability of corn being planted. These empirical results are consistent with behavioral patterns in the conceptual model in which farmers sensitively respond to how hot/dry it was in the previous growing season regardless of when (i.e., mid-season or end-season) it was hot/dry.

Evidence of recency effects on crop choice gives reasons for both hope and concern. On the positive side, it is encouraging that farmers tend to respond immediately to past climatic experience. In a rapidly changing climate, recency effects could turn out to be an effective way of mitigating weather-induced agricultural shocks but only to the extent that weather becomes serially correlated from year to year. For the similar reason, however, heavy reliance on recent information may lead to sub-optimal decisions if growing-season weather in one year is independent of that in the prior years. In the paper, I find little evidence that farmers anticipate yield shocks. This is not surprising given that growing-season weather has been highly random from year to year and past weather does not predict the current-year crop yields.

This paper took an initial step to make use of fine-resolution satellite-derived data to explore behavioral aspects of farmers in the major US Corn Belt states. There are multiple directions that future research could take. Future research could utilize an experiment and/or survey to complement the results found in this study. As weather-driven shocks could manifest in multiple ways and farmers' choice could be nuanced, considering local contexts would be essential to drawing valuable insights. Another direction is on margins other than crop choice: soil management, water use, and technology adoption. Lastly, future studies in other agricultural and climatic settings (e.g., specialty or perennial crops and multi-year droughts) would provide their own unique findings and implications.

1.8. Supplementary Section

1.8.1. Estimating time-varying effects of extreme heat on crop yields (for figure 1.1).

I use county-by-year panel data spanning from 1991 to 2020 for Iowa, Illinois, and Indiana to estimate the time-varying effects of extreme heat on crop yields to generate figure 1.1 using the following panel regression:

$$log(y_{it}) = g(h_{0it}, ..., h_{Dit}) + \alpha_i + f_i(t) + \epsilon_{cit}, \tag{1.14}$$

where y_{it} is crop-specific or relative yield in county i and year t. h_{dit} is degree days above $30°C$ in $d \in \{1, ..., D\}$. D is the number of days from May 1 to September 30. The top subfigure uses corn yield, the middle subfigure uses soybean yield, and the bottom subfigure uses corn yield relative to soybean yield.

For $g()$, I use a natural cubic spline with $K = 4$ knots. The basis matrix associated with the spline allows us to reduce the number of parameters to be estimated. (See Ortiz-Bobea 2021 for detailed discussion.) In my setting, the basis matrix maps daily exposure to extreme heat with D bins to K bins ($K < D$). Let H be a $(nT \times D)$ matrix of daily exposure to extreme heat over the growing season. The basis matrix for a natural cubic spline with K knots, B, allows us to reduce dimensionality as follow:

$$\underset{nT\times K}{\widetilde{H}} = \underset{nT\times D}{H} \times \underset{D\times K}{B}.$$

The effect of extreme heat over the growing season could be then modeled as:

$$g(h_{0it}, ..., h_{Dit}) \tag{1.15}$$

$$= \sum_{k=1}^{K} \sum_{d=1}^{D} \gamma_k B_{dk} h_{dit} \tag{1.16}$$

$$= \sum_{k=1}^{K} \gamma_k \underbrace{\sum_{d=1}^{D} B_{dk} h_{dit}}_{\widetilde{h}_{kit}} \tag{1.17}$$

Using the newly constructed regressors $\widetilde{H}$, I can approximate the time-varying effect of extreme heat on crop yield over the entire growing season with K parameters rather than D parameters. The crop yield model that allows for time-varying coefficients on extreme heat can be written as:

$$log(y_{it}) = \sum_{k=1}^{K} \gamma_k \sum_{d=1}^{D} \widetilde{h}_{kit} + \alpha_i + f_i(t) + \epsilon_{it}, \tag{1.18}$$

I recover the marginal effect of extreme heat at each d by pre-multiplying the $\hat{\Gamma}$ the basis matrix:

$$\underset{D\times 1}{\hat{\beta}_1} = \underset{D\times K}{B} \times \underset{K\times 1}{\hat{\Gamma}}$$

I use standard errors clustered by year to allow for spatial dependence in the error term. Using standard errors for γ, $var(\hat{\gamma})$ and a basis matrix, I derive the standard error for the marginal effect of extreme heat at each day d from the diagonal elements of

$$\underset{D\times D}{var(\beta_1)} = \underset{D\times K}{B} \, \underset{K\times K}{var(\hat{\Gamma})} \, \underset{K\times D}{B'}.$$

1.8.2. Bias due to the inclusion of field fixed effects.

In what follows, I show that exploiting within-field variation introduces almost no bias on my estimates. I first formally express the bias using generic notations and I then approximate the bias using data. I assume that coefficients are homogeneous across fields within MLRAs. Once coefficients are homogeneous across cross-section units, the issue boils downs to the Nickell (1981)'s problem. Nickell (1981) derived an explicit expression for bias on OLS estimates caused by individual fixed effects in a dynamic model for large N and small T. The bias on the lagged dependent variable is widely known as Nickell bias. The explanatory variable of interest in my econometric analysis is exogenous variables not the lagged dependent variable. Because this bias is less known, it would be worthwhile to include an explicit bias expression here.

I follow the notations of Nickell (1981):

$$\widetilde{y_t} = [y_{it} - y_{i\cdot}], \quad N \times 1 \text{ vector,} \tag{1.19}$$

$$\widetilde{y}_{t-1} = [y_{it-1} - y_{i\cdot -1}], \quad N \times 1 \text{ vector,} \tag{1.20}$$

$$\widetilde{X}_t = \left[x_{ijt} - x_{ij\cdot}\right], \quad N \times J \text{ vector,} \tag{1.21}$$

$$\widetilde{\epsilon}_t = [\epsilon_{it} - \epsilon_{i\cdot}], \quad N \times 1 \text{ vector,} \tag{1.22}$$

$$\widetilde{b} = \left[\beta_j\right], \quad J \times 1 \text{ vector,} \tag{1.23}$$

where for any variable z_{it}, $z_{i\cdot} = (1/T)\sum_{t=1}^{T} z_{it}$ and $z_{i\cdot-1} = (1/T)\sum_{t=0}^{T-1} z_{it}$.

Consider the dynamic model with exogenous variables that exploits within-individual variation:

$$\overbrace{\begin{bmatrix} \widetilde{y}_1 \\ \widetilde{y}_2 \\ \vdots \\ \widetilde{y}_T \end{bmatrix}}^{\widetilde{y}} = \rho \overbrace{\begin{bmatrix} \widetilde{y}_0 \\ \widetilde{y}_1 \\ \vdots \\ \widetilde{y}_{T-1} \end{bmatrix}}^{\widetilde{y}_{-1}} + \overbrace{\begin{bmatrix} \widetilde{X}_1 \\ \widetilde{X}_2 \\ \vdots \\ \widetilde{X}_T \end{bmatrix}}^{\widetilde{X}} b + \overbrace{\begin{bmatrix} \widetilde{\epsilon}_1 \\ \widetilde{\epsilon}_2 \\ \vdots \\ \widetilde{\epsilon}_T \end{bmatrix}}^{\widetilde{\epsilon}}. \tag{1.24}$$

Let $\hat{\rho}, \hat{b}$ denote the OLS estimates. Nickell (1981) derives

$$\operatorname*{plim}_{N\to\infty}(\hat{\rho}-\rho) = \left(\operatorname*{plim}_{N\to\infty} \frac{1}{NT}\widetilde{y}'_{-1} M \widetilde{y}_{-1}\right)^{-1} \operatorname*{plim}_{N\to\infty} \frac{1}{NT}\widetilde{y}'_{-1}\widetilde{\epsilon} \tag{1.25}$$

$$\operatorname*{plim}_{N\to\infty}(\hat{b}-b) = -\operatorname*{plim}_{N\to\infty}\left[(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'\widetilde{y}_{-1}\right] \operatorname*{plim}_{N\to\infty}(\hat{\rho}-\rho) \tag{1.26}$$

where $M = I - \widetilde{X}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'$.

For each MRLA, I calculate the magnitude of bias on exogenous variables in Eq.(1.26) assuming that my data have sufficiently large N, which is close to true. In my setting, the dependent variable is a binary variable (1 for corn and 0 for soybeans) and exogenous variables are planting-time soil moisture, current-year crop-specific expected revenues, realized revenue shocks and one of (lagged) average, relative and crop-specific yield shocks, exposure to extreme heat. $(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'\widetilde{y}_{-1}$ is a vector of coefficients from the regression of the lagged dependent variable on exogenous variables. In any cases, these coefficients are smaller than .2 in absolute value. If $\operatorname*{plim}_{N\to\infty}(\hat{b}-b)$ is sufficiently small, then we can conclude that the bias on b is small.

$\widetilde{y}'_{-1} M \widetilde{y}_{-1}$ is the mean of squared residuals from the regression of the lagged dependent variable on exogenous variables. Denoted as $\widetilde{e}$, the sample analog of $\widetilde{\epsilon}$ in $\widetilde{y}'_{-1}\widetilde{\epsilon}$ is the residuals from the dynamic model Eq.(1.24). I approximate $\operatorname*{plim}_{N\to\infty}(\hat{\rho}-\rho)$ with $(\widetilde{y}'_{-1} M \widetilde{y}_{-1})^{-1}\widetilde{y}'_{-1}\widetilde{e}$. The values of $(\widetilde{y}'_{-1} M \widetilde{y}_{-1})^{-1}\widetilde{y}'_{-1}\widetilde{e}$ computed using data are smaller than $\frac{1}{100000}$. Thus, I conclude the bias on my estimates due to field fixed effects is negligible.

1.8.3. Supplementary Figures.

1.8.3.1. *Noncanonical Cases.*

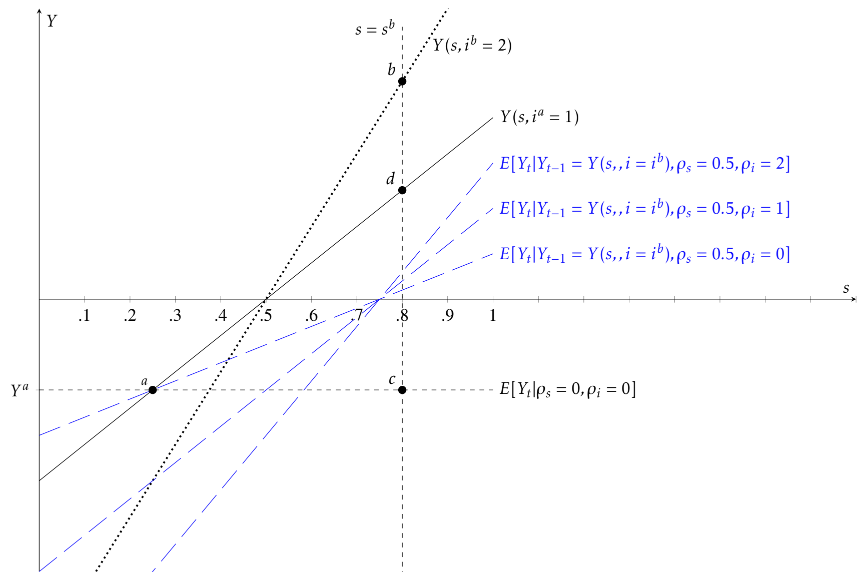


FIGURE S1.1. (*Conceptual Framework*) $\rho_s = 0.5$

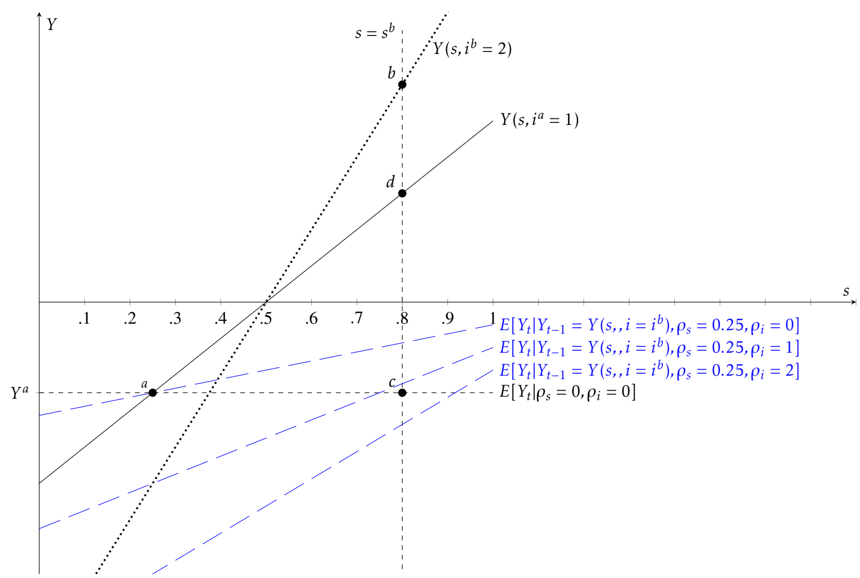
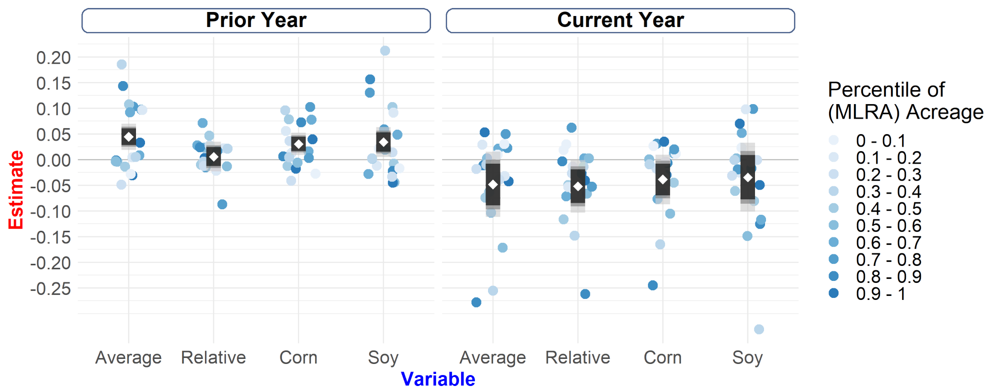


FIGURE S1.2. *(Conceptual Framework)* $\rho_s = 0.25$

1.8.3.2. *Results for Fields Planted to Soybeans in the Previous Year.*

(a) Yield Shocks



(b) Degree Days above Extreme Heat

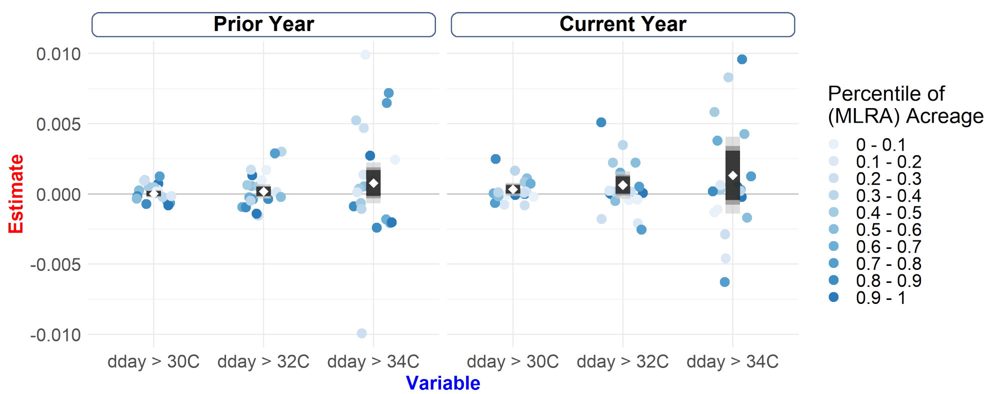


FIGURE S1.3. Effects of (lagged) Yield and Weather Shocks on the Probability of Planting Corn

*Note:*The figure visualizes OLS results from the linear probability model: probability of planting corn conditional on soybean being planted in the previous year. *Average* denotes average yield shocks weighted by acreage across corn and soybean, *Relative* denotes corn-relative-to-soybean yield shocks. *dday* means total degree days from April to September. White dots represent average coefficients weighted by the size of MRLAs and dark lines show their 90%, 95% and, 99% confidence intervals obtained from standard errors from a wild bootstrap whereby data were resampled by year. Jittered points show the point estimates for individual MLRAs with color indicating the size of MLRA.

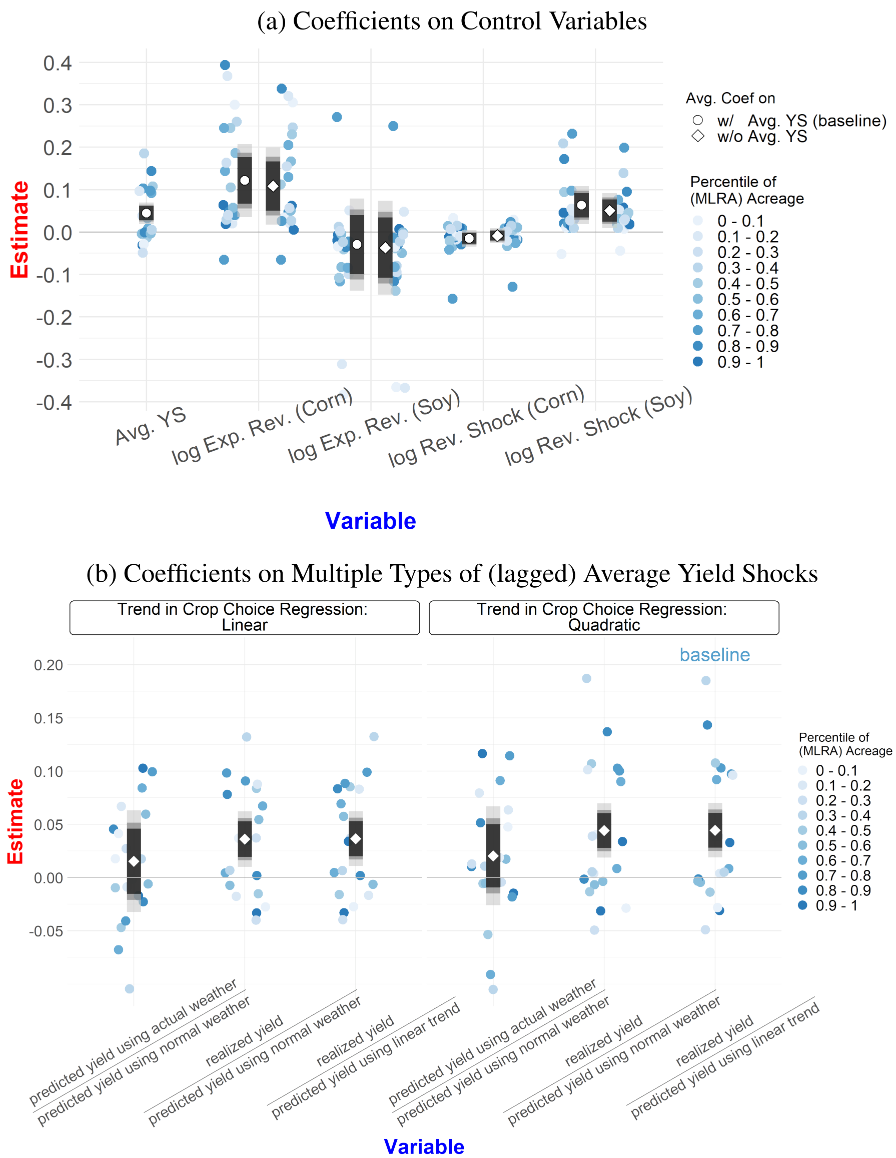
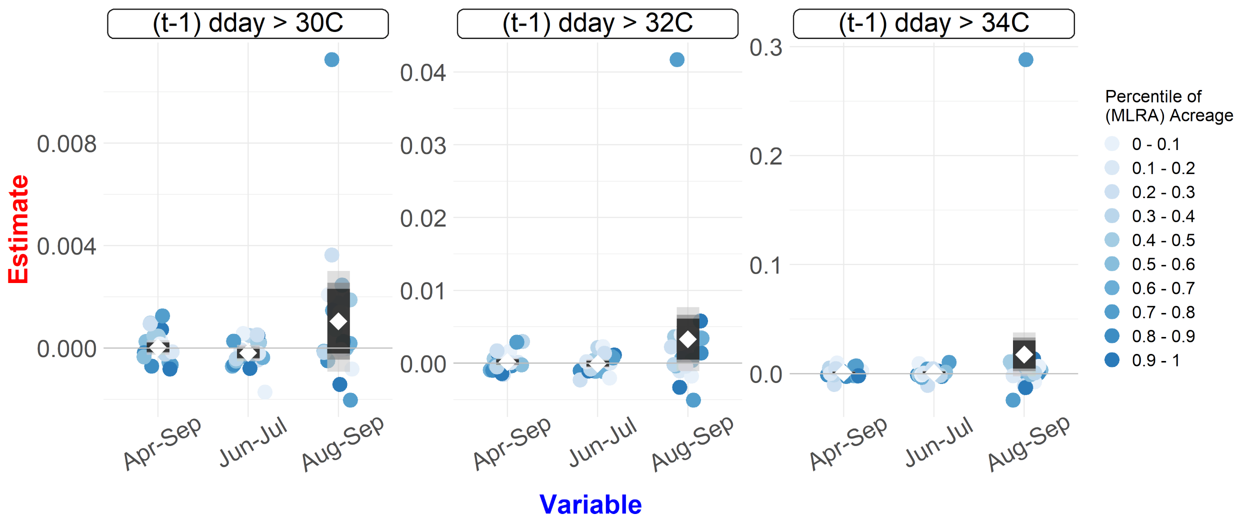


FIGURE S1.4. Results for Fields That Were Planted to Soybeans in the Previous Year I

Note: The figure visualizes OLS results from the linear probability model: probability of planting corn conditional on soybean being planted in the previous year. White dots represent average coefficients weighted by the size of MRLAs and dark lines show their 90%, 95% and, 99% confidence intervals obtained from standard errors from a wild bootstrap whereby data were resampled by year. Jittered points show the point estimates for individual MLRAs with color indicating the size of MLRA. *Avg. YS* denotes (lagged) average yield shock. *log Exp. Rev.* means log of expected revenue.

(a) Coefficients on (lagged) Extreme Heat



(b) Coefficients on Dummy Variables Interacted with (lagged) Average Yield Shock

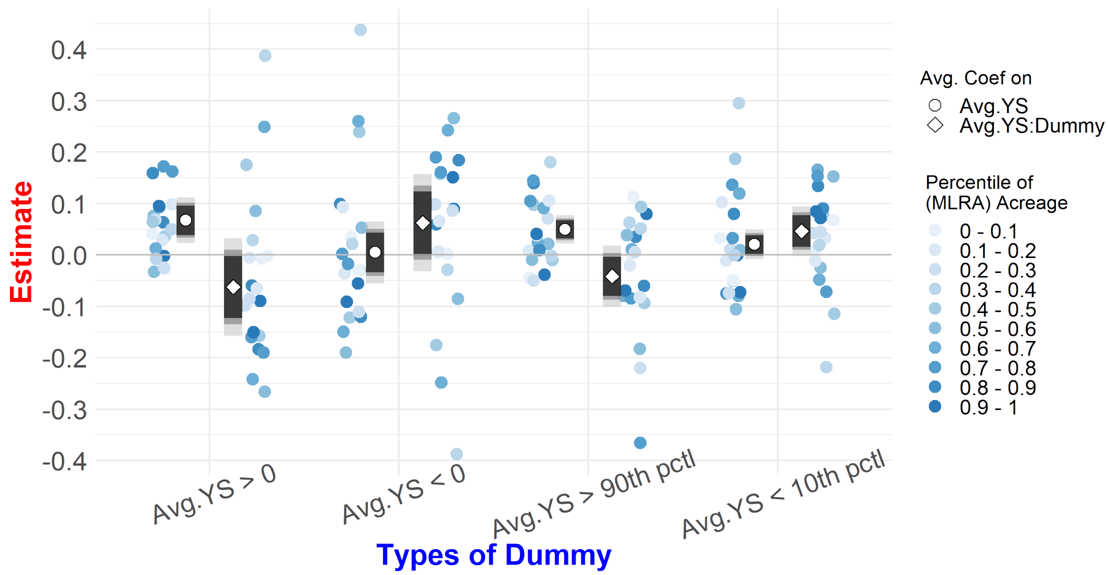


FIGURE S1.5. Results for Fields That Were Planted to Soybeans in the Previous Year II

Note: The figure visualizes OLS results from the linear probability model: probability of planting corn conditional on soybean being planted in the previous year. White dots represent average coefficients weighted by the size of MRLAs and dark lines show their 90%, 95% and, 99% confidence intervals obtained from standard errors from a wild bootstrap whereby data were resampled by year. Jittered points show the point estimates for individual MLRAs with color indicating the size of MLRA. *Avg. YS* denotes (lagged) average yield shock.

CHAPTER 2

Estimating the Impacts of Extreme Weather Events on Crop Revenues Accounting for the Correlation between Price and Yield

2.1. Introduction

Quantification of the sector-specific economic impacts of extreme weather events is essential to informing policy implementation and evaluation in a changing climate (Diffenbaugh et al. 2021). Like any other natural disasters, extreme weather events can result in numerous victims, and damages are often localized to the directly affected areas. From farmers' perspective, however, extreme weather events like droughts and heatwaves in crop-producing regions can generate both winners and losers. When such events hit these areas, they can lead to a substantial heterogeneity in the effects on crop yields across regions but commensurate price increases in response to production declines transmit to broad areas. This combination of yield and price responses to extreme weather events makes it interesting but not straightforward to quantify economic impacts on individual crop producers.

In this study, I quantify the impacts of the two historic US Midwest droughts—i.e., in 1988 and 2012—on the crop-specific revenues for corn and soybeans at the county and national level. To this end, I first propose a panel-data econometric approach to estimating the effects of local and nationwide weather shocks on county-level crop-specific revenues. My approach accounts for the correlation between price and yield, and spatially heterogeneous crop yield responses to weather. Accounting for the correlation is important in the context of the US corn and soybean production because the correlation is negative in many parts of the Corn Belt. Importantly, estimating only yield effects of weather shocks would overestimate the welfare effects of extreme weather events on producers for three reasons. First, as emphasized by Mendelsohn et al. (1994), measuring yield effects alone does not capture changes in crops and inputs (i.e., it ignores behavior). Second, it does not account for price effects. Third, it does not allow for policy responses: governments invariably respond to widespread disasters with disaster assistance policies.

I focus on crop revenues because they are closely linked to profits generated from crop production. Admittedly, weather-induced changes in crop profits are the ultimate outcome of interest. However, spatially and temporally granular data on expenses are not available at least across the US. In the conceptual framework, I first show that weather-induced changes in crop profits are identical to those in crop revenues under the following conditions. First, weather does not change the prices of inputs that farmers use during the growing season. Second, farmers can adjust their inputs in response to weather during the growing season. The close relationship between crop revenues and profits implies that empirically quantifying crop revenue effects could have significant implications for properly formulating and evaluating policies, such as agricultural assistance programs and ad hoc disaster payments. If policy makers determine the amount of disaster payments based on weather or yield shocks, they may end up overcompensating farmers' profit losses.

The most relevant work to this study is Hornbeck (2012). Hornbeck (2012) examines the immediate impacts of and the long-term economic adjustment to the 1930s American Dust Bowl, an environmental catastrophe that greatly eroded sections of the Plains. One of the adjustment margins considered in the paper was agricultural revenue. While both the Dust Bowl and the 1988 and 2012 droughts affected many parts of major crop-producing regions in the US, the former had enduring effects on agricultural productivity due to soil erosion but the latter did not have such persistent productivity effects.

Methodologically, his empirical framework is based on average changes for more-eroded counties relative to changes for less-eroded counties in the same state and with similar pre-1930s characteristics. Because of this empirical setting, his study focuses on between-region relative impacts. As he mentions in the paper, his estimates on relative changes in production—which do not account for price responses—may overstate the absolute degree of adjustment in eroded counties. This is because soil erosion can discourage certain types of production in the eroded location, increase shared output prices, and encourage those types of production in the noneroded location. Unlike Hornbeck (2012), I exploit within-county variation and quantify drought-induced changes in crop revenues at the county level by comparing crop revenues predicted by counterfactual normal and observed weather. In doing so, I explicitly account for price responses to nationwide yield shocks.

In the literature on the economics of climate change, agricultural economic outcomes—such as crop revenue and farm profit—have received academic attentions since economists started studying the impacts of climate change on agriculture (Burke and Emerick 2016; Deschênes and Greenstone 2007; Mendelsohn et al. 1994). The focus of these studies have been on a gradual warming trend rather than extreme weather events. However, climate change means not only changes in the mean temperature but also more frequent and severe extreme weather events like heatwaves (Collins et al. 2013). Because these extreme weather events in major crop-producing regions often lead to an increase in crop prices, it is important to account for the role of prices in quantifying the economic impacts of weather events on crop producers. Furthermore, much prior work focuses on ex-ante assessments of climate change impacts on agricultural outcomes accounting for adaptation, but I focus on ex-post assessments of the impacts of extreme weather events on crop revenues.

Using the Ricardian approach, Mendelsohn et al. (1994) sought to assess the impact of climate change on crop revenues. The approach exploits variation in climate across locations to establish the relationship between crop revenues and climate accounting for farmers' long-run adaptive responses given their local climate. One major drawback of this approach is that parameter estimates could be biased owing to potential omitted variables that are correlated with local climate. In contrast, Deschênes and Greenstone (2007) used a fixed-effects panel approach by exploiting year-to-year variation of weather within locations to understand how weather affects agricultural profits. Deschênes and Greenstone (2007) included year-fixed effects in all of their empirical analyses. Given that fluctuations in crop prices in the US are similar across regions, useful variation needed for accounting for price adjustments in response to production declines is soaked up by year-fixed effects in estimation. Burke and Emerick (2016) has also conducted some analysis of crop revenues using a long-difference approach. Because the approach uses temporally averaged variables, it would not be suitable for quantifying the impacts of an extreme weather event in a particular year on crop revenues.

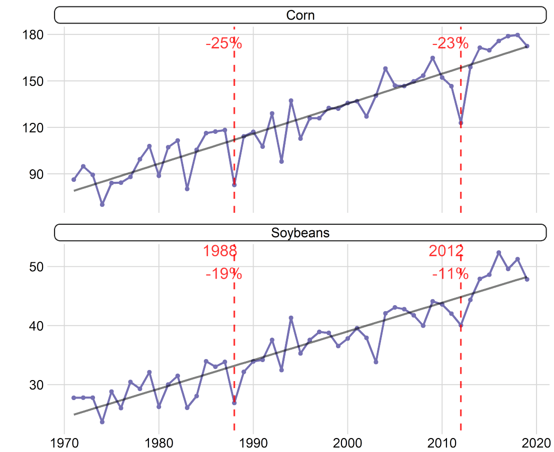
Unlike previous studies, I do not include year-fixed effects in my regression model. Instead, I construct a variable I call the weather-induced nationwide yield shock (WINYS), and include this variable as one of the explanatory variables in models of crop revenue. This variable has two appealing features. First, the coefficient on the variable in the crop revenue equation can be interpreted as the price flexibility of demand conditional on other covariates including local weather variables. Second, the variable can account for heterogeneous responses to weather across regions. The second feature is important in years of extreme weather because the contribution of irrigated counties to crop production tends to be high in those years. For this reason, my estimation includes both irrigated and nonirrigated counties. My preferred specification allows the coefficients on WINYS to vary over space and time. My purpose in allowing for spatially heterogeneous coefficients is to account for the spatially varying degrees of correlation between WINYS and local weather, while temporally varying coefficients would capture the period-specific conditional price flexibilities of demand.

In 1988, the US saw a substantial decrease in crop yields: 29% for corn and 21% for soybeans. Similarly in 2012, the national crop yields dropped by 25% for corn and 12% for soybeans. I estimate that, despite devastating yield losses, national-level crop revenues were less negatively affected because of the induced increases in prices; to be specific, by (-)11% for corn and (+)1% for soybeans in 1988, and (+)11% for corn and 0% in 2012. The major factor that differentiates corn revenue impacts in 1988 and 2012 stems from the fact that my estimates of conditional price flexibilities are much higher in the post-2005 period particularly for corn. To reinforce the idea, I ask what would have happened if the 1988 drought—more devastating in terms of crop yields—occurred in 2012. I estimate that the hypothetical 1988 drought in 2012 would have increased crop revenues by 21% for corn and 4% for soybeans. Last but not least, I also document that, in the two years, regional inequality of crop revenues substantially increased compared to years of normal weather.

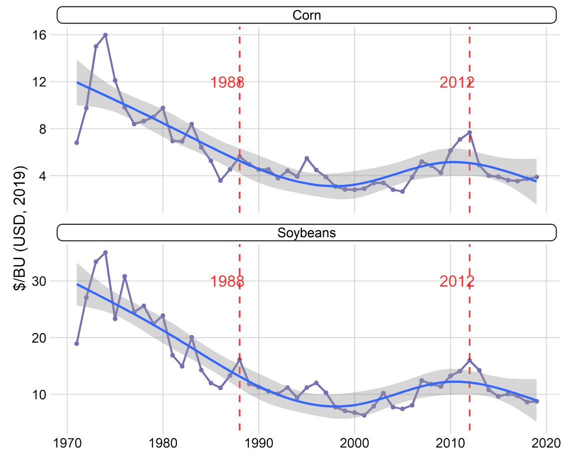
2.2. Background

2.2.1. Definition of Drought.

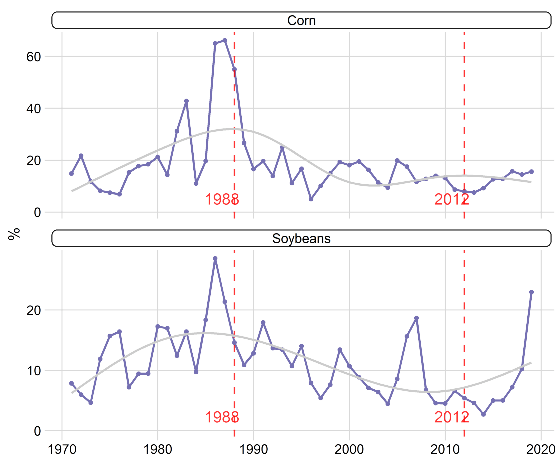
In this chapter, I use the term 'drought' to represent both heat and water stress that are associated with crop yields. Drought is a complex phenomenon and it is difficult to define a meaningful measure of characteristics of droughts in terms of intensity, magnitude, duration, and spatial extent. Not surprisingly, different studies define drought in different ways, making it very difficult to establish a universal drought index (Heim 2002). In the climatology and meteorology literature, drought tends to strictly refer to water deficit. Based on the time scale over which water deficits accumulate, the literature often functionally separates hydrological, environmental, agricultural, and other droughts. In the agronomy literature on crop yields, drought has been used to mean both dry and/or hot weather conditions. Since the seminal work by Schlenker and Roberts (2009), however, it is more common to use heat-related terms, such as heat waves and extreme temperatures particularly when researchers statistically model crop yields. As dryness and hotness are often intertwined, it is not always useful to distinguish the two concepts, at least in the statistical framework of crop yields. In



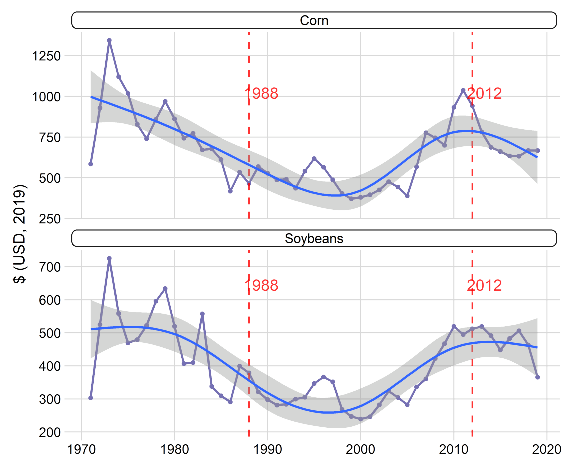
(a) Yield



(b) (market-year) Price



(c) Stocks as Percent of Use



(d) (per-acre) Crop Revenue

FIGURE 2.1. US Time-Series of Agricultural Outcomes for Corn and Soybeans

the statistical framework of crop yields for corn and soybeans in the US, exposure to extreme heat during the growing season is the most influential driver of crop yields and partially picks up the effect of water stress on crop yields when water stress is not included as explanatory variables (Berry et al. 2014; Ortiz-Bobea et al. 2019).

2.2.2. 1988 and 2012 US Midwest Droughts on Corn and Soybean Production.

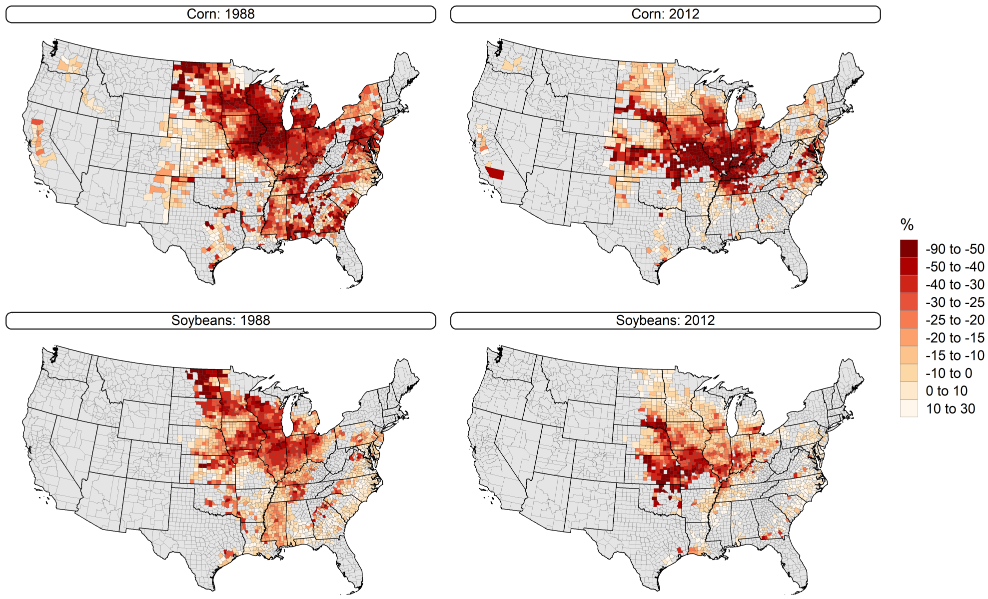
In 1988 and 2012, the US Midwest experienced the most catastrophic droughts over the past half century. Previous studies have documented that the two droughts are comparable in terms of severity and spatial extent (e.g., Karl et al. 2012; Rippey 2015). As figure 2.1a shows, national crop yields, simply measured by deviation from a linear trend, dropped by 25% (23%) for corn and 18% (11%) for soybeans in 1988 (2012). Figure 2.2b shows the spatial distribution of planted area in 1988 and 2012 and, similarly, figure 2.2a shows the spatial distribution of county-level yield shocks measured by a percentage deviation from the linear trend yield. It is noticeable that the two droughts impacted many parts of the US Corn Belt.

As figure 2.1 shows crop prices were above the flexible trend lines for both crops in the two marketing years of 1988 and 2012. It is apparent that, at the beginning of the marketing year of 1988, there were abundant carryover stocks of corn that could buffer a production decline. To be specific, as figure 2.1c shows, stocks as a percent of use were over 50% for corn in 1988 while it was lower than 20% for corn in 2012 as well as soybeans in 1988 and 2012.

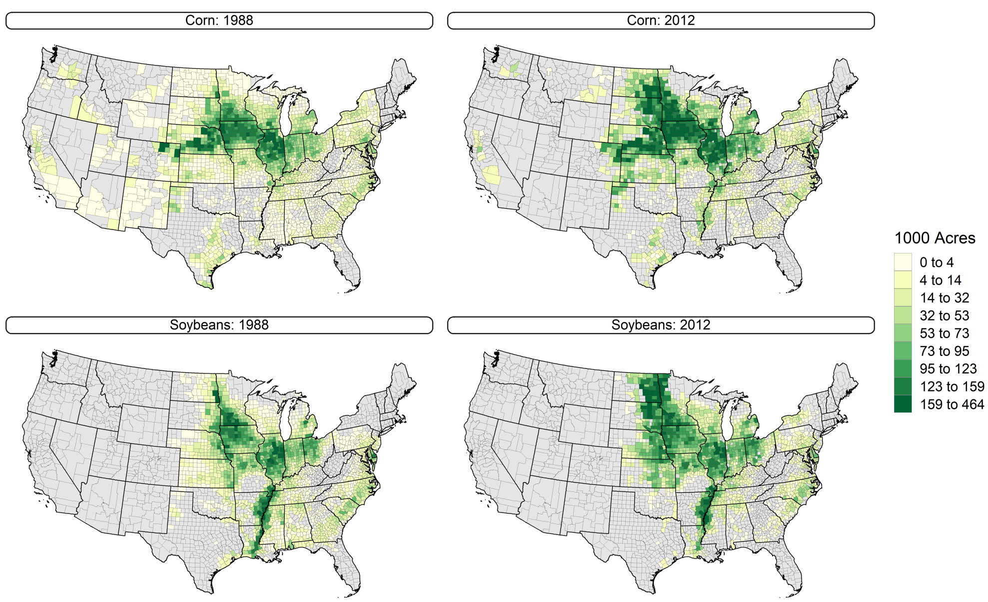
One may wonder whether farmers were able to anticipate the two droughts before planting their crops. I find little evidence that many farmers anticipated the droughts. Figure 2.1 shows the prior years, namely, 1987 and 2011, were normal in terms of crop yields for both crops, suggesting that the two droughts were flash droughts rather than multi-year droughts. Simple exploratory analysis also reveals that planted area did not noticeably drop in the area severely affected by the droughts. A report released in 2012 May by the World Agricultural Supply and Demand Estimates (WASDE) of the US Department of Agriculture (USDA) suggests that the 2012 drought was not foreseeable at least by the agency until a month before the drought started becoming severe. For example, according to the report, the projected corn yield in bushels per acre was 166.0 but the actual yield dropped by 26%, ending up with 123.1. In the same report, the WASDE projected the average prices received by farmers in the 2012/2013 marketing year to be 4.20 – 5.00 $/bu (in nominal value) for corn, but their counterparts ended up with 6.89 $/bu.

2.3. Conceptual Framework

In this section, I present a simple conceptual model to show weather-induced changes in profits are identical to those in crop revenues under some conditions. To be consistent with my empirical setting, I focus on the post-planting profit-maximizing problem associated with crop production in a single crop year. This means that the model also does not consider planting costs (e.g., rents and seeds) but consider within-season input adjustment costs. The following model can be applicable for farmers in both rain-fed and



(a) Yield Shock (deviation from trend)



(b) Planted Acres

FIGURE 2.2. Maps of Planted Acres and Yield Shocks

irrigated regions. To be consistent with my empirical analysis, I do not consider storage decisions, although they are an important means of buffering year-to-year price fluctuations.

Consider a risk-neutral representative crop producer's post-planting *per-acre* profit-maximizing problem:

$$\pi_i\,(I_i; w_i, \mathbf{w_{-i}}) = \max_{I_i} p(q) y_i(I_i(w_i); w_i) - c_i I_i(w_i) \tag{2.1}$$

where p is the producer's expected harvest price and I_i is the input quantity, which is the only decision variable in this model. w_i represents a generic weather variable that positively affects yield y_i during the growing season. $\mathbf{w_{-i}}$ denotes a vector of nonlocal weather associated with the other producers' yields. For simplicity, assume that yield depends only on input and exogenous weather. Assume, for all k, y_k is concave in I_k and w_k, and let c_k denote the input price. I assume that the input price is independent of weather conditions. q is the total quantity supplied. In the context of corn or soybean production in the US, contribution of an individual farmer to the total quantity produced is negligible. That is,

$$\begin{aligned} q &= q_i + \sum_{j \neq i} q_j \\ &\approx \sum_{j \neq i} q_j \\ &= \sum_{j \neq i} a_j y_j(I_j; w_j), \end{aligned}$$

where a_i denotes harvested acreage. Thus, the producer behaves as a price taker. The first order condition then gives $p(\sum_{j \neq i} a_j y_j(I_j; w_j)) \frac{\partial y_i(I_i; w_i)}{\partial I_i} = c_i$. For given w_i and $\mathbf{w_{-i}}$, let $I^*(w_i, \mathbf{w_{-i}})$ maximizes her per-acre profit with the corresponding profit $\pi_i\left(I_i^*(w_i, \mathbf{w_{-i}}); w_i, \mathbf{w_{-i}}\right)$. That is, the optimal input quantity depends on both local and nonlocal weather. To illustrate the link between the optimal input quantity and local weather, consider a farmer with an irrigated farm who would decide how much to irrigate her crop taking into consideration precipitation or water demand (evapotranspiration). Nonlocal weather can affect the optimal input choice indirectly through changes in price. If price is expected to increase due to bad weather in other regions, the marginal revenue of production increases and thereby the optimal input quantity would change.

Now consider effects of a widespread weather event like the 2012 Midwest drought on her profit. Using the Envelope theorem and taking total derivatives gives

(2.2)

$$\Delta\pi_i = \underbrace{y_i \frac{\partial p}{\partial q} \sum_{j \neq i} a_j \left(\frac{\partial y_j}{\partial I_j} \frac{\partial I_j}{\partial w_j} + \frac{\partial y_j}{\partial w_j} \right) \Delta w_j +}_{\text{price component due to changes in } \mathbf{w}_{-i}} \underbrace{p \frac{\partial y_i}{\partial w_i} \Delta w_i}_{\text{yield component}} \underbrace{+ a_i y_i \left(\frac{\partial y_i}{\partial w_i} + \frac{\partial y_i}{\partial I_i} \frac{\partial I_i}{\partial w_i} \right) \frac{\partial p}{\partial q} \left(\sum_{j \neq i} a_j \frac{\partial y_j}{\partial I_j} \frac{\partial I_j}{\partial p} \right) \Delta w_i}_{\text{price component due to changes in } w_i}.$$

(See appendix 2.7 for full derivation.) The first underbrace is the magnitude of (per-acre) profit change through price driven by changes in the other producers' weather conditions. Holding acreage fixed, the direct and indirect effects of nonlocal weather on yields in other regions affect the representative producer's expected price. The second underbrace shows how local weather affects (per-acre) profit through yield. The third underbrace represents a very indirect effect of local weather on price. To be specific, as her local weather changes her crop production (i.e., $\frac{\partial y_i}{\partial w_i} + \frac{\partial y_i}{\partial I_i} \frac{\partial I_i}{\partial w_i}$), this changes other producers' expected price and they adjust their input quantities (i.e., $\sum_{j \neq i} a_j \frac{\partial y_j}{\partial I_j} \frac{\partial I_j}{\partial p}$), which in turn change total market production and price (i.e., $\frac{\partial p}{\partial q}$).

This simple equality provides useful insights into my empirical framework in several ways. In the simple model, a change in (per-acre) profit is identical to a change in (per-acre) revenue if the following conditions hold. First, weather should not directly affect input price. Second, the farmer should be able to adjust their inputs during the growing-season without constraints. In the estimation section, I use (per-acre) revenue as a proxy for (per-acre) crop-specific profit, on which crop-specific panel data are not publicly available at the county level in the US. It is unlikely that the close relationship between revenue and profit holds in the long run, as input prices such as rent can also adjust over time. In addition, if weather directly affects input prices to a large extent, the difference between revenue and profit becomes larger as the Envenlope theorem predicts.

There are a few intuitively apparent but noteworthy points. First, although droughts can reduce a farmer's yield, the resulting loss of income can be offset by an increase in price if competitors in other areas also experience adverse weather. Second, counties with more harvested acreage are more likely to affect price. Lastly, as the first term ($\frac{\partial y_j}{\partial I_j} \frac{\partial I_j}{\partial w_j}$) in the bracket of the first price component suggests, farmers more capable of adjusting within-season input quantities in response to weather shocks are more likely to influence price, *ceteris paribus*. In occurrence of drought, irrigated counties would have more capacity to mitigate negative effects of drought on yields than rainfed counties. Thus, it is important to take into account irrigated counties in the empirical framework.

2.4. Data

The unit of my empirical analysis is crop-by-county-by-year. I use panel data spanning from 1971 to 2019 in my main estimation.

TABLE 2.1. Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Max
(Corn) WINYS	90,477	0.953	0.092	0.711	1.091
(Corn) (per-acre) Revenue ($, 2019)	90,477	589.518	270.518	13.950	2,652.632
(Corn) Yield (bushel/acre)	90,477	104.919	41.341	4.500	270.200
(Corn) Acreage	90,477	41,807.870	52,865.730	10	397,000
(Corn) Stock-to-Use Ratio (SUR)	90,477	0.188	0.135	0.050	0.661
(Corn) (market-year) Price ($/bushel, 2019)	90,477	6.211	3.226	2.284	19.579
Deg. Days (>29C)	90,477	0.596	0.561	0.000	7.131
Deg. Days (10C-29C)	90,477	18.606	4.203	7.371	30.361
Total Precipitation (100mm)	90,477	5.911	1.891	0.027	14.733
(Soybeans) WINYS	71,794	0.972	0.073	0.794	1.144
(Soybeans) (per-acre) Revenue ($, 2019)	71,794	449.328	194.077	5.093	1,465.263
(Soybeans) Yield (bushel/acre)	71,794	33.072	11.043	0.700	80.400
(Soybeans) Stock-to-Use Ratio (SUR)	71,794	0.112	0.056	0.027	0.285
(Soybeans) (market-year) Price ($/bushel, 2019)	71,794	14.564	7.172	5.870	38.421
Deg. Days (>30C)	71,794	0.372	0.353	0.000	2.825
Deg. Days (10C-30C)	71,794	19.233	4.085	7.375	30.407

2.4.1. Agriculture and Weather.

The agricultural data in this study come from the National Agricultural Statistics Service (NASS) of the USDA. I use annual data on county-level yields (bushels per harvested acre) and planted acres, and state-level marketing year prices. When there are missing values in planted acres, I use harvested acres as a proxy for planted acres. I construct county-level (per-acre) revenues by multiplying county-level yields and state-level prices. I use state-level prices because county-level prices are not available for corn and soybeans. Although there can be measurement errors in crop revenues in the dependent variable due to coarseness of price data, the magnitude would be small and parameters of interest will be still consistently estimated.

In my regression models, I control for the stocks-to-use ratio at the beginning of the marketing year, which is from September 1 to August 31 for both corn and soybeans. Stock-to-use ratio is measured by the level of carryover stock for any given commodity as a percentage of the total use over the previous marketing year. In practice, I calculate utilization in a given marketing year t, which is the denominator in the stock-to-use variable, by using the following identity:

$$I_{t-1} + Q_t = U_t + I_t \tag{2.3}$$

$$\Leftrightarrow U_t = Q_t - \Delta I_t \tag{2.4}$$

I_t = US carryover stock in t, Q_t = US production (harvest) in t, and U_t = utilization in t (US consumption + US net exports).

For weather variables, I use county-level data on cumulative precipitation, the square of precipitation, moderate degree days (mdd), and heat degree days (hdd) during the growing season (April–September). Based on the results of Schlenker and Roberts (2009), for corn, I define moderate degree days as degree days from 10°C to 29°C and heat degree days as degree days above 29°C. I similarly define moderate degree days and heat degree days for soybeans but use 30°C as the threshold temperature. To construct these county-level data, I first obtain the gridded dataset with the spatial resolution of 4km from the PRISM (Daly et al. 1997). Following Schlenker and Roberts (2009), I then aggregate weather data to the county level using farmland areas as weights.[1]

2.4.2. Irrigated Acreage.

To explore the potential mitigating effects of irrigation on drought impacts, I use additional data from the USDA NASS that allows me to assign counties to either irrigated or rainfed counties. I follow Ortiz-Bobea et al. (2019) for designating irrigated and rainfed counties. I first calculate the proportion of irrigated harvested acreage for a crop in a county in the US Census of Agriculture for 2002, 2007, and 2012. When missing, I supplement this information with survey data from USDA NASS for 1981–2012. I then take the largest proportion observed for each county and use it as a time-invariant variable that describes the potential for the crop in the county to be irrigated. I define a county to be irrigated if the proportion of irrigated harvested acreage is greater than 50%. When no information about the irrigation proportion is available for a county with the USDA data, I assigned a county to irrigated counties if located to the west of the 100th meridian and to rainfed counties otherwise.[2] Figure 2.3 shows the resulting maps of irrigated and rainfed counties.

[1] I used data and code available at Wolfram Schlenker's webpage (`http://www.columbia.edu/~ws2162/links.html`).
[2] This is because counties west of 100th meridian tend to be highly irrigated.

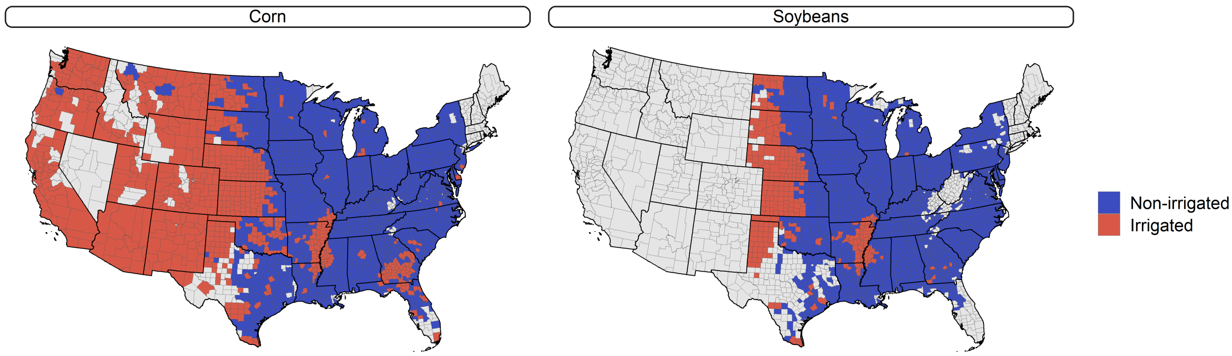


FIGURE 2.3. Irrigation Status by Crop

2.5. Empirical Analysis

2.5.1. Regression Model.

For ease of discussion, I first introduce a simplified version of my regression equation, which does not allow for any coefficient heterogeneity:

$$log(\widetilde{p_{cs(i)t}y_{cit}}) = \beta_c log\left(\widetilde{Q_{ct}}\right) + \Gamma_c X_{cit} + \lambda_c SUR_{ct} + \alpha_{ci} + \epsilon_{cit} \tag{2.5}$$

$\widetilde{Q_{ct}}$ is the weather-induced nationwide yield shock (WINYS) for crop c in time t, which I discuss in detail in the next subsection. $\widetilde{p_{cs(i)t}y_{cit}}$ denotes per-*harvested* acre (hereinafter, per-acre) revenue for crop c in county i of state s in the crop year t, relative to average of per-acre revenues over the previous and subsequent two years (hereinafter, relative revenue): $\frac{p_{cs(i)t}y_{cit}}{\frac{1}{4}\sum_{\tau\in\{-1,-2,1,2\}} p_{cs(i),t-\tau}y_{ci,t-\tau}}$. $p_{cs(i)t}$ is the state-level market-year price and y_{cit} is the crop yield. Thus, variation in the dependent variable comes from deviation of the current-year per-acre revenue from average of temporally nearby per-acre revenues. Such variable construction can be viewed as detrending the dependent variable as a way of removing variation in the dependent variable that is not explained by independent variables. One may also include a flexible time trend on the right-hand side rather than detrending the dependent variable directly. The main reason for detrending the dependent variable using temporally nearby realized values is that, unlike crop yields, per-acre revenues tend to be highly nonlinear as can be seen in figure 2.1d. This implies that, if one fits a flexible time trend, it is not very clear where the residual variation in the dependent variable comes from.

Identifying source of variation is particularly important and relevant when quantifying the drought impacts on crop revenues in the postestimation section.

X_{it} is a vector of local weather variables including $prec_{it}$, $precsq_{it}$, mdd_{it}, hdd_{it}, hdd_Jul_{it}, and hdd_Aug_{it}. $prec_{it}$ is the total precipitation during the growing season (April to September), mdd_{it} is total number of degree days of beneficial heat ($10°C - 29°C$ for corn and $10°C - 30°C$ for soybeans.), hdd_{it} is total number of degree days of extreme heat ($29°C$ for corn and $30°C$ for soybeans). I additionally include total number of degree days of extreme heat in July and in August to parsimoniously allow for within-season time-varying effects of extreme heat (Berry et al. 2014; Ortiz-Bobea et al. 2019). SUR_{ct} is the stock-to-use ratio at the beginning of the crop year t. α_{ci} is the county fixed effects, which will capture time-invariant location characteristics, such as soil types. ϵ_{cit} is the error term.

2.5.2. Weather-Induced Nationwide Yield Shock (WINYS).

The main goal of my empirical strategy is to estimate the impact of local and national weather shocks on crop-specific county-level per-acre crop revenue using panel data. Accounting for price changes in (per-acre) revenue due to weather shocks across a broad area is not straightforward, because price received by farmers in one location is influenced by weather conditions of other areas to a large extent. This is more so in the US, since agricultural transportation is efficient and, thereby, crop markets are highly spatially integrated.

To address this issue, I construct a variable I call the weather-induced nationwide yield shock (WINYS) as follows:

$$\widetilde{Q}_{ct} = \frac{\sum_i a_{cit} y_{cit}}{\sum_i a_{cit} \tilde{y}_{cit}} \tag{2.6}$$

where a_{cit} is the planted acres for crop c in county i in year t, y_{cit} is the realized yield, and $\tilde{y}_{cit}$ is the predicted yield under normal weather conditions, which I will discuss more in detail. The WINYS variable essentially captures crop-specific percentage yield shock weighted by acreage. The WINYS has two main appealing features. First, it permits economically intuitive interpretation. β_c in eq.(2.5) measures the price flexibility of demand, partialling out the effects of local weather in variation of crop revenues (Hereinafter, I call β_c the conditional price flexibility.) This is because variation of price in crop revenues will be captured by WINYS conditional on local weather, while variation of local yield will be explained by local weather. Second, it can account for heterogeneous crop yield responses to weather. This can be done either by simply using realized yields or by using predicted yields from a crop yield regression model that allows for heterogeneity in coefficient on weather variables across subgroups. Crop yield responses are substantially heterogeneous between rain-fed and irrigated counties and between northern and southern states (Berry et al. 2014). This point is important in my empirical setting because the relative contribution of irrigated counties to national production tends to be larger in drought years than in normal years, as figure 2.4 shows.

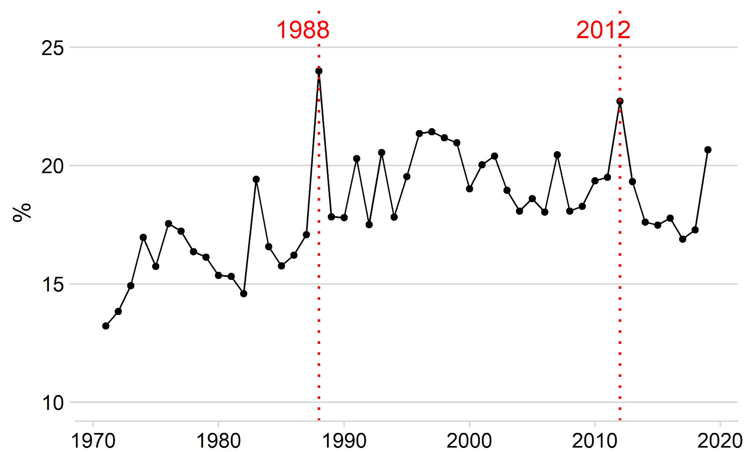


FIGURE 2.4. Proportion of Corn Production by Irrigated Counties

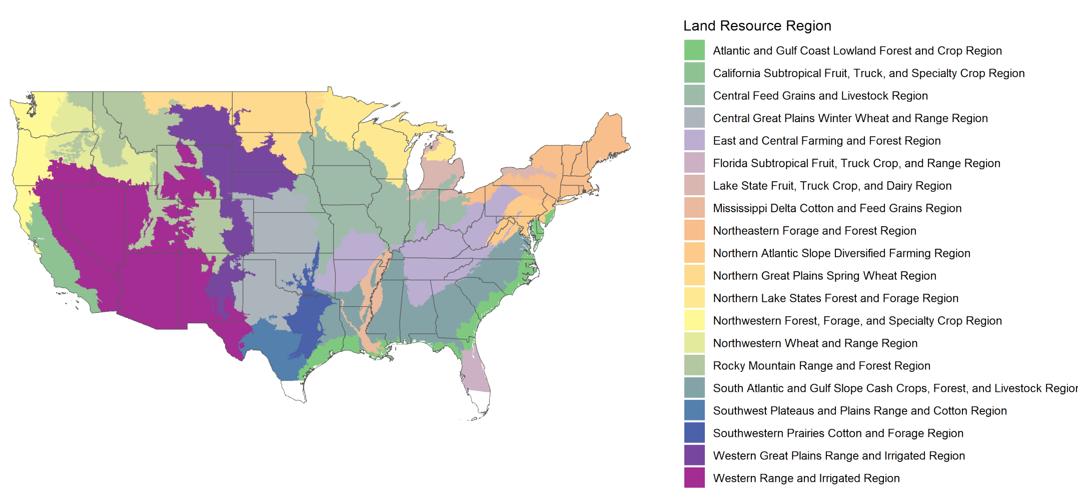


FIGURE 2.5. Land Resource Region
Source: the USDA National Resource Conservation Service (NRCS)

For predicted yields $\tilde{y}_{cit}$ in the denominator of 2.6, I first run the regression of log yield on weather variables:

(2.7)
$$log(y_{cit}) = \beta_{c1} prec_{it} + \beta_{c2} precsq_{it} + \beta_{c3} mdd_{it} + \beta_{c4} hdd_{it} + \beta_{c5} hdd\_Jul_{it} + \beta_{c6} hdd\_Aug_{it} + \alpha_{ci} + f(t) + \epsilon_{it}$$

I allow flexible time trends, $f(t)$, and coefficients on weather variables to vary by irrigation status and Land Resource Region (LRR) shown in figure 2.5. I then obtain predicted values of y_{cit} using the regression results from the estimation of eq.(2.7) combined with crop-specific 20-year rolling average weather for normal weather conditions.

In the numerator of eq.(2.6), I use realized yields rather than fitted yields because the panel approach in eq.(2.7) does not always fit well. For example, in 1993 when there was a historic flooding in parts of the US Corn Belt, there is a substantial discrepancy between two types of WINYS (see figure 2.6): one that used realized yield and the other that used fitted yields under observed weather in the numerator of WINYS. Fortunately, the model in eq.(2.7) fits reasonably well in the two drought years at least at the aggregate level, as figure 2.6 shows.

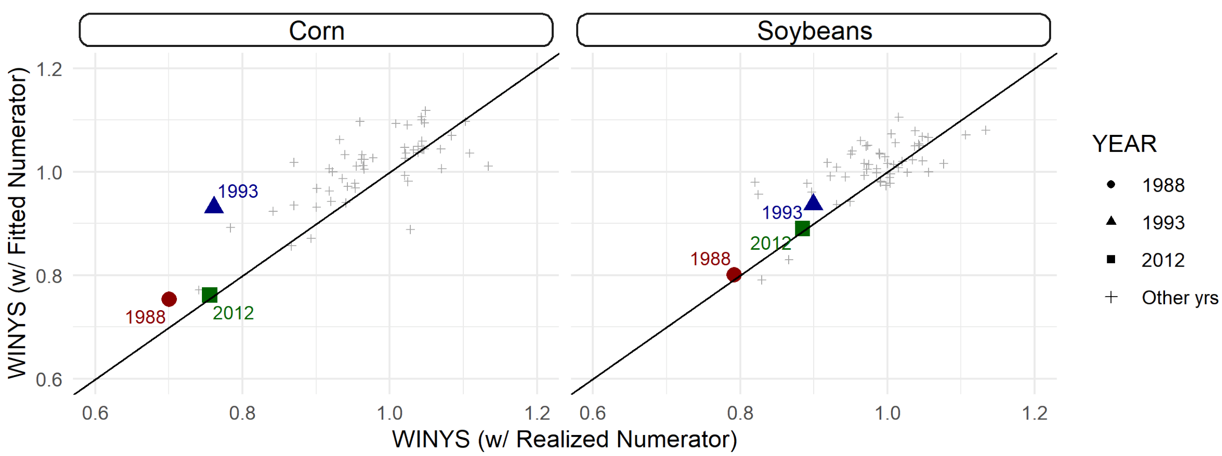


FIGURE 2.6. WINYS

2.5.3. Natural Hedge and WINYS.

The main motivation for estimating (per-acre) revenue rather than estimating price and yield separately is to account for the correlation between price and yield. As the US is one of the major corn and soybean producers, the correlation between price and yield for the two crops is negative in many parts of the Corn Belt. Such negative correlation is often called as ”natural hedge” (Du et al. 2014; Harwood 1999; Zulauf et al. 2016). Prior knowledge about how the natural hedge varies across space provides useful insights into developing my preferred specification and is an integral part of understanding what determines the spatial distribution of crop revenue variability.

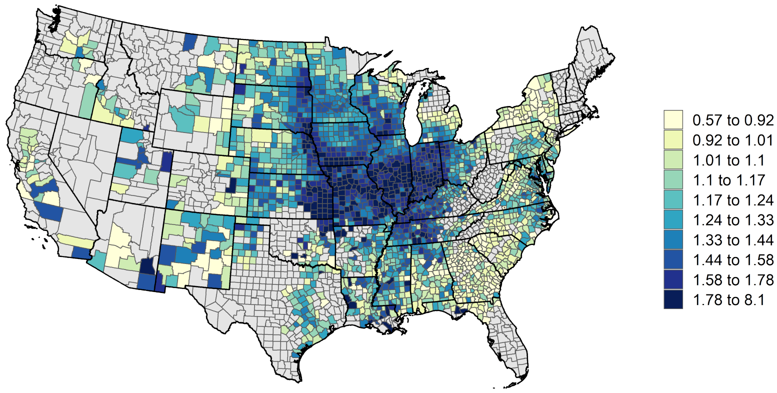
To draw key insights into the role of the natural hedge in relation to WINYS, I conduct analysis of (per-acre) revenue variability decomposition using the time series data on national (market-year) corn price and county-level corn yields spanning from 1971 to 2019. Variance of per-acre revenue is determined by variance of price, variance of yield, and covariance of price and yield. Each time-series of price, yield, and per-acre revenue has its own time trend over the period. As a time trend can distort the actual variance leading to a misleading conclusion, I work with detrended log price and detrended log yield. I approximate variance of detrended log per-acre revenue as follows: $var(\widehat{p_t y_t}) = var(\widehat{p_t}) + var(\widehat{y_t}) + 2cov(\widehat{p_t}, \widehat{y_{it}})$, where $\widehat{A}$ denotes detrended $log(A)$.

Figure 2.7b shows the covariance between national price and county yield for corn over 1971–2019: $cov(\widehat{p_t}, \widehat{y_{it}})$. It is clear that, in the central part of the US, the natural hedge is strong. Figure 2.7a shows that the sum of variance of price and variance of yield divided by variance of per-acre revenue $\frac{var(\widehat{p_t})+var(\widehat{y_{it}})}{var(\widehat{p_t y_{it}})}$. The figure suggests that, for half of the corn-growing counties in the US, revenue variability would have been at least 24% higher had not been the natural hedge, simply measured by $cov(\widehat{p_t}, \widehat{y_{it}})$.

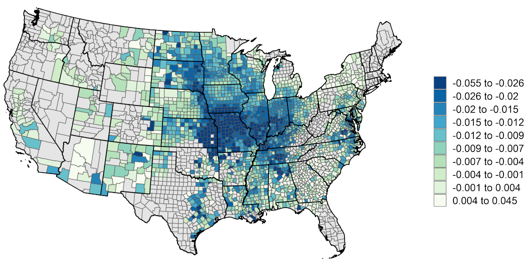
Here, I show that there is a close connection between natural hedge and the covariance between national and county yield. If we express the national corn price as a function of national corn production as $p_t = AQ_t^{\theta_1} e^{\eta_t}$, then we have $log(p_t) = \theta_0 + \theta_1 log(Q_t) + \eta_t$, where $\theta_0 \equiv log(A)$ and η_t is the idiosyncratic error term. If $E[\eta_t log(y_{cit})] = 0$ and we ignore any time trends, we have

$$cov(log(p_t), log(y_{cit})) = \theta_1 cov(log(Q_t), log(y_{cit})). \tag{2.8}$$

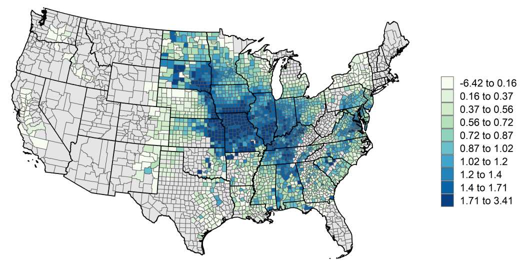
This simple expression provides us two valuable insights into my empirical strategy. First, the larger θ_1 in magnitude is, the higher the degree of natural hedge will be. θ_1 captures the price flexibility of demand (Adjemian and Smith 2012; Moore 1919). Second, the natural hedge would be higher in magnitude in the regions where local yield shock is highly correlated with national yield shocks (weighted by acreage). Figure 2.7c shows county-specific coefficients from the regression of county-level log yield shock on national-level



(a) $\frac{var(\widetilde{p_t})+var(\widetilde{y_{it}})}{var(\widetilde{p_t y_{it}})}$



(b) $cov(\widetilde{p_t}, \widetilde{y_{it}})$



(c) Coefficients on WINYS

FIGURE 2.7. Crop Revenue Variability Decomposition

Notes: $\widetilde{A}$ denotes detrended log A. The figure uses data spanning from 1971-2019 for corn. p_t is a market-year US corn price. y_{it} is a county-level corn yield.

log yield shock:

$$\log(y_{it}/\hat{y}_{it}) = \psi_{0i} + \psi_{1i} log(Q_t/\hat{Q}_t) + e_{it}, \tag{2.9}$$

where $Q_t/\hat{Q}_t (\equiv \sum_i a_{cit} y_{cit} / \sum_i a_{cit} \hat{y}_{cit})$ is the WINYS for corn in eq.(2.6). As eq.(2.8) suggests, figures 2.7b and 2.7c show similar spatial patterns meaning that the natural hedge is stronger in the region where local yield shocks are highly correlated with national yield shocks. Figure 2.8 shows that, for a given yield variability, revenue variability is smaller if local yield shocks are highly correlated with national yield shocks (i.e., high $\hat{\psi}_{1i}$).

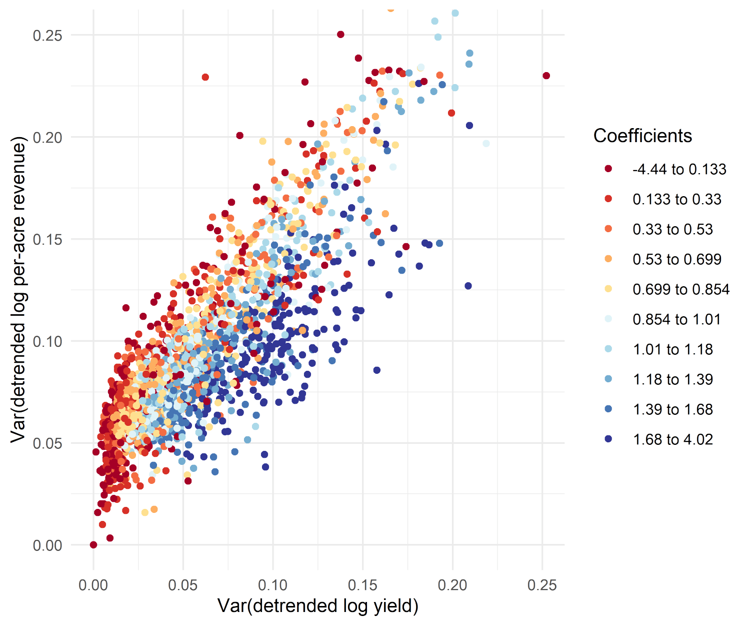


FIGURE 2.8. Revenue vs Yield Variability

Notes: The figure uses data on national corn price and county-level corn yield data from 1971 to 2019. The figure plots county-specific variance of detrended log yield and variance of detrended log per-acre revenue. Color represents county-specific coefficients from the regression in eq.(2.9).

2.5.4. Heterogeneity in Coefficient on WINYS.

My preferred specification allows coefficients on $log\left(\widetilde{Q_{ct}}\right)$ to vary over space and time. As figure 2.7c suggests, the correlation between local and national yield shocks varies across areas. In the context of the regression model in eq.(2.5), this implies that different regions would have a different correlation between the national variable $log\left(\widetilde{Q_{ct}}\right)$ and local weather variables X_{cit} and thus the coefficient on $log\left(\widetilde{Q_{ct}}\right)$ would vary across regions. To allow for heterogeneity of coefficient over space, I group counties into relatively homogeneous groups using the geographic boundaries used to define the Major Land Resource Area (MLRA) from the United States Department of Agriculture (USDA) Natural Resources Conservation Service (NRSC). Each MLRA shares similar soil and climate characteristics. Some MLRAs are small encompassing only a few counties. If a MLRA contains fewer than 10 counties, I spatially merge these counties to a nearby large MLRA that contains more than 10 counties. Figure 2.9 shows the final region groups across which coefficients on $log\left(\widetilde{Q_{ct}}\right)$ are allowed to vary.

A time-invariant coefficient on log WINYS assumes constant price flexibility of demand for corn and soybeans. However, economists have documented that agricultural prices started showing unusual behaviors

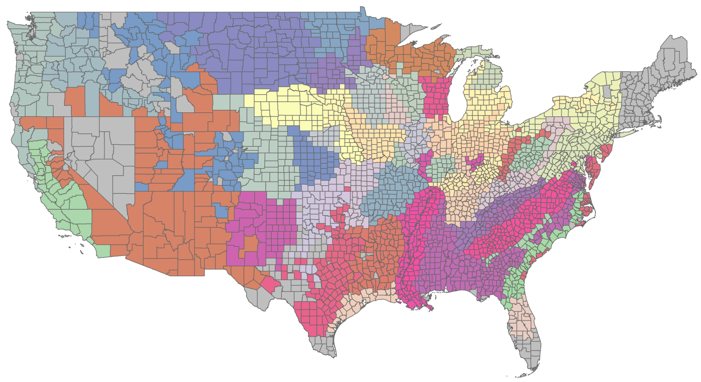


FIGURE 2.9. Region Groups across Which Coefficients on WINYS Vary

since the mid 2000s: high crop prices, high volatility, and a puzzling relationship between stock-to-use ratios and crop prices, to name a few (Carter et al. 2017; Etienne et al. 2015; Wright 2014). I do not attempt to delve into how behaviors of agricultural prices changed and what caused these changes. However, it would seem to be worthwhile to allow for potential heterogeneity of coefficient before and after the mid 2000s. In practice, I allow coefficients on log WINYS to vary between the pre-2005 and post-2005 periods.

2.5.5. Results.

2.5.6. Regression Results.

Table 2.2 shows the regression results of the estimation of eq.(2.5) by parsimoniously allowing for coefficient heterogeneity. Columns 1 and 3 are for corn and columns 2 and 4 are for soybeans. Column 1 (2) indicates that, for an average US corn (soybean) county, a one percent increase in WINYS leads to a decrease in per-acre crop revenue by 0.5% (1%) for corn (soybeans) conditional on local weather and SUR. This result suggests that it is important to account for price responses to widespread production shocks if one's interest is on the effects of weather events on crop revenues in crop-growing regions. Given that the demand for crops is inelastic in general, the smaller-than-unity magnitude of the coefficients on log WINYS might appear counter-intuitive. However, because, for many parts of the study region, local weather and WINYS are correlated, the coefficients on log WINYS do not fully reflect the relationship between price and production. Besides the estimation results on WINYS, the coefficients on local weather variables agronomically make

TABLE 2.2. Regression of Crop Revenue on WINYS and Local Weather Variables

	Log of (per-acre) Crop Revenue			
	(1) Corn	(2) Soybeans	(3) Corn	(4) Soybeans
log(WINYS)	-0.4778**	-0.9406***		
	(0.2052)	(0.2620)		
log(WINYS) × I(YEAR≤2005)			-0.3336	-0.9134***
			(0.2185)	(0.2753)
log(WINYS) × I(YEAR>2005)			-1.533***	-1.114*
			(0.2999)	(0.6189)
SUR	-0.2091	-0.5567*		
	(0.1631)	(0.3317)		
SUR × I(YEAR≤2005)			-0.1606	-0.5382*
			(0.1725)	(0.3131)
SUR × I(YEAR>2005)			-0.5814	-0.7320
			(0.3926)	(0.7275)
prec [rainfed]	0.1430***	0.1078***	0.1510***	0.1093***
	(0.0384)	(0.0231)	(0.0350)	(0.0220)
prec [irrigated]	-0.0291	0.0706	-0.0223	0.0732
	(0.0381)	(0.0488)	(0.0373)	(0.0471)
precsq [rainfed]	-0.0091***	-0.0068***	-0.0096***	-0.0068***
	(0.0027)	(0.0017)	(0.0025)	(0.0016)
precsq [irrigated]	0.0031	-0.0032	0.0028	-0.0034
	(0.0029)	(0.0039)	(0.0028)	(0.0038)
mdd [rainfed]	0.0275	0.0279*	0.0236	0.0301*
	(0.0197)	(0.0165)	(0.0198)	(0.0156)
mdd [irrigated]	0.0230	0.0305*	0.0172	0.0328*
	(0.0220)	(0.0181)	(0.0209)	(0.0173)
hdd [rainfed]	-0.6967***	-0.6350***	-0.6942***	-0.6401***
	(0.1618)	(0.1647)	(0.1600)	(0.1657)
hdd [irrigated]	-0.4092***	-0.3547**	-0.4166***	-0.3622**
	(0.0954)	(0.1686)	(0.0949)	(0.1705)
hdd_Jul [rainfed]	-0.0016	0.0033	-0.0028	0.0031
	(0.0025)	(0.0029)	(0.0026)	(0.0028)
hdd_Jul [irrigated]	0.0014	0.0016	0.0012	0.0015
	(0.0012)	(0.0029)	(0.0013)	(0.0028)
hdd_Aug [rainfed]	0.0078***	0.0002	0.0085***	0.0005
	(0.0029)	(0.0028)	(0.0028)	(0.0028)
hdd_Aug [irrigated]	0.0047**	0.0004	0.0049***	0.0006
	(0.0018)	(0.0025)	(0.0018)	(0.0026)
Observations	88,076	69,903	88,076	69,903
R^2	0.20392	0.20292	0.22619	0.20403
Within R^2	0.20078	0.20060	0.22314	0.20172
County fixed effects	✓	✓	✓	✓

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Notes: The table shows the results from the estimation of eq.(2.5). Standard errors were clustered by years.

sense. For example, extreme heat has strong negative effects and the magnitudes are higher for rainfed counties.

Columns 3 and 4 suggest that responsiveness of price to production shocks was quite different before and after 2005 particularly for corn. For both crops, the coefficient on log WINYS is greater than one in magnitude for the post-2005 period. This means that, if a county experienced adverse weather shock that led to a negative yield shock equivalent to the national average yield shock, then that county's crop revenue would increase because price increases by more than yield drops.

The magnitude of the coefficients on log WINYS during the pre-2005 period is much higher in magnitude for soybeans than that for corn (-0.91 for soybeans and -0.32 for corn). One possible explanation is that, unlike for corn, the contribution of US to the global soybean supply used to be much larger than recent decades. For example, the US produced 58% of the world's soybeans in 1980 whereas the US proportion dropped to 35% in 2010.

In what follows, I present results from the regression of the form in eq.(2.5) allowing for coefficient heterogeneity to a greater extent. To be specific, I allow the coefficients on log WINYS to vary across MLRAs and between the periods before and after 2005. I also allow the coefficients on local weather variables to vary by irrigation status and LRRs. Lastly, I allow the coefficients on stock-to-use ratios to vary between the periods before and after 2005. Figure 2.10 shows the region-specific coefficients on log WINYS. Again, coefficients tend to be larger in magnitude for the post-2005 period for both crops. It is also noticeable that there is substantial heterogeneity of coefficients over space. This spatial pattern is driven by the fact that different regions show different degrees of correlation between local weather and WINYS. For example, for corn in the post-2005 period, the coefficient on log WINYS is close to 0 in southern Illinois but most of outer areas (e.g., North Dakota, Minnesota, or states along the coastal area) have coefficients greater than 1 in magnitude.

2.5.7. Postestimation.

In this section, I quantify the impacts of the 1988 and 2012 droughts on crop revenues. To do so, I use the regression results of my preferred specification, which allows for heterogeneous coefficients on log WINYS and weather variables among counties. My initial step is to construct measures of counterfactual revenues for each county and crop. For WINYS, I use the average of WINYS for each

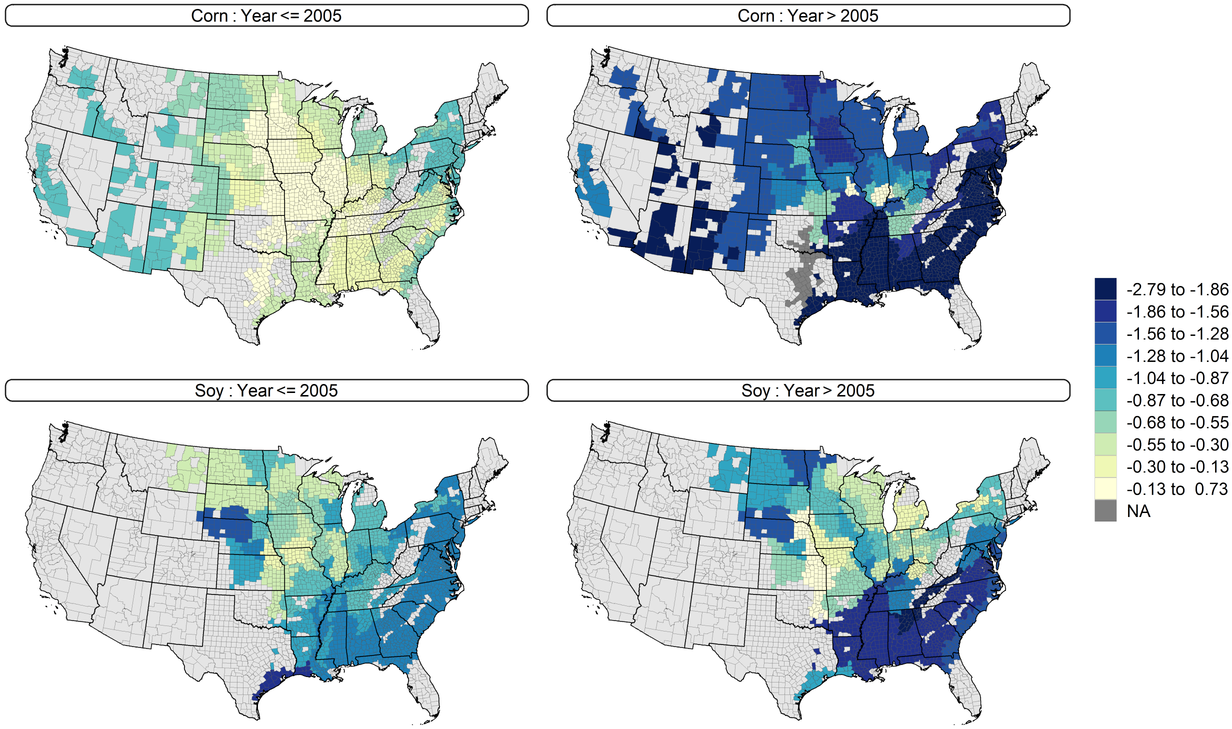


FIGURE 2.10. Coefficients on log(WINYS)

Notes: The figure shows coefficients on log WINYS from the regressions of the form in eq.(2.5).

crop from 1971 to 2019. The averages for both corn and soybeans are close to 1. For counterfactual weather, I use county-level rolling averages of the previous 20-year weather realizations. Applying the estimated parameters to these counterfactual and realized values of 1988 and 2012, and taking exponential transformation of the predicted dependent variables[3], I obtain predicted values of relative revenues ($i.e.,$ predicted values of $p_{\widetilde{cs(i)t}y_{cit}}$ $\left(\equiv \frac{p_{cs(i)t}y_{cit}}{\frac{1}{4}\sum_{\tau\in\{-1,-2,1,2\}} p_{cs(i),t-\tau}y_{ci,t-\tau}}\right)$) for corn and soybeans in 1988 and 2012 under observed and counterfactual scenarios. Let $p_{\widetilde{cs(i)t}y_{cit}}{}^{O}$ and $p_{\widetilde{cs(i)t}y_{cit}}{}^{C}$ denote predicted values of relative revenues under observed and counterfactual scenarios, respectively. I estimate county-level percentage per-acre revenue impacts by $\left(\frac{p_{\widetilde{cs(i)t}y_{cit}}{}^{O}}{p_{\widetilde{cs(i)t}y_{cit}}{}^{C}} - 1\right) \times 100$. County-level revenue changes are computed using temporally nearby crop revenues as the base: $a_{cit} \times \left(\frac{p_{\widetilde{cs(i)t}y_{cit}}{}^{O}}{p_{\widetilde{cs(i)t}y_{cit}}{}^{C}} - 1\right) \times \frac{1}{4}\sum_{\tau\in\{-1,-2,1,2\}} p_{cs(i),t-\tau}y_{ci,t-\tau}$. I compute national-level revenue changes simply by aggregating the county-level counterparts to the US.

[3]Predicted relative revenues are computed as $exp(log(p_{\widetilde{cs(i)t}y_{cit}}) + \frac{\sigma^2}{2})$ to account for the convexity of the exponential function

Using averages of temporally nearby crop revenues as weights, I also obtain national-level percentage revenue impacts by $\left(\frac{\sum_i \widetilde{p_{cs(i)t}y_{cit}}^{O} a_{cit} \sum_{\tau\in\{-1,-2,1,2\}} p_{cs(i),t-\tau}\, y_{ci,t-\tau}}{\sum_i \widetilde{p_{cs(i)t}y_{cit}}^{C} a_{cit} \sum_{\tau\in\{-1,-2,1,2\}} p_{cs(i),t-\tau}\, y_{ci,t-\tau}} - 1\right)\times 100.$

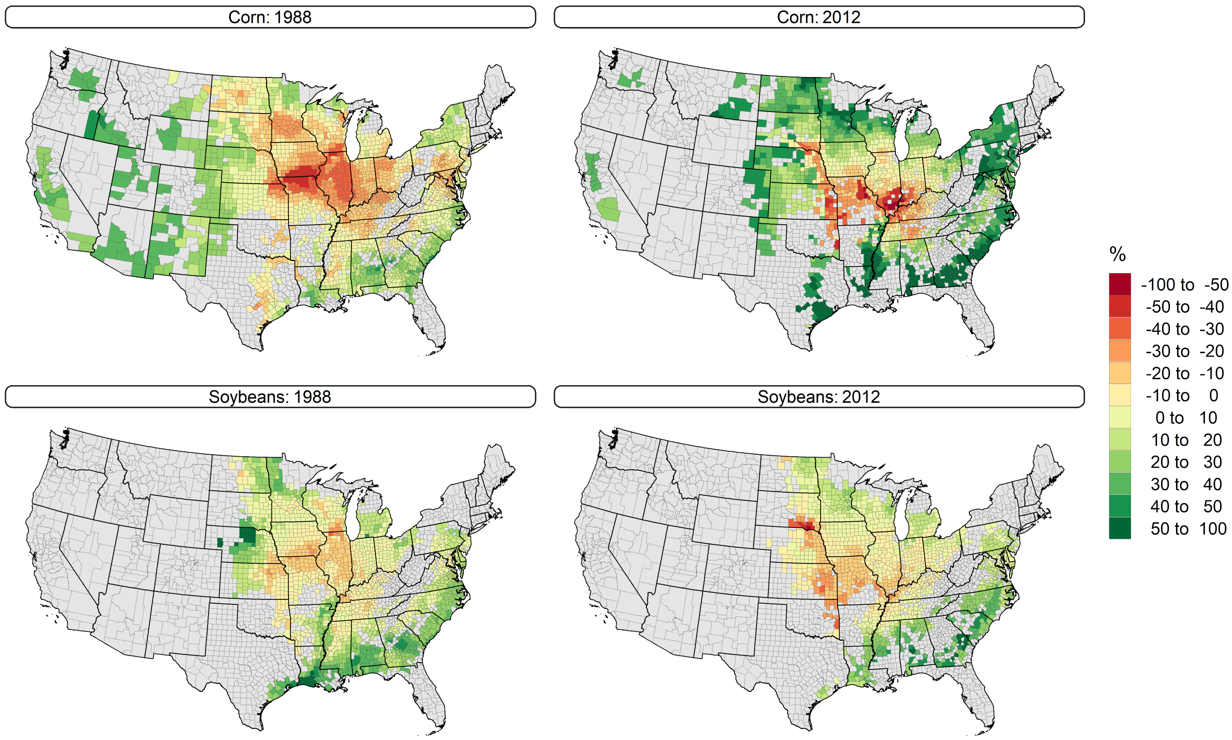


FIGURE 2.11. (per-acre) Crop Revenue Impact (%)

Notes: The figure shows estimated county-level percentage per-acre revenue impacts: $\left(\frac{\widetilde{p_{cs(i)t}y_{cit}}^{O}}{\widetilde{p_{cs(i)t}y_{cit}}^{C}} - 1\right) \times 100$, where $\widetilde{p_{cs(i)t}y_{cit}}^{O}$ and $\widetilde{p_{cs(i)t}y_{cit}}^{C}$ denote predicted values of relative revenues ($i.e.$, $\widetilde{p_{cs(i)t}y_{cit}} \left(\equiv \frac{p_{cs(i)t}\,y_{cit}}{\frac{1}{4}\sum_{\tau\in\{-1,-2,1,2\}} p_{cs(i),t-\tau}\, y_{ci,t-\tau}}\right)$) under observed and counterfactual scenarios, respectively.

Figure 2.11 shows the spatial distribution of drought-induced percentage changes in per-acre revenue. The figure indicates that, although counties impacted heavily by droughts experienced negative revenue impacts, other regions tend to benefit indirectly from droughts due to a higher price received. Table 2.3 summarizes the impacts of the 1988 and 2012 droughts. At the national level, the percentage impacts on crop revenues are estimated to be -11% (-3.83 billion US dollars) for corn and 1% (.26 billion US dollars) for soybeans in 1988, whereas they were 11% (8.80 billion US dollars) for corn and 0% (.16 billion US dollars)

TABLE 2.3. Impacts of the 1988 and 2012 US Midwest Droughts

	Year	Crop	Yield (%)	Revenue (%)	Billion $ (2019)	% of Counties w/ (+)
1	1988	corn	-28.98	-11.45	-3.83	44.31
2	2012	corn	-25.33	10.94	8.80	70.66
3	1988	soybeans	-20.69	1.01	0.26	62.05
4	2012	soybeans	-11.54	0.29	0.16	58.00

Notes: Yield (%) is the national-level percentage yield shock. Revenue (%) is the estimated national-level percentage revenue impacts. Billion $ (2019) is the national-level crop revenue impacts measured in 2019 US billion dollars. % of Counties w/ (+) is the proportion of counties that earned a higher crop revenue than what it would have earned under its counterfactual normal weather.

for soybeans in 2012. The results reveal that the 1988 drought made the US corn farmers worse off as a whole but the 2012 counterpart made them better off. This result is not surprising given that the conditional price flexibility was much higher in magnitude for the post-2005 period (see table 2.5). In 1988, less than one-half of corn-growing counties earned a higher per-acre revenue than what they would have earned under normal weather conditions but, in 2012, more than 70% of the counties did so. Such temporally different revenue impacts did not happen for soybeans, as the conditional price flexibility of soybeans has remained stable.

TABLE 2.4. Impacts of the Hypothetical 2012 US Midwest Drought Identical to the 1988's

	Year	Crop	Yield (%)	Revenue (%)	Billion $ (2019)	% of Counties w/ (+)
1	1988	corn	-28.98	-11.45	-3.83	44.31
2	2012	corn	28.98	21.24	16.73	84.04
3	1988	soybeans	-20.69	1.01	0.26	62.05
4	2012	soybeans	-20.69	3.72	1.65	61.51

Notes: Yield (%) is the national-level percentage yield shock. Revenue (%) is the estimated national-level percentage revenue impacts. Billion $ (2019) is the national-level crop revenue impacts measured in 2019 US billion dollars. % of Counties w/ (+) is the proportion of counties that earned a higher crop revenue than what it would have earned under its counterfactual normal weather.

Figure 2.12 presents comparisons between predicted relative revenues under observed and counterfactual normal weather at the county level. While normal relative revenues are concentrated around 1, there is substantial variation in the predicted relative revenues under observed weather. Figure 2.13 shows the percentage per-acre revenue impacts plotted against the percentage yield impacts at the county level. The figure suggests that, for a given yield impact, percentage revenue impacts tend to be much higher in 2012 for corn but such a pattern does not appear to be the case for soybeans.

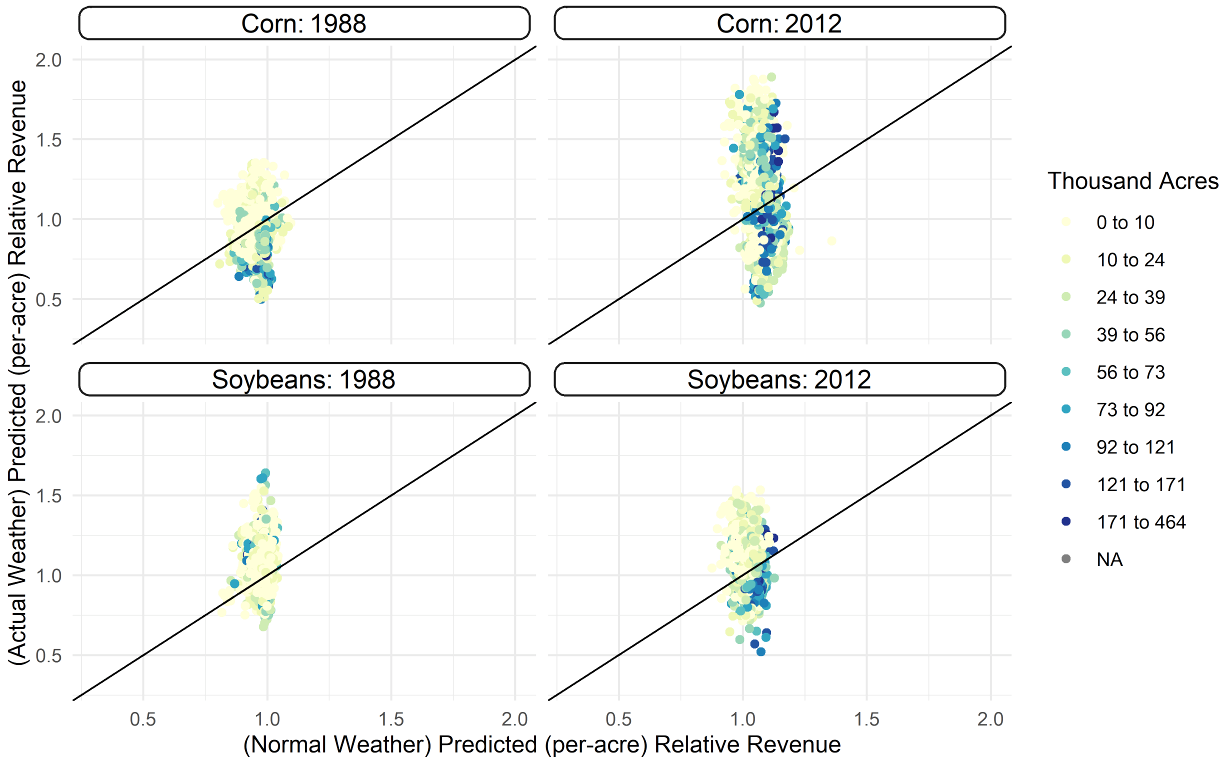


FIGURE 2.12. Predicted County-Level (per-acre) Revenues (Realized vs Normal Weather)

Next, I ask what would have happened if the 1988 drought happened in 2012. I assume that counties experienced local weather in 2012 identical to that of 1988 and the same level of WINYS in 1988 is also realized in 2012. Table 2.4 summarises the results of this exercise. The results for corn suggest that if the 1988 drought happened in 2012, the national-level revenue impacts would have been much higher in 2012 than in 1988 (21% vs -11%) and 84% of corn-growing counties would have earned a higher revenue than their counterfactual revenues whereas only 44% of them did so in 1988. This exercise underscores the importance of the role of market conditions in determining farmers' crop revenues.

Finally, I document that extreme weather events in crop-growing regions can lead to regional inequality in terms of crop revenue impacts. To formally present this idea, I first calculate relative revenue shocks $\left(\frac{\widetilde{p_{cs(i)t}y_{cit}}^{O}}{\widetilde{p_{cs(i)t}y_{cit}}^{C}} - 1\right) \times 100$ for all years in 1971–2019. For each crop and year, I then compute the difference between the top and bottom quintile of the percentage county-level per-acre revenue impacts and use this statistic as the regional inequality measure. In normal years for which I simply exclude major drought years (1970, 1974, 1980, 1983, 1988, and 2012), the difference is less than 20 percentage points (pp.) for both

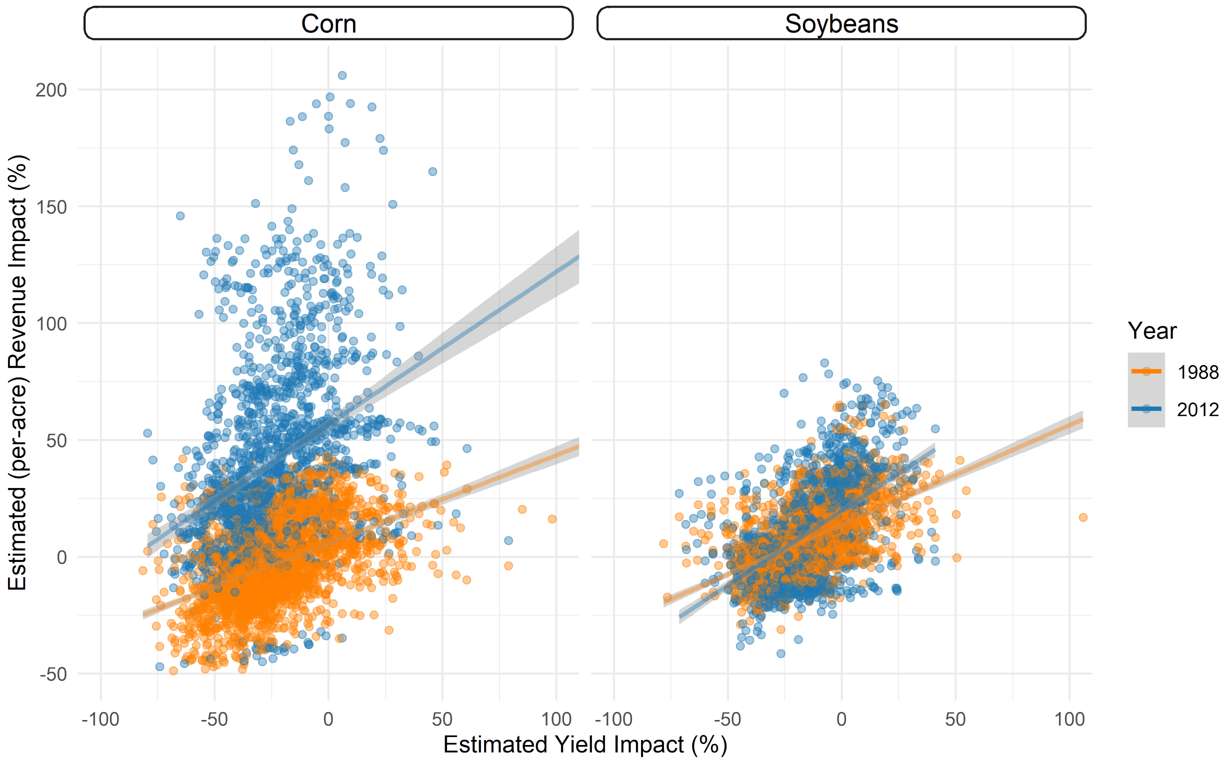


FIGURE 2.13. (county-level) Yield Yield Impact (%) vs Estimated (per-acre) Revenue Impact (%)

crops: 18 pp. for corn and 15 pp. for soybeans. This means that, if the top quintile county earns 10% more per-acre revenue than what it would have earned under normal weather, the bottom quintile county would earn more than 90% of its counterfactual revenue. In drought years, however, the quintile difference increases to 36 pp. for corn and 28 pp. for soybeans on average. Figure 2.14 visualizes the distribution of yield and per-acre revenue impacts for some relatively normal years: two drought years (1988 and 2012) and three relatively normal years (2013, 2014, and 2015).

2.6. Conclusion

What are the economic costs of large-scale droughts to crop producers? The main goal of this paper is to propose a panel approach to estimating the impacts of large-scale weather events in major crop-producing regions on crop revenues at the local and national levels. In these areas, crop prices received by farmers are often negatively correlated with local yield shocks. This paper provides a new way of accounting for the correlation between price and yield in estimating weather-induced crop revenue impacts in a econometric

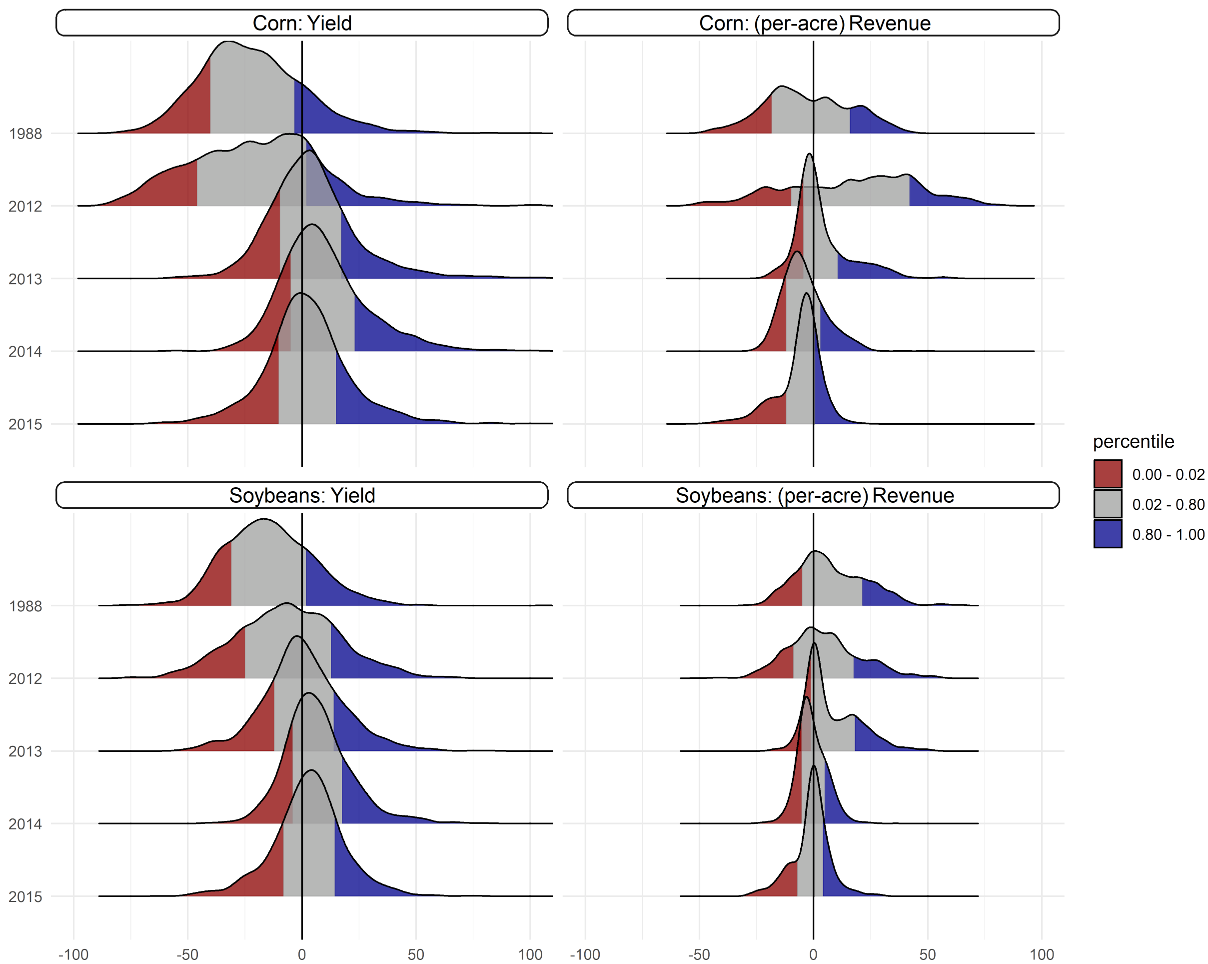


FIGURE 2.14. Distribution of (county-level) Yield and per-acre Revenue Impacts (%)

panel setting. My approach utilizes the fact that the spatially varying relationship between price and yield is largely determined by the extent to which a local yield shock is correlated with the corresponding national-level yield shock.

In the conceptual model, I first showed that weather-induced changes in post-planting crop revenues could be a good proxy for weather-induced changes in post-planting crop profits. The model reveals that inclusion of irrigated counties in empirical analysis is particularly important if drought years are of the researcher's main interest. My crop revenue variability decomposition analysis underscores the importance of accounting for temporally and spatially varying degree of conditional price flexibilities of demand.

After proposing the method, I quantified the impacts of the 1988 and 2012 US Midwest droughts on corn and soybean revenues at the county and national levels. I find that despite large yield shocks, price increases in response to widespread yield shocks tend to offset negative yield shocks. These price increases indirectly make better off farmers who did not experience droughts. I estimate crop revenue impacts of the two droughts to be -11% (11%) in 1988 and 1% (0%) in 2012 for US corn (soybeans). I also showed that extreme weather events in major crop-growing regions tend to worsen the regional inequality of crop revenues relative to normal years.

I leave the following tasks for future work. i) Although the focus of this paper is welfare effects to producers, it would be worth quantifying welfare losses to consumers from the higher prices and lost productions. ii) I will try additional empirical specifications. One of the examples is adding the interaction of SUR and WINYS in eq.(2.5). iii) Although the 1988 and 2012 droughts were flash droughts, drawing implications of consecutive extreme droughts would be also useful given that frequency of these events would increase in a changing climate.

2.7. Supplementary Section: Derivation

In the following, I derive $\frac{\partial \pi_i}{\partial w_j}$ and $\frac{\partial \pi_i}{\partial w_i}$, which are essential components in eq.(2.2). I assume a representative farmer i behaves as a price taker so that she considers $q = \sum_{j \neq i} a_j y_j$. Taking total derivatives implies

$$\Delta \pi_i = \sum_{j \neq i} \frac{\partial \pi_i}{\partial w_j} \Delta w_j + \frac{\partial \pi_i}{\partial w_i} \Delta w_i.$$

Taking derivative of π_i with respect to w_j gives

$$\begin{aligned}
\frac{\partial \pi_i}{\partial w_j} &= \frac{\partial p}{\partial q}\frac{\partial q}{\partial w_j} y_i + p\frac{\partial y_i}{\partial w_j} - c_i \frac{\partial I_i}{\partial w_j} \\
&= y_i \frac{\partial p}{\partial q} a_j \left(\frac{\partial y_j}{\partial I_j}\frac{\partial I_j}{\partial w_j} + \frac{\partial y_j}{\partial w_j} \right) + \left(p\frac{\partial y_i}{\partial I_i} - c_i \right) \frac{\partial I_i}{\partial p}\frac{\partial p}{\partial q} a_j \left(\frac{\partial y_j}{\partial w_j} + \frac{\partial y_j}{\partial I_j}\frac{\partial I_j}{\partial w_j} \right) \\
&= y_i \frac{\partial p}{\partial q} a_j \left(\frac{\partial y_j}{\partial I_j}\frac{\partial I_j}{\partial w_j} + \frac{\partial y_j}{\partial w_j} \right).
\end{aligned}$$

The first equality says that w_j affects q, y_i, and I_i indirectly. If we specify each channel using the chain rule, we get the second equality. The third equality holds because the second term in the second equality is zero at the optimal level of input $I_i^*(w_i; \mathbf{w_{-i}})$.

Taking derivative of π_i with respect to w_i gives

$$\begin{aligned}
\frac{\partial \pi_i}{\partial w_i} &= \frac{\partial p}{\partial q}\frac{\partial q}{\partial w_i} y_i + p\frac{\partial y_i}{\partial w_i} + \left(p\frac{\partial y_i}{\partial I_i} - c\right)\frac{\partial I_i}{\partial w_i} \\
&= y_i \frac{\partial p}{\partial q}\left(\sum_{j\neq i} a_j \frac{\partial y_j}{\partial I_j}\frac{\partial I_j}{\partial p}\frac{\partial p}{\partial q} a_i \left(\frac{\partial y_i}{\partial w_i} + \frac{\partial y_i}{\partial I_i}\frac{\partial I_i}{\partial w_i}\right)\right) + p\frac{\partial y_i}{\partial w_i} \\
&= a_i y_i \left(\frac{\partial y_i}{\partial w_i} + \frac{\partial y_i}{\partial I_i}\frac{\partial I_i}{\partial w_i}\right)\frac{\partial p}{\partial q}\left(\sum_{j\neq i} a_j \frac{\partial y_j}{\partial I_j}\frac{\partial I_j}{\partial p}\right) + p\frac{\partial y_i}{\partial w_i}.
\end{aligned}$$

Again the last term in the first equality becomes zero by the first order condition. The first term in the second equality is obtained by specifying all the channels through which w_i affect q. Rearranging terms gives the third equality.

CHAPTER 3

Effects of Water Balance on Prevented Planting in the US Corn Belt for Corn and Soybeans

3.1. Introduction

Springtime adverse weather conditions, such as excess moisture, can prevent crops from being planted. For example, in 2019, anomalously high precipitation in the spring across the Corn Belt led to record-high prevented-planting cropland of nearly 20 million acres according to the U.S. Department of Agriculture (USDA). As agricultural production relies heavily on climatic factors, there is a rich literature on the link between climatic conditions and agricultural outcomes. Most of the prior research focuses on impacts of growing-season weather on crop yields (e.g., Gammans et al. 2017; Lobell and Field 2007; Ortiz-Bobea et al. 2019; Schlenker and Roberts 2009). Another strand of literature has sought to understand acreage responses to past local growing-season weather (Cui 2020; Ramsey et al. 2021). Although planting-season weather can also have a large impact on crop production by reducing planted area, the effects of planting-season weather on prevented planting have been understudied.

Previous studies (e.g., Hendricks et al. 2014b; Miao et al. 2016) have documented that springtime heavy rainfall tends to reduce planted acres–for example, due to the difficulty of operating machinery in wet soils. These studies include planting-season weather variables as controls in estimating acreage responses to economic variables (e.g., crop prices). For example, Miao et al. (2016) used monthly cumulative precipitation in March-May, and Hendricks et al. (2014b) used a dummy variable equal to one if precipitation in April is above a certain threshold level (e.g., 75th percentile). To our knowledge, the study by Boyer et al. (2022) is the only study that seeks to statistically characterize the relationship between planting-season weather and prevented planting. Using monthly precipitation and temperature data, they find that monthly precipitation from January to May affects the prevented-planting ratio of corn while only precipitation in May and June impacts prevented-plant ratio of soybeans.

In this study, we use the data on water balance measured by precipitation minus reference evapotranspiration to shed light on the intra-season acreage sensitivities to planting-season weather for corn and soybeans in the US Corn Belt, which produces around 20% of the world's corn and soybeans. This measure can be viewed as non-standardized version of the Standardized Precipitation Evapotranspiration Index (SPEI) (Vicente-Serrano et al. 2010). The motivation for using water balance in the present study is the following. The ability of farmers to plant their intended crop depends not only on precipitation but also evapotranspiration – how quickly water transfers from the land to the atmosphere. Accounting for water demand for evapotranspiration is essential because farmers have to wait until fields are dry enough for the seeder to operate properly without causing soil compaction (Sacks et al. 2010).

Accounting for water demand for evapotranspiration is also important for climate change impact assessments. Considering only precipitation may overestimate the impacts of planting-season weather on prevented planting in a changing climate and may lead to misleading adaptation strategies or polices, as evaporative demand is generally projected to increase (Vicente-Serrano et al. 2020). Admittedly, modeled soil moisture data could be alternative appealing weather data for our research questions. However, we opted to use water balance because there are more extensive downscaled climate change projections on precipitation and reference evapotranspiration than on soil moisture. Lastly, water balance provides a climatologically intuitive interpretation of parameter estimates compared to Boyer et al. (2022) that separately estimated parameters on precipitation, precipitation squared, temperature, and the interaction of precipitation and temperature.

In the US Corn Belt, the most common cause (> 90%) of prevented planting has been excess moisture (USDA RMA, 2021). Growing evidence suggests that parts of the US Corn Belt are projected to experience more frequent precipitation extremes as anthropogenic global warming continues (e.g., Zhou et al., 2022). However, warming temperature in spring due to global warming can increase evapotraspiration and can offset increases in precipitation, making net effects on water balance less clear. To the extent to which changes in water balance over the planting season leave more land unplantable, efforts to meet future food demand will be undermined.

This study consists of two parts. In the first part, we ask how planting-season weather affects prevented-planting in the US Corn Belt by examining the relationship between monthly water balance and prevented planting during the planting season across counties in the US corn Belt. In the second part, using a model developed from observations and downscaled climate change projections, we ask how the risk of prevented-planting of corn and soybeans will change during the mid 21st century (2036-2065) to identify at-risk areas. Answers to these questions will help stakeholders and policy-makers to make informed decisions to reduce the economic and food security impacts of prevented planting.

3.2. Methods

3.2.1. Study Region.

We focus our analysis on a major crop-producing region made up of 12 states in the US (See figure S3.1.). The region produces most of the corn and soybeans in the US. In the region, the most active planting season is around May for both crops, although corn tends to be planted a few weeks earlier. To explore the potential for heterogeneous responses between sub-regions in our analysis, we divide the region into northern and southern sub-regions by equally spaced latitude as seen in figure S3.1.

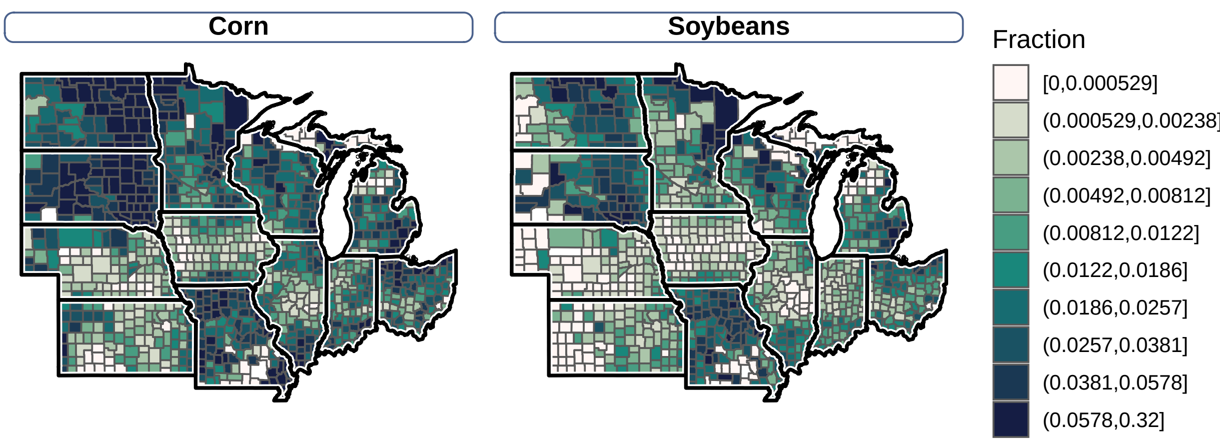


FIGURE 3.1. Average prevented-planting share in 2012–2021

3.2.2. Data.

Acreage

We source annual county-level crop acreage data from the USDA Farm Service Agency (FSA) for 2012–2021. The FSA requires producers who participate in the federal programs including crop insurance to file an annual report of all cropland use on their farms. Producers are required to report *prevented*, *planted* (and

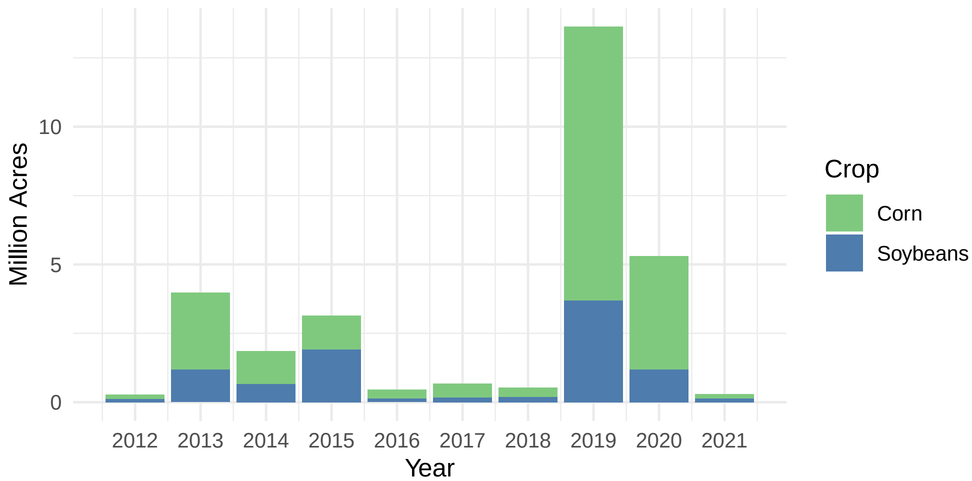


FIGURE 3.2. Time series of total prevented-planting acres in 12 states

successfully harvested), and (planted but) *failed* acres by crop. Failed acreage is acreage that was planted with the intent to harvest but failed before it could be brought to harvest because of adverse conditions. Under the US federal crop insurance program, prevented planting refers to the inability to plant the intended crop acreage with proper equipment by the final planting dates designated by the USDA Risk Management Agency (RMA). In our study region, the final planting dates are around the end of May for corn and mid-June for soybeans.

For the dependent variable, we take a ratio of prevented-planting acres to total acres—sum of prevented, planted, and failed acres—for each crop and year. We refer to this measure as the prevented-planting share. Figure 3.1 shows spatial distribution of the average prevented-planting share in 2012–2021, whereas figure S3.2 shows maps of the average planted acres in 2012–2021. Both for corn and soybeans, prevented-planting share is high in the Prairie Pothole Region, which covers parts of North and South Dakotas and Minnesota. Figure 3.2 shows time series of total prevented-planting acres for corn and soybeans in the 12 states over the period. Prevented-planting acres were highest in 2019 reaching 13.6 (9.9 for corn and 3.6 for soybeans) million acres in the region. In percentage terms, these numbers are about 12% of total corn acres (i.e., sum of planted, prevented, and failed acres) and 5.6% of total soybean acres (See figure S3.3.).

Water Balance

We obtain daily (4km-resolution) gridded data on precipitation and reference evapotranspiration from gridMET (Abatzoglou 2013). Reference evapotranspiration was calculated using the Penman-Monteith equation for a reference grass crop. We define water balance as precipitation (mm) minus reference evapotranspiration (mm). To match the acreage data, we aggregate the gridded data to the county level using the fraction of cropland as the weight in each GridMET grid. We derived cropland using land cover data from the National Land Cover Database (NLCD). For empirical analysis with monthly water balance data, we compute water balance for a month as sum of daily water balance in that month. This measure is essentially a non-standardized version of the standardized precipitation evapotranspiration index (SPEI) with one-month time window (Vicente-Serrano et al. 2010). We use raw water balance rather than SPEI because we are interested in approximating actual soil moisture level, which biophysically may be a more accurate assessment of local water balance than SPEI. To explore more detailed time-varying effects, we also conduct our analysis using daily water balance data. Table S3.1 shows descriptive statistics of the data for empirical analysis. Figure 3.3 provides a preliminary insight into how monthly water balance influences the changes in prevented-planting shares. The figure suggests that, for corn (soybeans), prevented-planting share tends to be high when there is high excess water in May (June).

Multivariate Adaptive Constructed Analogs (MACA)

To assess climate change impacts, we utilize 20 downscaled Global Climate Models (GCMs) of Coupled Model Intercomparison Project Phase 5 (CMIP5) from the Multivariate Adaptive Constructed Analogs (MACA) dataset (Abatzoglou and Brown 2012). MACA is a statistical method for downscaling GCMs from their native coarse resolution to a higher spatial resolution that reflects observed patterns of daily near-surface meteorology and simulated changes in GCMs experiment. We first obtain 4km-resolution monthly precipitation and reference evapotranspiration data for the historical period (1950-2005) and the mid of century (2036-2065) for a moderate emission scenario (RCP4.5). We calculate monthly water balance data as input into the aforementioned models that were empirically developed with observed data. Downscaled MACA data allow for interoperable use in climate impact assessments as they were created using gridMET as training data. As we did for observed weather data, we aggregate the gridded data to the county level using the fraction of cropland as the weight in each MACA grid.

3.2.3. Regression Model.

To investigate the relationship between planting-season weather and prevented-planting share, we use

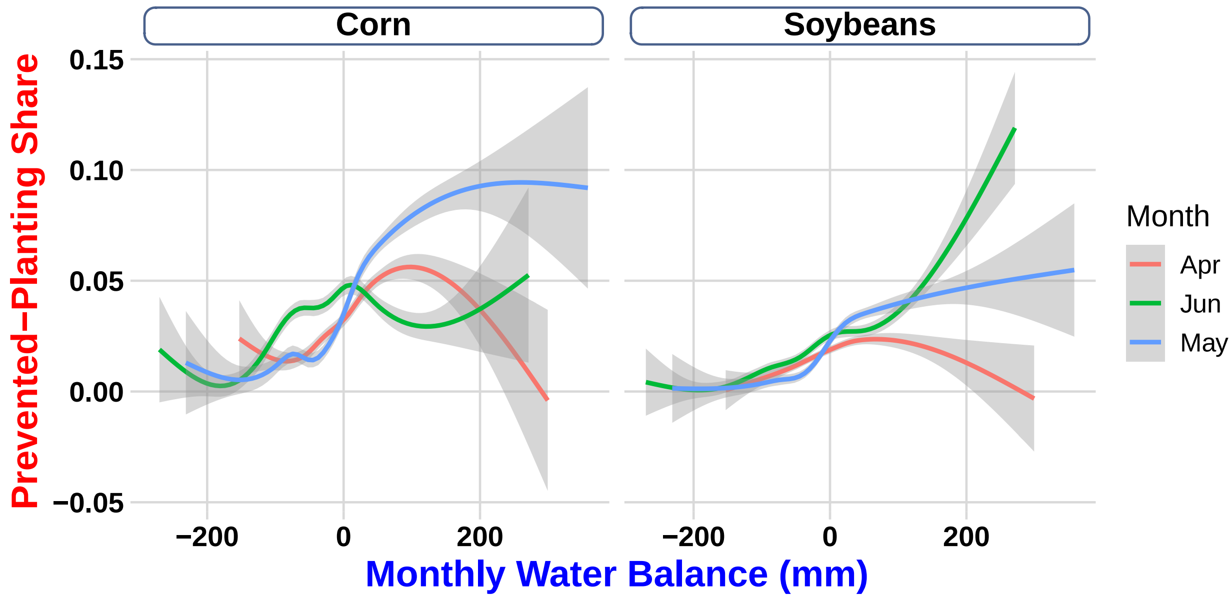


FIGURE 3.3. Prevented-Planting Share against Monthly Water Balance in 2012–2021

Note: The figure uses all observations in the data from 2012–2021 and then fits a smoothing function on monthly water balance data. Grey areas show 95% confidence intervals.

the fixed-effects Poisson pseudo-maximum likelihood (PPML) estimator (Hausman et al. 1984; Wooldridge 1999). We use the estimator because more than 30% of the observations of the dependent variable of our data are zero, and naturally accommodating zero values without any transformation, the Poisson model produces consistent estimates that are interptretable as semi-elasticities. Even if the data generating process does not follow a Poisson distribution, the estimator is still consistent as long as the mean is correctly specified. Admittedly, this robustness does not extend to estimated covariance matrices, but bootstrap could be used to resolve this concern. In addition, the PPML does not suffer from the incidental parameter problem in the presence of fixed effects, unlike other maximum likelihood estimators. The model can be also applied to a fractional or continuous variable (Gaule and Piacentini 2013; Silva and Tenreyro 2006; Zhao et al. 2013).

Parametric estimation using monthly water balance data

Our initial step is to estimate the following simple model:

$$a_{it} = exp(\sum_{m \in M} \beta_m w_{mit} + \lambda_i + f(t))\epsilon_{it}, \tag{3.1}$$

where a_{it} is prevented-planting share in county i in year t. w_{mit} is monthly water balance in month m in year t, and $M \equiv \{March, April, May, June\}$. λ_i is county fixed effects, $f(t)$ denotes a linear time trend, and ϵ_{it} is the error term. We run Eq.(3.1) separately for each crop and region. All of our regressions use crop-by-county total acres as the weight. For standard errors, we use year-block bootstrapped standard errors whereby we randomly resample our data by year 1,000 times. Resampling data by year assumes independence between years but accounts for any cross-sectional correlation.

Nonparametric estimation using binned monthly water balance data

So far, we assumed the response of prevented-planting share to water balance to be constant across the level of water balance, conditional on timing of water balance. However, the response could be nonlinear. For example, extremely wet conditions could have a more detrimental effect than moderately wet conditions. We explore potential nonlinear responses nonparametrically using a binned model. To do so, we bin monthly water balance data into ten indicator variables using deciles of each county-specific historical (1979–2021) distribution of monthly water balance, and estimate the following model:

$$a_{it} = exp(\sum_{m\in M, d\in\{1,2,3,4,6,7,8,9,10\}} \beta_{m,d}\mathbb{I}\{w_{mit} \in Q_{mi,d}\} + \lambda_i + f(t))\epsilon_{it}, \tag{3.2}$$

where $\mathbb{I}$ is an indicator function that equals 1 if $w_{mit} \in Q_{mi,d}$ and 0 otherwise. $Q_{mi,d}$ denotes the interval of the dth decile of the historical distribution of water balance for month m in county i. To avoid perfect multicollinearity, we drop the bin for the 5th decile in each month.

Semiparametric estimation using daily water balance data

To investigate more detailed time-varying effects, we perform our analysis using daily water balance. We estimate a regression of the form:

$$a_{it} = exp(g(w_{1it}, ..., w_{Dit}) + \lambda_i + f(t))\epsilon_{it}, \tag{3.3}$$

where D is the number of days between March 1 and July 30. We include July simply because estimated parameters are sometimes sensitively influenced by boundary values in semiparametric estimation. $g()$ is a natural cubic spline. Unlike the analysis using monthly water balance variables, Eq.(3.3) restricts coefficients to vary smoothly over the planting season. (See Supplementary Materials for more details.)

We optimally choose the degrees of freedom ($df \in \{3,4,5,6,7,8\}$) for $g()$ for each crop and region using the root mean square error (RMSE) of year-block 10-fold cross validation.

3.3. Results

3.3.1. Regression Results.

Figure 3.4 shows the response of prevented-planting share to monthly water balance with 95%

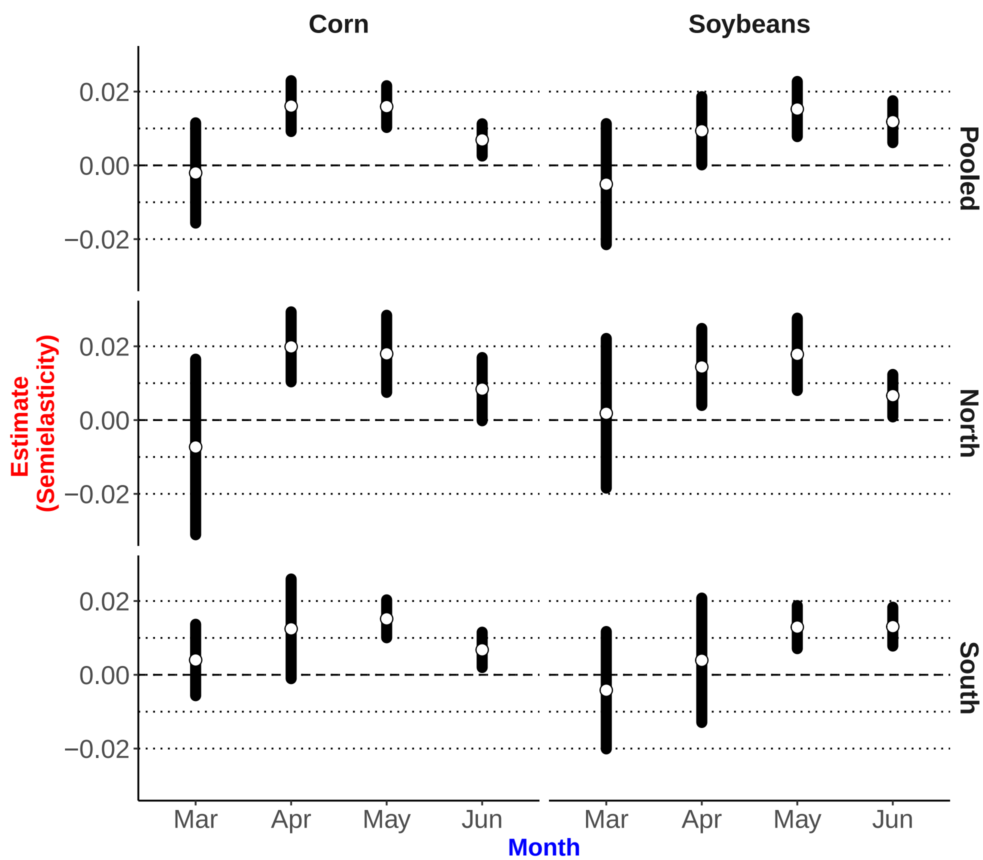


FIGURE 3.4. Results from the Parametric Estimation Using Monthly Water Balance Data.

Note: The figure visualizes the regression results from the parametric estimation in Eq.(3.1) by crop and region. White dots represent point estimates from regressions and vertical black lines represent 95% confidence intervals around them. Point estimates can be interpreted as semielasticities. For standard errors, we use year-block bootstrapped standard errors whereby we randomly resample our data by year for 1000 times. The pooled region include all counties in 12 states in the study region. The northern region includes counties located above northern Iowa. The southern region includes counties located below northern Iowa and above southern Missouri.

confidence intervals. Point estimates can be interpreted as semi-elasticities. The figure suggests that water balance in April-May (May-June) has large and statistically significant positive effects on prevented-planting share of corn (soybeans) while water balance in March has limited effects. We first interpret our results using the pooled region. Specifically, we find that a one standard deviation increase in water balance in April (53mm), May (64mm), and June (66mm) leads to increases in prevented-planting share of corn by 139%, 177%, and 58%, respectively. We calculated these numbers using the following formula: $[e^{\hat{\beta}_m \sigma_m} - 1] \times 100\%$, where σ_m is standard deviation of water balance for month m in 1979–2021. Again using the pooled region, the effects of monthly water balance on prevented-planting share for soybeans tends to be less pronounced than those for corn. However, there is a noticeable difference in the timing of critical months with a more significant influence in May and June. For soybeans, a one standard deviation increase in water balance in April, May, and June would lead to increases in prevented-planting share by 65%, 165%, and 120%, respectively. Our results also suggest that, when we use monthly water balance rather than precipitation, goodness of fit—measured by McFadden's pseudo R^2—increases by 5-7% (McFadden et al., 1973). Similarly, goodness of fit also increases by 5% when we use monthly water balance rather than total water balance from March to June. The results from the pooled regions obscure heterogeneous responses between regions. In general, the northern region shows a stronger positive relationship between water balance and prevented-planting share in April and May for both crops than its counterpart. On the other hand, in the southern region, the effect of water balance remains relatively salient even in June and this tendency is particularly true for soybeans.

Figure 3.5 shows the regression results from the nonparametric estimation in Eq.(3.2). It visualizes the percentage impacts of an event of monthly water balance being in each decile on prevented-plant share relative to that of monthly water balance being in a 5th decile. The numbers from 1 to 10 in each month represents bins of 1st to 10th decile of the historical distribution of monthly water balance. (In figure S3.5, we show the coefficients and confidence intervals used to calculate the percentage changes.) The figure indicates that an event of extremely wet conditions in the southern region can be highly damaging for both crops. To be specific, an event of monthly water balance in 10th decile in May increases prevented-planting share by 15 times for corn and by 11 times for soybeans relative to an event of 5th decile. In the northern region, extremely wet conditions appear to be less damaging than those in the southern region. However, for soybeans in the northern region, extremely wet condition in May (i.e., 10th decile) could increase prevented-planting share more than by six-fold relative to an event of 5th decile in May.

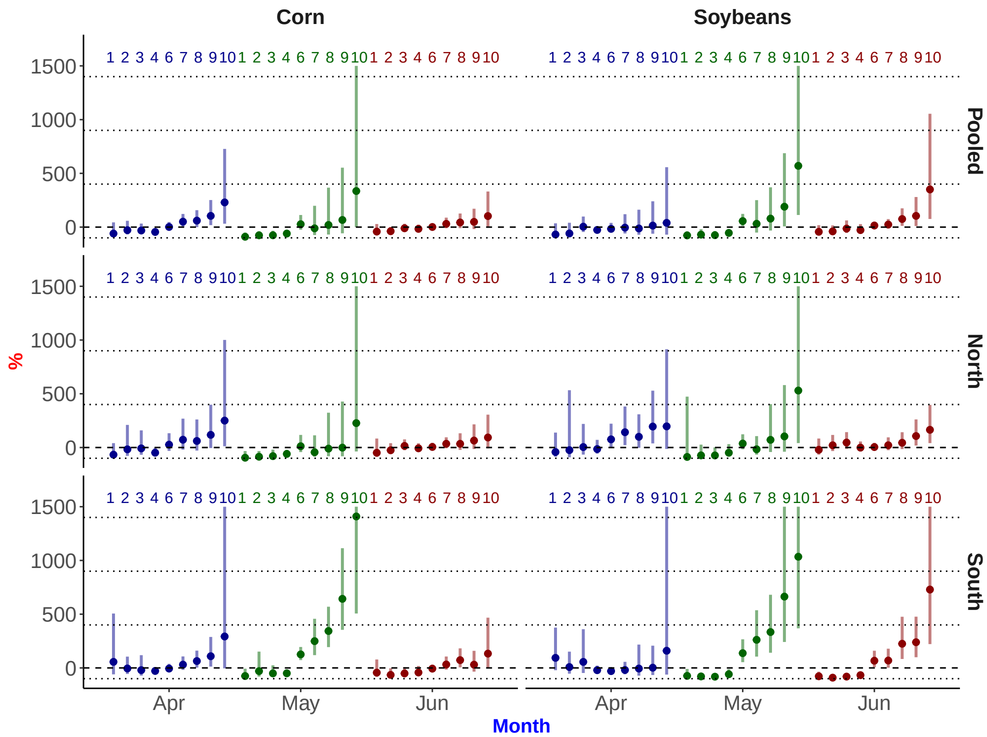


FIGURE 3.5. Percentage Impacts of an Event of Monthly Water Balance Being in Each Decile Bin

Note: The figure visualizes the regression results from the nonparametric estimation in Eq.(3.2) by crop and region, in which we binned monthly water balance data into deciles of county-specific distribution of monthly water balance for each county. The variable for 5th decile was dropped to avoid perfect multicollinearity. Each dot represents a percentage impacts of monthly water balance being in each decile relative to 5th decile. For standard errors, we use year-block bootstrapped standard errors whereby we randomly resample our data by year for 1000 times. The pooled region include all counties in 12 states in the study region. The northern region includes counties located above northern Iowa. The southern region includes counties located below northern Iowa and above southern Missouri.

Figure 3.6 shows time-varying effects of daily water balance on prevented-planting share. The results for the pooled region suggest that corn is most sensitive to water balance around early May whereas soybeans is most sensitive to water balance around late May. To be specific, one standard deviation increase in water balance on the estimated most critical date for corn (soybeans), May 6 (May 23), increases prevented-planting share by 8.7% (9%). (See table S3.3 for more detailed calculation and results.) Based on the results for subregions, we also find that, while early May is the most critical time window for corn and soybeans

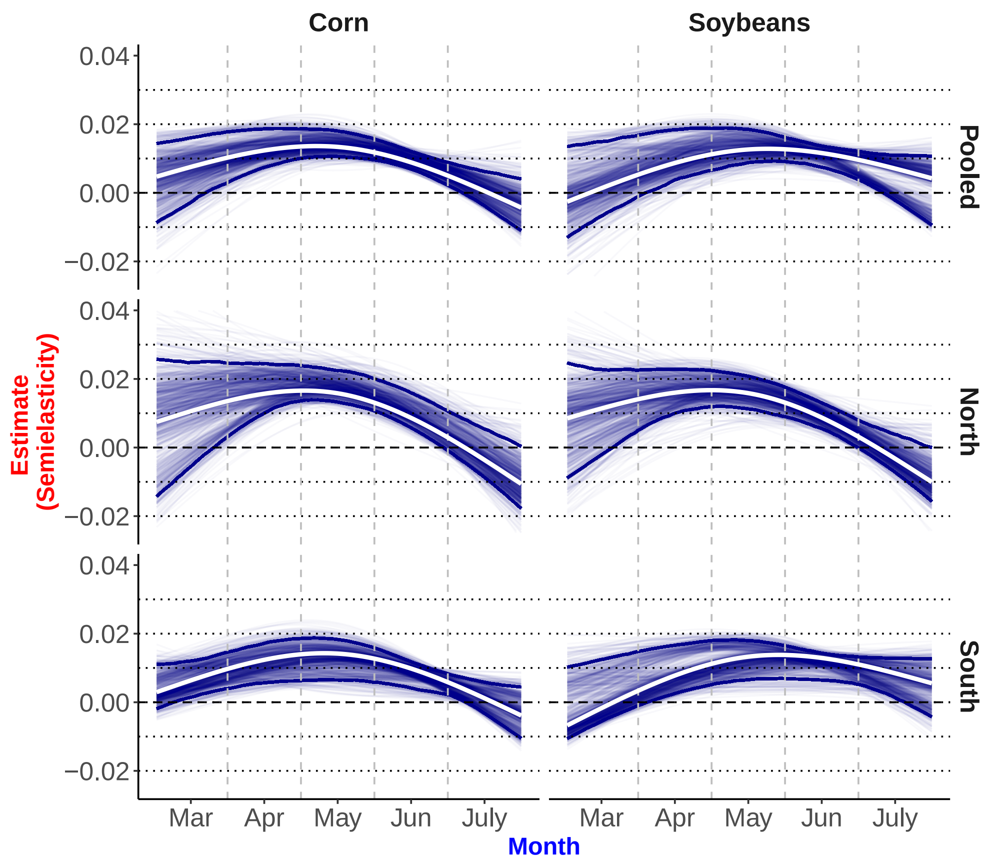


FIGURE 3.6. Results from the Semiparametric Estimation using Daily Water Balance Data

Note: The figure visualizes the regression results from the semiparametric estimation in Eq.(3.4) by crop and region. Each line represents a response function of prevented-planting share to daily water balance. White lines correspond to the response functions obtained from the full-sample (2012–2021) model. The overlay of blue response functions corresponds to 1000 response functions derived from bootstrapped regressions in which years of data (2012–2021) were sampled with replacement. 95% confidence bands derived from the bootstrapped regressions are represented in thicker blue lines. The pooled region include all counties in 12 states in the study region. The northern region includes counties located above northern Iowa. The southern region includes counties located below northern Iowa and above southern Missouri.

in the northern region and for corn in the southern region, late May is the most critical time window for soybeans in the southern region.

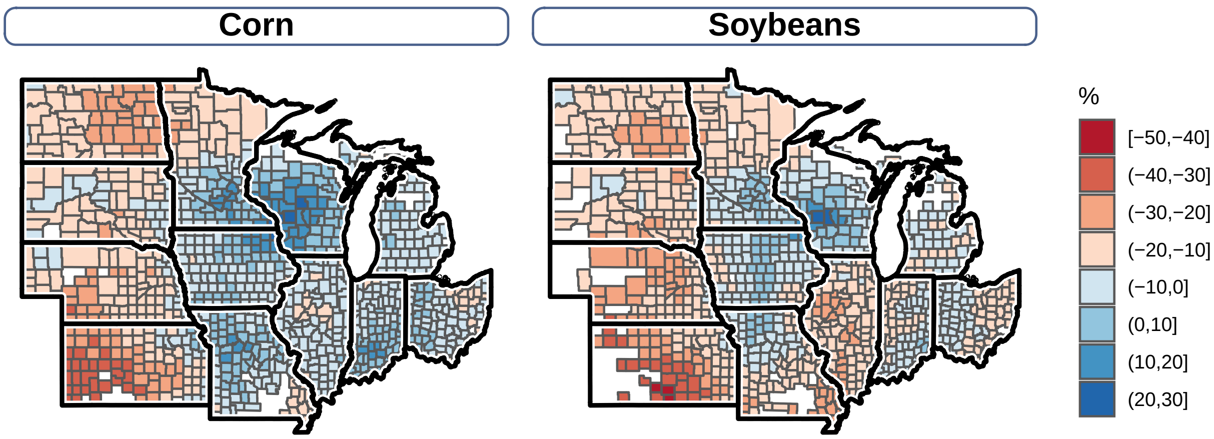
3.3.2. Climate Change Impact Assessment.

Figure 3.7b shows the spatial distribution of percentage impacts of climate change on prevented-planting share under the moderate emission scenario (RCP4.5) during the mid 21st century (2036-2065) compared to the historical (1950-2005) levels. Each value in the map represents the county-specific median of the projected percentage impacts in prevented-planting share from 20 downscaled climate models. Our results suggest that prevented-planting share of corn is projected to increase in parts of Iowa, Minnesota, and Wisconsin by 0-30%. Located in the northern part of the US Corn Belt, many of these counties have experienced acreage expansion over the past few decades (Cui 2020). Our results suggest that wetter spring conditions in these areas would put a strain on farmers' ability to plant crops in a changing climate, even if growing-season weather becomes more favorable for planted crops. Conversely, modeled results suggest that prevented-planting share would decrease by 10-40% in North and South Dakotas, in which prevented planting has been historically a major concern. This finding is of great importance and relevance in a changing climate given that the two states—particularly their eastern regions—have seen the most extensive growth of corn and soybean acreage in the past decades due to a warming trend (Cui 2020).

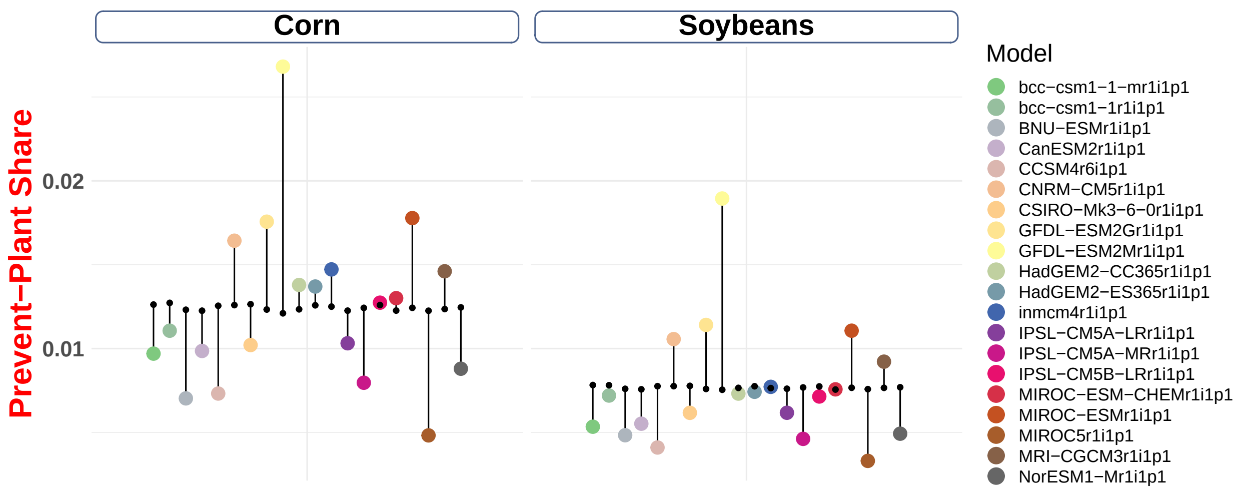
Given that much of the northern region is expected to see an increase in precipitation during the spring, the spatial heterogeneity of the projected changes in the prevented-planting share is due to net water supply (i.e., water balance). Figure S3.4 shows the county-specific median of the projected changes in water balance under the moderate emission scenario in the mid-century. As the figure shows, many parts of North and South Dakotas are expected to see a decrease in water balance in April and May. The opposite is true for the region expected to see an increase in prevented-planting shares. Because reference evapotranspiration is projected to increase in a warmer climate, we posit that failure to account for water demand in the planting period would suggest widespread increases in prevented planting.

It is worth emphasizing that farmers' adaptive responses to projected changes could mitigate acreage losses. For example, warmer temperature in late winter or early spring in the future might allow farmers to start planting a few weeks earlier, which could potentially provide farmers with more flexibility to shift planting dates. Besides planting date adjustments, farmers can implement variety of adaptation strategies to reduce acreage losses: planting cover crops to improve water infiltration of the soil, installing tile drains, and investing in machinery that could reduce the total hours of planting time.

It should be also noted that we assessed climate change impacts using county-specific median projections but there are substantial differences in projected impacts across climate models. Figure 3.7b shows aggregate impacts for the study region across 20 downscaled climate models under the moderate (RCP4.5) during the mid-century (2036-2065). (Figure S3.6 shows projected prevented-planting shares



(a) Climate Change Impacts (%) on Prevent-Plant Share for the Did-Century (2036-2065) under RCP4.5.



(b) Climate Change Impact Projections on Prevent-Plant Share

Note: (a) The values in the map represent the county-specific median of the projected climate change impacts (%) on prevented-planting share from 20 downscaled climate models (MACA) relative to the historical period (1950-2005). The climate change impact estimates uses the response function for the pooled region from the parametric estimation using monthly water balance data. (b) Each colored dot represents an acreage-weighted predicted prevented-planting share from each 20 downscaled climate model under the mid-century (2036–2065) and under the moderate emission (RCP4.5) scenario. Each black dot represents an acreage-weighted predicted prevented-planting share for the historical period (1950-2005) from the MACA.

FIGURE 3.7. Projected Climate Change Impacts on Prevented Planting

across 20 downscaled climate models under the moderate (RCP4.5) and high (RCP8.5) emission scenarios during the mid-century (2036-2065) and the late-century (2070-2099).) The figure suggests that there is substantial uncertainty about how future changes in water balance during the springtime period would affect prevented-planting share.

3.4. Conclusion

In this paper, we sought to establish the relationship between planting-season water balance and prevented-planting of corn and soybeans in a wide expanse of the US Corn Belt. Using water balance that accounts for both water supply and demand, we find that wet conditions in April-May (May-June) tends to increase prevented-planting of corn (soybeans). Using nonparametric estimation, we also document that extremely wet events in the critical planting window can detrimentally increase prevented-planting share, particularly in the southern portion of the region.

We have a caveat. Our study region has a unique institutional context not to mention biophysical environment. Thus, our results might not hold in other contexts or regions. A large portion of farmers participate in the federal crop insurance program, which has its unique prevented-planting provisions. This means that our results reflect farmers' behavioral responses driven by incentives they face under the current prevented-planting provisions in the federal crop insurance policies, although, of course, weather is the key driver of prevented-planting acreage (Kim and Kim 2018; Rejesus et al. 2005; Wu et al. 2020).

Unlike Boyer et al. (2022), we do not find that planting-season weather in March plays an important role in determining the prevented-planting share of corn. One possible explanation is that their standard errors do not account for the correlation across cross-sectional units (here, counties) in the data even though the amount or fraction of prevented-planting area is highly spatially correlated. Their assumption of spatially uncorrelated errors could have led Boyer et al. (2022) to over-rejecting the null hypothesis of no effect in March. In addition, our result indicates that monthly water balance in June has a statistically significant effect on prevented-planting share of soybeans, although Boyer et al. (2022) found mixed results on precipitation and temperature in June. We posit that our different finding stems from the fact that we harness a weather variable that simultaneously accounts for water supply and demand.

Our climate change impacts assessments suggest that prevented-planting share will decrease in the Prairie Pothole Region—where prevented-planting acres tended to be high—in the future especially for corn under the moderate emission scenario (RCP4.5) during the mid of the century (2035-2065). Such decline is due to a warming trend that increases water demand captured by reference evapotranspiration. In some northern parts of the US Corn Belt, it appears that increasing water demand is not enough to offset increasing precipitation in the planting season. As a result, prevented-planting share of corn is projected to increase by 0-30% in the region.

It should be noted that these assessments were based on the median projected impacts among 20 climate models and there is a substantial difference among the projected impacts from different climate models. In addition, our projections are based on the average planting season over multiple years meaning that our results are silent about the effect of changing inter-annual variability of planting season weather. One future research direction would be investigation into changes in the intensity and frequency of extreme planting-season weather and their corresponding consequences on prevented-planting cropland.

3.5. Supplementary Section

3.5.1. Semiparametric Estimation.

We express our regression model using daily water balance as:

$$a_{it} = exp(g(w_{1it}, ..., w_{Dit}) + \lambda_i + f(t))\epsilon_{it}, \tag{3.4}$$

where a_{it} is prevented-planting share in county i in year t. $d \in \{1, ..., D\}$, where D is the number of days between March 1 to July 30. $g()$ is a natural cubic splines. We optimally choose the degrees of freedom for $g()$ among 3,4,5,6,7, and 8 for each crop and region using the Root Mean Square Error (RMSE) of year-block 10-fold cross validation. Unlike in the analysis using monthly water balance variables, $g()$ restricts coefficients to vary smoothly over the planting season.

The natural cubic spline function $g()$ in Eq.(3.4) involves a basis matrix, which allows us to reduce the number of parameters to be estimated (For detailed discussion, see Ortiz-Bobea (2021)). In our setting, the basis matrix maps daily water balance with D bins to K bins ($K < D$). Let W be a ($nT \times D$) matrix of daily water balance over the planting season. The basis matrix for a natural cubic spline with K degrees of freedom, B, allows us to reduce dimensionality as follow:

$$\underset{nT\times K}{\widetilde{W}} = \underset{nT\times D}{W} \times \underset{D\times K}{B}.$$

The effect of water balance over the planting season could be then modeled as:

$$g(w_{0it},...,w_{Dit}) \tag{3.5}$$

$$= \sum_{k=1}^{K}\sum_{d=1}^{D}\gamma_k B_{dk} w_{dit} \tag{3.6}$$

$$= \sum_{k=1}^{K}\gamma_k \underbrace{\sum_{d=1}^{D} B_{dk} w_{dit}}_{\widetilde{w}_{kit}} \tag{3.7}$$

Using the newly constructed regressors $\widetilde{W}$, we can approximate the time-varying effect of daily water balance on prevented-planting share over the entire planting share with K parameters rather than D parameters. The prevented-planting model that allows for time-varying coefficients on water balance can be written as:

$$a_{it} = exp(\sum_{k=1}^{K}\gamma_k \widetilde{w}_{kit} + \alpha_i + f_i(t))\epsilon_{it}, \tag{3.8}$$

We recover the marginal effect of water balance at each d by pre-multiplying $\hat{\Gamma}$ by the basis matrix:

$$\underset{D\times 1}{\hat{\beta}} = \underset{D\times K}{B} \times \underset{K\times 1}{\hat{\Gamma}}$$

3.5.2. Supplementary Figures.

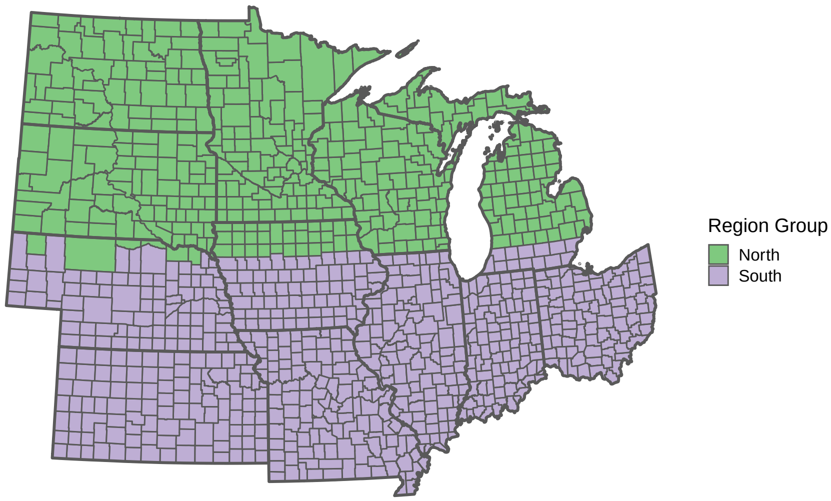


FIGURE S3.1. Study Region and Regression Group

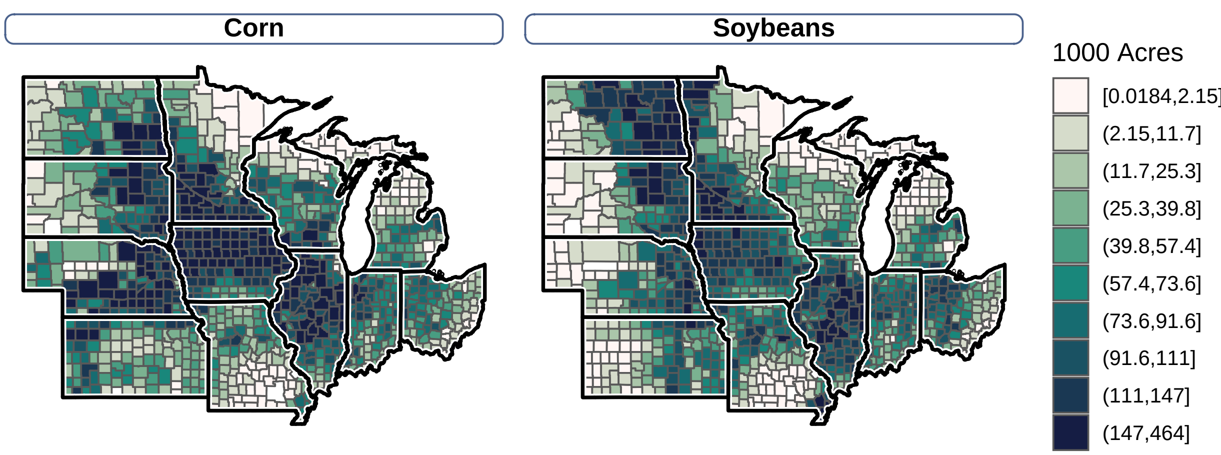


FIGURE S3.2. Average Planted Acres in 2012–2021

Note: The values in the map include *planted* (and successfully harvested) and (planted but) *failed* acres from the the USDA FSA.in 2012–2021

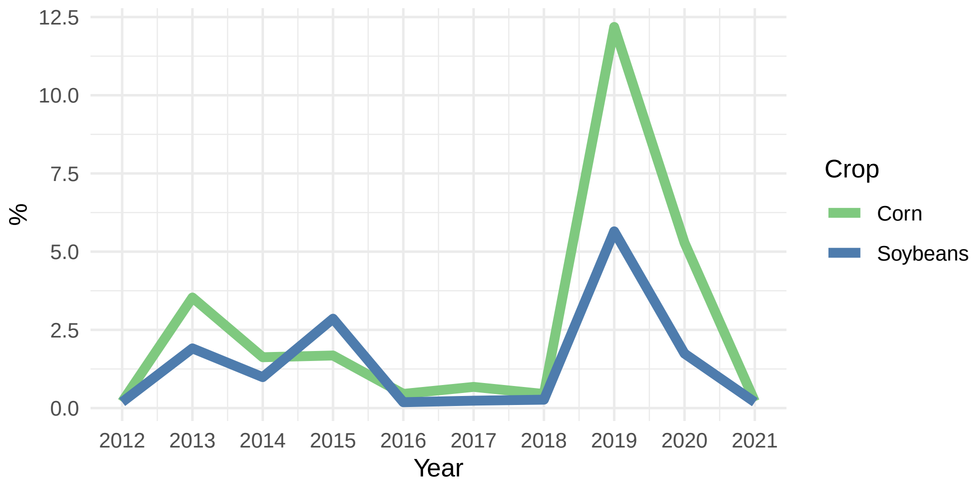


FIGURE S3.3. Prevent-Plant Acres as a Percent of Total Acres in 12 States

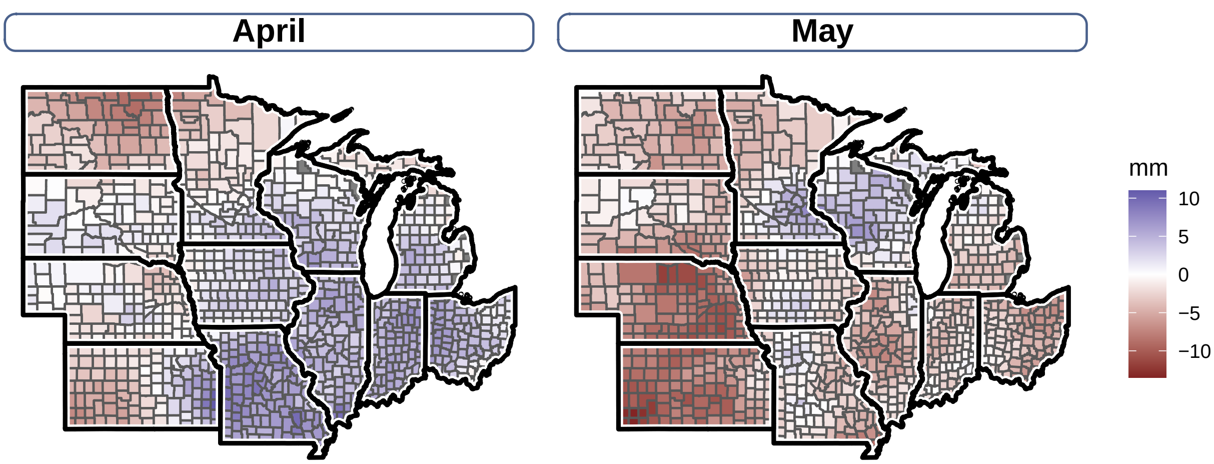


FIGURE S3.4. Projected Change in Water Balance under RCP4.5 in the Mid-Century

Note: Each value represents the county-specific median of the projected changes in monthly water balance (mm) from 20 downscaled climate models (MACA) for the mid-century (2036–2065) under the moderate emission scenario (RCP4.5) relative to the historical (1950-2005) water balance.

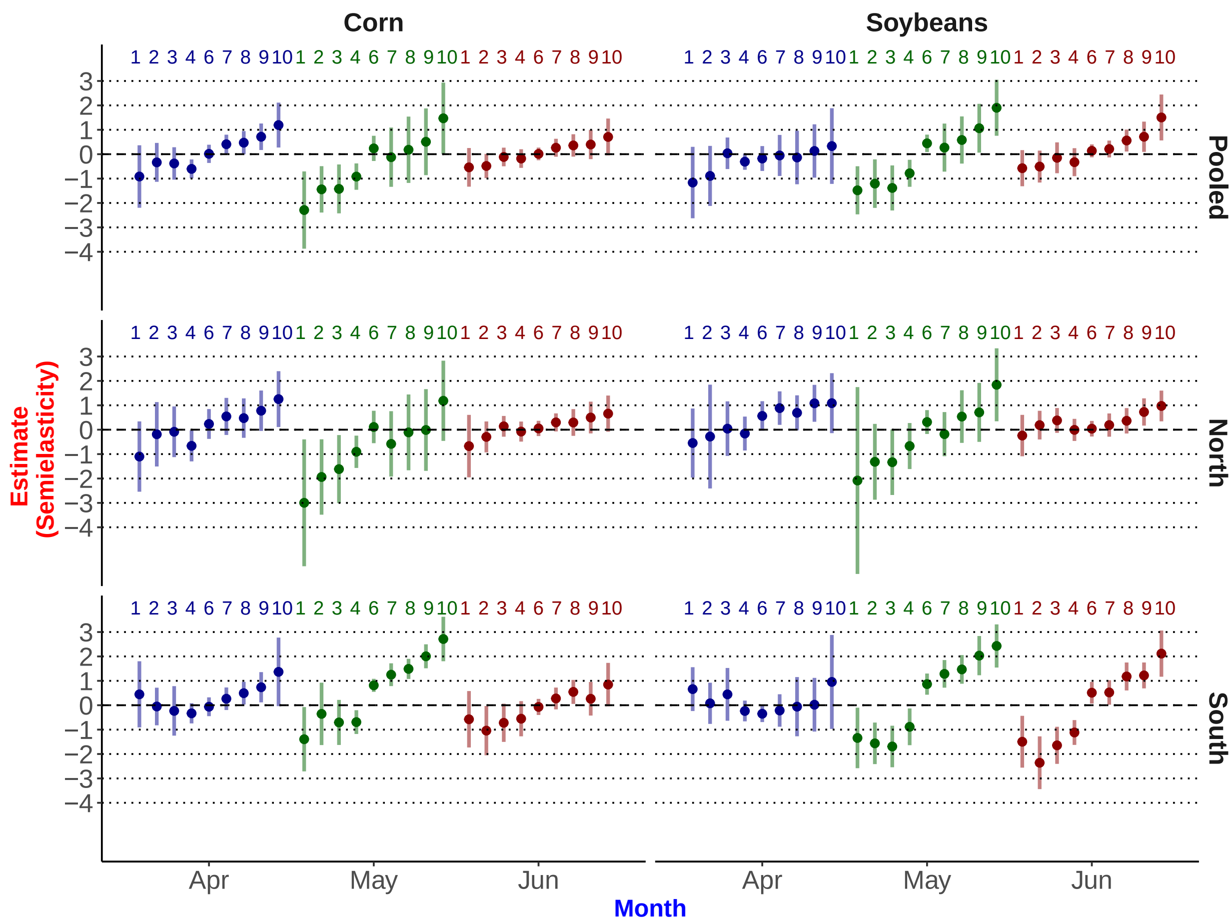


FIGURE S3.5. Results from the Nonparametric Estimation using Binned Monthly Water Balance Data

Note: The figure visualizes the regression results from the nonparametric estimation in Eq.(3.2) by crop and region, in which we binned monthly water balance data into deciles of county-specific distribution of monthly water balance for each county. The variable for 5th decile was dropped to avoid perfect multicollinearity. Each dot represents a marginal effect of monthly water balance being in each decile relative to 5th decile. Point estimates can be interpreted as semielasticities. For standard errors, we use year-block bootstrapped standard errors whereby we randomly resample our data by year for 1000 times. The pooled region include all counties in 12 states in the study region. The northern region includes counties located above northern Iowa. The southern region includes counties located below northern Iowa and above southern Missouri.

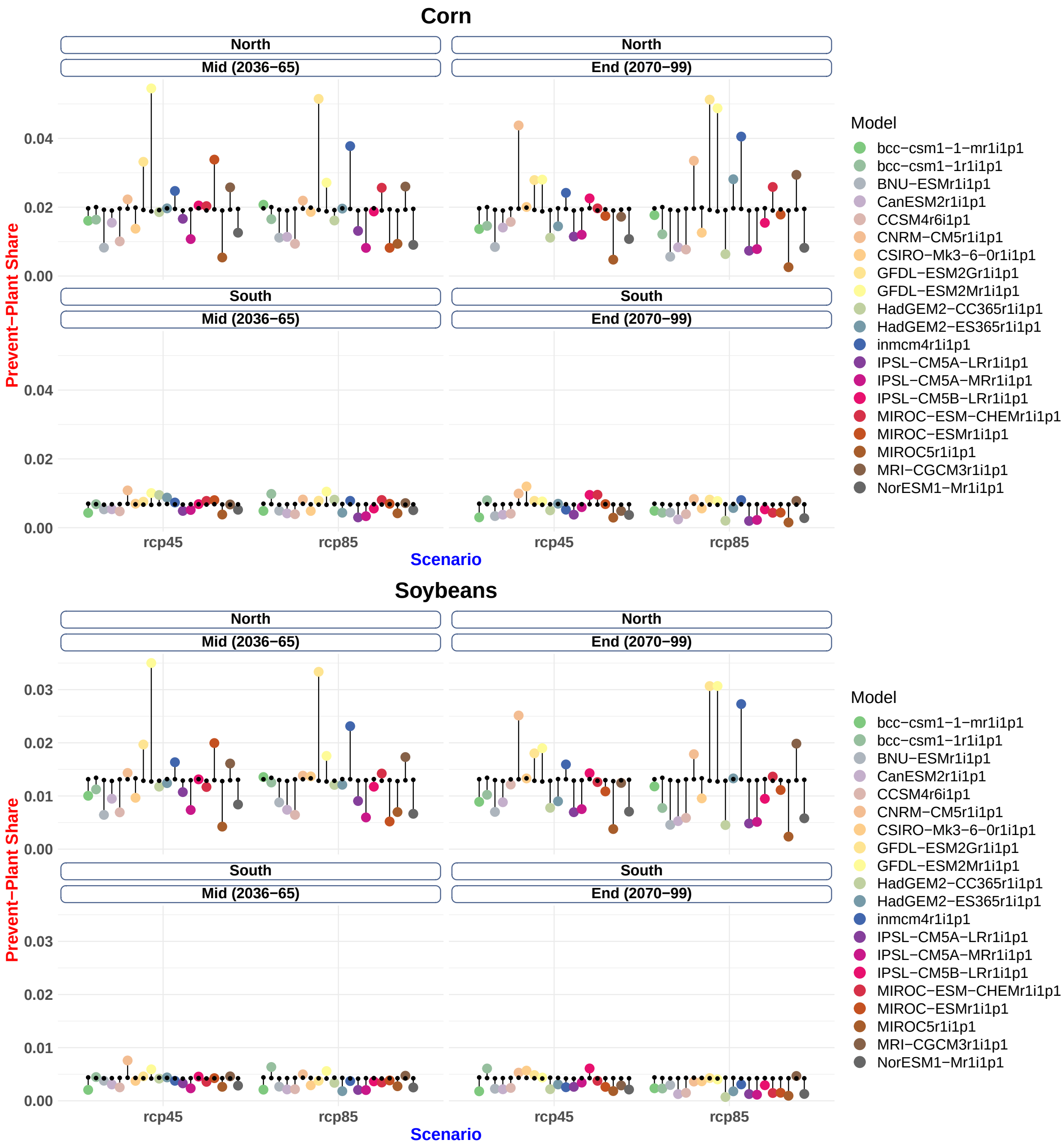


FIGURE S3.6. Climate Change Impact Projections on Prevent-Plant Share

Note: Each dot is acreage-weighted predicted prevented-planting share using the region-by-crop specific coefficients from the parametric estimation in Eq.(3.1) and 20 downscaled climate projections from the MACA. Each colored dot represents a predicted value for the mid-century (2036–2065) or the late-century (2070–2099). Each black dot represents a predicted value for the historical period (1950-2005).

3.5.3. Supplementary Tables.

TABLE S3.1. Descriptive Statistics

variable	group	mean	sd	mean (w/o 2019)	sd (w/o 2019)
Water Balance in Mar (mm)	Pooled	-8.833	42.253	-9.507	42.293
Water Balance in Mar (mm)	North	-12.534	29.306	-12.878	29.434
Water Balance in Mar (mm)	South	-6.715	47.993	-7.578	48.024
Water Balance in Apr (mm)	Pooled	-22.050	53.394	-22.725	53.277
Water Balance in Apr (mm)	North	-26.963	44.886	-27.838	44.735
Water Balance in Apr (mm)	South	-19.238	57.513	-19.799	57.394
Water Balance in May (mm)	Pooled	-32.099	63.848	-34.140	62.259
Water Balance in May (mm)	North	-46.696	51.983	-48.279	51.014
Water Balance in May (mm)	South	-23.745	68.348	-26.049	66.524
Water Balance in Jun (mm)	Pooled	-52.184	66.287	-52.941	66.375
Water Balance in Jun (mm)	North	-50.511	58.408	-50.738	58.778
Water Balance in Jun (mm)	South	-53.141	70.383	-54.202	70.325
Water Balance in Jul (mm)	Pooled	-74.155	68.571	-74.519	68.917
Water Balance in Jul (mm)	North	-76.087	57.966	-76.953	58.030
Water Balance in Jul (mm)	South	-73.049	73.937	-73.127	74.400
prevented-planting Share (Corn)	Pooled	0.030	0.085	0.020	0.065
prevented-planting Share (Corn)	North	0.044	0.110	0.032	0.093
prevented-planting Share (Corn)	South	0.023	0.066	0.013	0.040
prevented-planting Share (Soybeans)	Pooled	0.016	0.057	0.011	0.048
prevented-planting Share (Soybeans)	North	0.022	0.072	0.016	0.063
prevented-planting Share (Soybeans)	South	0.013	0.046	0.008	0.037

Notes: Water balance data span from 1979 to 2021 while data on prevented-planting share cover 2012–2021.

TABLE S3.2. Impacts of One Standard Deviation Increase in Monthly Water Balance

	Region	Crop	Month	Marginal Effect	SD of WB	Effect of One SD Increase (%)
1	Pooled	Corn	Mar	-0.002	42.253	-8.322
2	Pooled	Corn	Apr	0.016	53.394	135.813
3	Pooled	Corn	May	0.016	63.848	176.572
4	Pooled	Corn	Jun	0.007	66.287	58.027
5	Pooled	Soybeans	Mar	-0.005	42.253	-19.359
6	Pooled	Soybeans	Apr	0.009	53.394	64.970
7	Pooled	Soybeans	May	0.015	63.848	165.132
8	Pooled	Soybeans	Jun	0.012	66.287	119.218
9	North	Corn	Mar	-0.007	29.306	-19.229
10	North	Corn	Apr	0.020	44.886	143.761
11	North	Corn	May	0.018	51.983	154.431
12	North	Corn	Jun	0.008	58.408	63.189
13	North	Soybeans	Mar	0.002	29.306	5.541
14	North	Soybeans	Apr	0.014	44.886	90.865
15	North	Soybeans	May	0.018	51.983	152.667
16	North	Soybeans	Jun	0.007	58.408	47.199
17	South	Corn	Mar	0.004	47.993	21.244
18	South	Corn	Apr	0.012	57.513	104.725
19	South	Corn	May	0.015	68.348	182.146
20	South	Corn	Jun	0.007	70.383	60.988
21	South	Soybeans	Mar	-0.004	47.993	-18.162
22	South	Soybeans	Apr	0.004	57.513	25.408
23	South	Soybeans	May	0.013	68.348	141.950
24	South	Soybeans	Jun	0.013	70.383	150.697

Notes: Marginal Effect is derived from the semiparametric estimation Eq.(3.1) and can be interpreted as semi-elasticity. *SD of WB* refers to the standard deviation of monthly water balance in 1979–2021. *Effect of One SD Increase (%)* was calculated by $(exp(\hat{\beta}_m \sigma_m) - 1) \times 100$, where β_m is the estimated marginal effect of monthly water balance on prevented-planting share for month m and σ_m is the standard deviation of water balance in month m.

TABLE S3.3. Impacts of One Standard Deviation Increase in Daily Water Balance

	Region	Crop	Apprx. Most Critical Day	Marginal Effect	SD of WB (mm)	Effect of One SD Increase (%)
1	Pooled	Corn	05-06	0.0136	6.1070	8.6817
2	Pooled	Soybeans	05-23	0.0128	6.6957	8.9754
3	North	Corn	05-01	0.0166	5.7677	10.0441
4	North	Soybeans	04-30	0.0167	6.1425	10.8195
5	South	Corn	05-09	0.0143	8.1217	12.3099
6	South	Soybeans	05-27	0.0138	9.7320	14.3841

Notes: Marginal Effect is derived from the semiparametric estimation Eq.(3.3) and can be interpreted as semi-elasticity. *Apprx. Most Critical Date* is the date that yields the highest marginal effect from March 1 to July 30. *SD of WB* refers to the standard deviation of water balance on the most critical date. *Effect of One SD Increase (%)* was calculated by $(exp(\hat{\beta}_d \sigma_d) - 1) \times 100$, where β_d is the estimated marginal effect of monthly water balance on prevented-planting share on date d and σ_d is the standard deviation of water balance on date d.

Bibliography

ABATZOGLOU, J. T. (2013): "Development of gridded surface meteorological data for ecological applications and modelling," *International Journal of Climatology*, 33, 121–131.

ABATZOGLOU, J. T. AND T. J. BROWN (2012): "A comparison of statistical downscaling methods suited for wildfire applications," *International Journal of Climatology*, 32, 772–780.

ADJEMIAN, M. K. AND A. SMITH (2012): "Using USDA forecasts to estimate the price flexibility of demand for agricultural commodities," *American Journal of Agricultural Economics*, 94, 978–995.

AMIR, E. AND Y. GANZACH (1998): "Overreaction and underreaction in analysts' forecasts," *Journal of Economic Behavior & Organization*, 37, 333–347.

BELTRÁN, A., D. MADDISON, AND R. J. ELLIOTT (2018): "Is flood risk capitalised into property values?" *Ecological Economics*, 146, 668–685.

BERRY, S. T., M. J. ROBERTS, AND W. SCHLENKER (2014): "Corn production shocks in 2012 and beyond: Implications for harvest volatility," in *The Economics of Food Price Volatility*, University of Chicago Press, 59–81.

BOYER, C. N., E. PARK, AND S. D. YUN (2022): "Corn and soybean prevented planting acres response to weather," *Applied Economic Perspectives and Policy*.

BURKE, M. AND K. EMERICK (2016): "Adaptation to climate change: Evidence from US agriculture," *American Economic Journal: Economic Policy*, 8, 106–40.

CAMERON, A. C., J. B. GELBACH, AND D. L. MILLER (2008): "Bootstrap-based improvements for inference with clustered errors," *The Review of Economics and Statistics*, 90, 414–427.

CARTER, C. A., G. C. RAUSSER, AND A. SMITH (2017): "Commodity storage and the market effects of biofuel policies," *American Journal of Agricultural Economics*, 99, 1027–1055.

CHE, Y., H. FENG, AND D. A. HENNESSY (2020): "Recency effects and participation at the extensive and intensive margins in the US Federal Crop Insurance Program," *The Geneva Papers on Risk and Insurance-Issues and Practice*, 45, 52–85.

COLLINS, M., R. KNUTTI, J. ARBLASTER, J.-L. DUFRESNE, T. FICHEFET, P. FRIEDLINGSTEIN, X. GAO, W. J. GUTOWSKI, T. JOHNS, G. KRINNER, ET AL. (2013): "Long-term climate change: projections, commitments and irreversibility," in *Climate change 2013-The physical science basis: Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change*, Cambridge University Press, 1029–1136.

CUI, X. (2020): "Climate change and adaptation in agriculture: Evidence from US cropping patterns," *Journal of Environmental Economics and Management*, 101, 102306.

DALY, C., G. TAYLOR, AND W. GIBSON (1997): "The PRISM approach to mapping precipitation and temperature," in *Proc., 10th AMS Conf. on Applied Climatology*, Citeseer, 20–23.

DEMNITZ, R. AND S. JOSLYN (2020): "The effects of recency and numerical uncertainty estimates on overcautiousness," *Weather, Climate, and Society*, 12, 309–322.

DESCHÊNES, O. AND M. GREENSTONE (2007): "The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather," *American Economic Review*, 97, 354–385.

DIFFENBAUGH, N. S., F. V. DAVENPORT, AND M. BURKE (2021): "Historical warming has increased US crop insurance losses," *Environmental Research Letters*, 16, 084025.

DING, D. K., C. CHAROENWONG, AND R. SEETOH (2004): "Prospect theory, analyst forecasts, and stock returns," *Journal of Multinational Financial Management*, 14, 425–442.

DOERING, C., GANNETT WASHINGTON BUREAU, AND U. TODAY (2013): "More drought in 2013 threatens Midwest farms," *USA today*.

DU, X., D. A. HENNESSY, AND H. FENG (2014): "A natural resource theory of US crop insurance contract choice," *American Journal of Agricultural Economics*, 96, 232–252.

DURAND, R., F. M. PATTERSON, AND C. A. SHANK (2021): "Behavioral biases in the NFL gambling market: Overreaction to news and the recency bias," *Journal of Behavioral and Experimental Finance*, 100522.

ETIENNE, X. L., S. H. IRWIN, AND P. GARCIA (2015): "Price explosiveness, speculation, and grain futures prices," *American Journal of Agricultural Economics*, 97, 65–87.

FEZZI, C., L. MENAPACE, AND R. RAFFAELLI (2021): "Estimating risk preferences integrating insurance choices with subjective beliefs," *European Economic Review*, 135, 103717.

FREDRICKSON, B. L. AND D. KAHNEMAN (1993): "Duration neglect in retrospective evaluations of affective episodes," *Journal of Personality and Social Psychology*, 65, 45.

GALLAGHER, J. (2014): "Learning about an infrequent event: Evidence from flood insurance take-up in the United States," *American Economic Journal: Applied Economics*, 206–233.

GAMMANS, M., P. MÉREL, AND A. ORTIZ-BOBEA (2017): "Negative impacts of climate change on cereal yields: statistical evidence from France," *Environmental Research Letters*, 12, 054007.

GAULE, P. AND M. PIACENTINI (2013): "Chinese graduate students and US scientific productivity," *Review of Economics and Statistics*, 95, 698–701.

HARWOOD, J. L. (1999): *Managing risk in farming: concepts, research, and analysis*, 774, US Department of Agriculture, ERS.

HAUSMAN, J., B. H. HALL, AND Z. GRILICHES (1984): "Econometric Models for Count Data with an Application to the Patents-R & D Relationship," *Econometrica*, 909–938.

HEIM, R. R. (2002): "A review of twentieth-century drought indices used in the United States," *Bulletin of the American Meteorological Society*, 83, 1149–1166.

HENDRICKS, N. P., S. SINNATHAMBY, K. DOUGLAS-MANKIN, A. SMITH, D. A. SUMNER, AND D. H. EARNHART (2014a): "The environmental effects of crop price increases: nitrogen losses in the US Corn Belt," *Journal of Environmental Economics and Management*, 68, 507–526.

HENDRICKS, N. P., A. SMITH, AND D. A. SUMNER (2014b): "Crop supply dynamics and the illusion of partial adjustment," *American Journal of Agricultural Economics*, 96, 1469–1491.

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

HERTWIG, R., G. BARRON, E. U. WEBER, AND I. EREV (2004): "Decisions from experience and the effect of rare events in risky choice," *Psychological Science*, 15, 534–539.

HOGARTH, R. M. AND H. J. EINHORN (1990): "Venture theory: A model of decision weights," *Management Science*, 36, 780–803.

HORNBECK, R. (2012): "The enduring impact of the American Dust Bowl: Short-and long-run adjustments to environmental catastrophe," *American Economic Review*, 102, 1477–1507.

KAHNEMAN, D. AND A. TVERSKY (1979): "Prospect theory: An analysis of decision under risk," *Econometrica*, 47, 263–292.

KALA, N. (2015): "Ambiguity aversion and learning in a changing world: The potential effects of climate change from Indian agriculture," *Working Paper*, available at http://create.usc.edu/research/50788.pdf. Accessed Feb 1. 2022.

KARL, T., B. GLEASON, M. MENNE, J. MCMAHON, R. HEIM JR, M. BREWER, K. KUNKEL, D. ARNDT, J. PRIVETTE, J. BATES, ET AL. (2012): "US temperature and drought: Recent anomalies and trends," *Eos, Transactions American Geophysical Union*, 93, 473–474.

KIM, T. AND M.-K. KIM (2018): "Ex-post moral hazard in prevented planting," *Agricultural Economics*, 49, 671–680.

KORNFELD, R. P., B. W. ARNOLD, M. A. GROSS, N. T. DAHYA, W. M. KLIPSTEIN, P. F. GATH, AND S. BETTADPUR (2019): "GRACE-FO: The gravity recovery and climate experiment follow-on mission," *Journal of Spacecraft and Rockets*, 56, 931–951.

KRISHNAMURTHI, L., T. MAZUMDAR, AND S. RAJ (1992): "Asymmetric response to price in consumer brand choice and purchase quantity decisions," *Journal of Consumer Research*, 19, 387–400.

LARK, T. J., N. P. HENDRICKS, A. SMITH, N. PATES, S. A. SPAWN-LEE, M. BOUGIE, E. G. BOOTH, C. J. KUCHARIK, AND H. K. GIBBS (2022): "Environmental outcomes of the US Renewable Fuel Standard," *Proceedings of the National Academy of Sciences*, 119, 1–8.

LARK, T. J., I. H. SCHELLY, AND H. K. GIBBS (2021): "Accuracy, bias, and improvements in mapping crops and cropland across the United States using the USDA Cropland Data Layer," *Remote Sensing*, 13, 968.

LOBELL, D. B. AND J. A. BURNEY (2021): "Cleaner air has contributed one-fifth of US maize and soybean yield gains since 1999," *Environmental Research Letters*, 16, 074049.

LOBELL, D. B. AND C. B. FIELD (2007): "Global scale climate–crop yield relationships and the impacts of recent warming," *Environmental Research Letters*, 2, 014002.

MCFADDEN, D. ET AL. (1973): "Conditional logit analysis of qualitative choice behavior," *Frontiers in Econometrica*, 105–142.

MENDELSOHN, R., W. D. NORDHAUS, AND D. SHAW (1994): "The impact of global warming on agriculture: A Ricardian analysis," *American Economic Review*, 84, 753–771.

METAXOGLOU, K. AND A. SMITH (2020): "Productivity spillovers from pollution reduction: Reducing coal use increases crop yields," *American Journal of Agricultural Economics*, 102, 259–280.

METAXOGOLOU, K. AND A. SMITH (2022): "Nutrient pollution and U.S. agriculture: Causal effects, integrated assessment, and implications of climate change," *Working Paper*.

MIAO, R., M. KHANNA, AND H. HUANG (2016): "Responsiveness of crop yield and acreage to prices and climate," *American Journal of Agricultural Economics*, 98, 191–211.

MOORE, F. C., N. OBRADOVICH, F. LEHNER, AND P. BAYLIS (2019): "Rapidly declining remarkability of temperature anomalies may obscure public perception of climate change," *Proceedings of the National Academy of Sciences*, 116, 4905–4910.

MOORE, H. L. (1919): "Empirical laws of demand and supply and the flexibility of prices," *Political Science Quarterly*, 34, 546–567.

NICKELL, S. (1981): "Biases in dynamic models with fixed effects," *Econometrica*, 1417–1426.

NOFSINGER, J. R. AND A. VARMA (2013): "Availability, recency, and sophistication in the repurchasing behavior of retail investors," *Journal of Banking & Finance*, 37, 2572–2585.

ORTIZ-BOBEA, A. (2021): "The empirical analysis of climate change impacts and adaptation in agriculture," in *Handbook of Agricultural Economics*, Elsevier, vol. 5, 3981–4073.

ORTIZ-BOBEA, A., H. WANG, C. M. CARRILLO, AND T. R. AULT (2019): "Unpacking the climatic drivers of US agricultural yields," *Environmental Research Letters*, 14, 064003.

PATES, N. J. AND N. P. HENDRICKS (2021): "Fields from afar: evidence of heterogeneity in United States corn rotational response from remote sensing data," *American Journal of Agricultural Economics*.

PESARAN, M. H. AND R. SMITH (1995): "Estimating long-run relationships from dynamic heterogeneous panels," *Journal of Econometrics*, 68, 79–113.

RAMSEY, S. M., J. S. BERGTOLD, AND J. L. HEIER STAMM (2021): "Field-Level Land-Use Adaptation to Local Weather Trends," *American Journal of Agricultural Economics*, 103, 1314–1341.

REJESUS, R. M., C. L. ESCALANTE, AND A. C. LOVELL (2005): "Share tenancy, ownership structure, and prevented planting claims in crop insurance," *American Journal of Agricultural Economics*, 87, 180–193.

RIPPEY, B. R. (2015): "The US drought of 2012," *Weather and Climate Extremes*, 10, 57–64.

SACKS, W. J., D. DERYNG, J. A. FOLEY, AND N. RAMANKUTTY (2010): "Crop planting dates: an analysis of global patterns," *Global Ecology and Biogeography*, 19, 607–620.

SCHLENKER, W. AND M. J. ROBERTS (2009): "Nonlinear temperature effects indicate severe damages to US crop yields under climate change," *Proceedings of the National Academy of sciences*, 106, 15594–15598.

SEIFERT, C. A., M. J. ROBERTS, AND D. B. LOBELL (2017): "Continuous corn and soybean yield penalties across hundreds of thousands of fields," *Agronomy Journal*, 109, 541–548.

SILVA, J. S. AND S. TENREYRO (2006): "The log of gravity," *The Review of Economics and statistics*, 88, 641–658.

SINGH, M. (2021): "Heuristics in the delivery room," *Science*, 374, 324–329.

TAPLEY, B. D., S. BETTADPUR, M. WATKINS, AND C. REIGBER (2004): "The gravity recovery and climate experiment: Mission overview and early results," *Geophysical Research Letters*, 31.

THALER, R. H. (2016): "Behavioral economics: Past, present, and future," *American Economic Review*, 106, 1577–1600.

TVERSKY, A. AND D. KAHNEMAN (1973): "Availability: A heuristic for judging frequency and probability," *Cognitive Psychology*, 5, 207–232.

USDA RMA (2021): "Summary of Business," `https://www.rma.usda.gov/SummaryOfBusiness/CauseOfLoss`, accessed: 2021-12-06.

VICENTE-SERRANO, S. M., S. BEGUERÍA, AND J. I. LÓPEZ-MORENO (2010): "A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index," *Journal of climate*, 23, 1696–1718.

VICENTE-SERRANO, S. M., T. R. MCVICAR, D. G. MIRALLES, Y. YANG, AND M. TOMAS-BURGUERA (2020): "Unraveling the influence of atmospheric evaporative demand on drought and its response to climate change," *Wiley Interdisciplinary Reviews: Climate Change*, 11, e632.

WALKINSHAW, M., T. O'GEEN, AND D. BEAUDETTE (2020): "Soil properties," *casoilresource.lawr.ucdavis.edu/soil-properties/*, California Soil Resource Lab, 1 Oct. 2020.

WOOLDRIDGE, J. M. (1999): "Distribution-free estimation of some nonlinear panel data models," *Journal of Econometrics*, 90, 77–97.

WRIGHT, B. (2014): "Global biofuels: Key to the puzzle of grain market behavior," *Journal of Economic Perspectives*, 28, 73–98.

WU, S., B. K. GOODWIN, AND K. COBLE (2020): "Moral hazard and subsidized crop insurance," *Agricultural Economics*, 51, 131–142.

ZHAO, Y., K. K. TANG, AND L.-L. WANG (2013): "Do renewable electricity policies promote renewable electricity generation? Evidence from panel data," *Energy Policy*, 62, 887–897.

ZHOU, W., L. R. LEUNG, AND J. LU (2022): "Seasonally dependent future changes in the US Midwest hydroclimate and extremes," *Journal of Climate*, 35, 17–27.

ZIPPER, S. C., J. QIU, AND C. J. KUCHARIK (2016): "Drought effects on US maize and soybean production: Spatiotemporal patterns and historical changes," *Environmental Research Letters*, 11, 094021.

ZULAUF, C., J. LIM, S.-H. KIM, V. DEMIRCAN, G. SCHNITKEY, AND K. HERBEL (2016): "Why is Revenue More Variable over the Growing Season in Kansas than Illinois?" *Farmdoc Daily*, 6.