ABSTRACT

WON, SUNJAE. Three Essays on the Agriculture Production Risk Management. (Under the direction of Dr. Barry K. Goodwin and Sujit K. Ghosh).

The three essays of this dissertation consider aspects of agricultural economics that are relevant to current agricultural policies, more specifically agricultural conservation programs and federal crop insurance programs in the United States. Econometric techniques, some of which are not standard in the literature, have been employed to provide additional insight into risks associated with crop production.

Chapter 1 examines the impact of planting cover crops on the risk of a failure to plant a crop (i.e., the so-called prevented planting losses). Using a novel data set that combines satellite-based cover crop information and county-level crop insurance data in the US Midwest, we find evidence that counties with higher cover crop adoption rates tend to have lower levels of crop insurance losses due to prevented planting. In addition, the resulting reduction in prevented planting risk also becomes larger with longer-term multi-year cover crop use. These results support two notions of cover crop adoption: (1) cover crops improve soil conditions such that the likelihood and magnitude of prevented planting losses decrease, (2) the benefits of cover crops accumulate over time. The study posits that the ability of cover crops to deal with excess moisture (i.e., through better water absorption and improved water infiltration in the soil) is the main factor in its ability to reduce prevented planting losses.

Chapter 2 builds crop yield risk modeling (four-directional simultaneous autoregressive model, FDSAR) to investigate how crop yield distribution varies in direction as well as distance. To accommodate spatial dependence that varies by direction, a new class of spatial models is developed using the spatial proximity matrix which different weights are assigned to neighbors in different directions. By accounting for the spatial anisotropy, the newly proposed model

generalizes the usual SAR model that gives different weights to all directions. Finally, the method is applied to a county-level data set on corn yield risks in the Corn Belt from 1951 to 2019 in the United States. These findings show that spatial dependence of corn yield risks varies by direction. These findings imply that it may be an opportune time to revisit actuarially fair crop insurance rates and redesign reinsurance contracts between spatially correlated regions.

Chapter 3 investigates the linkages between agriculture production and the incidence and nature of foodborne illnesses. By constructing county-level data by combining Centers for Disease Control and Prevention foodborne data, Bureau of Economic Analysis farm income and expense data, county health data, and Bureau of Labor Statistics food expenditure data, the findings of this study indicate that there is a significant correlation between food production and processing activity and risks associated with foodborne outbreaks. Intensive livestock farming appears to be particularly relevant to the risks of foodborne illnesses. Such production practices may weaken the immune system of farmers and neighboring residents who may be more likely to get ill from foodborne illnesses. Compared to agricultural producers, the food processing sector maintains relatively higher standards for hygiene which leads to relatively lower risks of foodborne illness. The results of this analysis have implications for assessing food safety risks. An important policy implication from the main findings is that there should be information dissemination efforts that impart knowledge to stakeholders about "the importance of occupational hygiene". This effort can enhance awareness and understanding of the benefit and likely encourage agricultural stakeholders to improve hygienic environments in their workplace.

Three Essays on the Agriculture Production Risk Management

by
Sunjae Won

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Economics

Raleigh, North Carolina
2022

APPROVED BY:

Barry K. Goodwin Co-Committee Chair	Sujit K. Ghosh Co-Committee Chair
Roderick M. Rejesus	Kathryn A. Boys

DEDICATION

This dissertation is dedicated to my wife, Chansong Woo, my newborn baby, Woori Won, and my family in South Korea for their endless support.

BIOGRAPHY

Sunjae Won began his undergraduate studies at Hanyang University in Seoul and graduated Cum Laude with a Bachelor of Arts degree in Finance and Economics. He decided to take his academic journey more through Boston University in Boston, MA with a Master of Arts degree in Economics. Before joining his Ph.D. program in 2017, he had worked at the nation's leading think tank, Korea Development Institute (KDI) in South Korea.

His research and teaching activities are mainly in the areas of applied microeconomics, especially agricultural economics. He is particularly interested in investigating how climate change would affect our society and how our society can mitigate the impact. He believes that the research experience and education at North Carolina State University can enable him to tackle productivity constraints in a sustainable manner.

ACKNOWLEDGMENTS

Joining this program took a life-changing event to get me to truly appreciate the precious and amazing life I have been blessed with. I would like to express my sincere gratitude to those who helped me carry on no matter what are the obstacles.

Dr. Barry K. Goodwin, my co-chair advisor, might have been the luckiest draw for me during my graduate studies. Dr. Goodwin provided his insights and shared his ideas for agricultural economics with me. I have learned immensely from him, and I am looking forward to doing more of the same. I am grateful to him beyond words not just for his academic support, but also for being a true friend.

A special thank goes to Dr. Sujit K. Ghosh, for contributing so much to my understanding of statistical analysis. I have been incredibly lucky to have his constant support, guidance, and patience, particularly in the most difficult times. I wish to express my deep appreciation to him for his time, effort, and expertise throughout my life at NC State University.

Dr. Roderick M. Rejesus has been a fantastic mentor to me. Dr. Rejesus always made himself available to answer questions, exchange ideas, and provide guidance. He consistently provided me with opportunities and advice to help me succeed and to cultivate myself professionally. I am truly lucky to count Dr. Rejesus as another great friend and I deeply appreciate his support, especially during the job market process. He has been and continues to be a role model.

Dr. Kathryn A. Boys has posed keen questions and intuitive suggestions which I should consider. I am very pleased to have Dr. Boys as an advisor who was always willing to lend me access to her vast knowledge. In summary, each of my committee members has contributed immensely to my success as a Ph.D. student and has helped me to blend non-trivial methodology with economic intuition. My progress would not have been possible without them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Understanding the Effect of Cover Crop Use on Prevented Planting Losses

1.1 Introduction

The agricultural sector of the United States (US) is considered one of the most vulnerable to the adverse impacts of climate change because it relies heavily on favorable weather conditions to achieve good crop yield outcomes. There is growing evidence that climate change has induced more frequent and more severe extreme weather events, such as floods and droughts, which in turn negatively affect US crop production (Karl et al., 2009; Walthall et al., 2013). Not only do these extreme weather events affect row crop production when these events occur in the growing season (i.e., May-September), more frequent and severe weather events in early spring (during planting time) also adversely affect row crop production — especially if extreme weather prevents farmers from planting their cash crops in a timely fashion. For example, the spring of 2019 was extremely wet in the US Midwest, which resulted in large numbers of farmers not being able to go to their fields and plant their row crops (i.e., prevented planting occurred on approximately 19 million acres that year).[1]

Crop insurance in the US has a provision that allows insured farmers to be compensated if they are prevented from planting due to naturally occurring, insurable causes of loss (e.g., excessive moisture). In 2019, prevented planting-related crop insurance payments reached a record high of $4.3 billion, which is almost twice the previous record ($2.2 billion in 2011). In general, the most common cause of prevented planting is excess moisture conditions (e.g., heavy rainfall), historically accounting for close to 90% of prevented-planting-related insurance payments (see footnote 1).

In light of the increasing frequency and severity of extreme weather events at planting time, use of agricultural practices that provide better resilience to extreme weather in the spring is critical to maintaining and enhancing US agricultural productivity. Planting cover crops is considered one such sustainable agricultural practice. Cover crops can help mitigate the adverse impacts of extreme weather events, reduce negative environmental externalities (e.g., reduced nutrient leaching), and at the same time provide productivity-enhancing soil health benefits to the farm operation (Snapp et al., 2005; Kladivko et al., 2014; Poeplau & Don, 2015; Kaye & Quemada, 2017; Myers et al., 2019; Rejesus et al., 2021). Cover cropping is a practice where crops – typically non-harvested legumes, grasses, or brassicas – are planted between the growing seasons of the main commodity crop (e.g., in the winter months in the US). Planting cover crops is primarily to protect and improve the soil in between periods of regular crop production (Schnepf & Cox, 2006; Arbuckle & Roesch-McNally, 2015). As an example of the potential resilience effect of cover

[1] The 2019 prevented planting acres reported here, as well as the reported proportion of prevented planting due to excess moisture in the next paragraph, are calculated based on the authors' calculations using the Risk Management Agency (RMA) "cause of loss" (COL) data.

crops, there were anecdotal reports in 2019 suggesting that farmers who have historically used cover crops generally were not prevented from planting even in the face of record rainfall levels (Sustainable Agriculture Research and Education-Conservation Technology Information Center (SARE-CTIC), 2020).[2] Therefore, providing more evidence about the relationship between cover crop use and prevented planting is important for enhancing the resilience of US agriculture.

The objective of this study is to determine whether planting cover crops help mitigate the risk of prevented planting losses in US crop production. Specifically, we are interested in exploring whether counties with higher cover crop adoption rates are more likely to have lower prevented planting-related crop insurance payments. To achieve this objective, we construct a unique county-level panel data set that includes rich information on prevented planting losses, cover crop adoption, and weather variables. The prevented planting loss information was primarily collected from the Risk Management Agency (RMA) "cause-of-loss" (COL) data, while a satellite-based remote sensing data set was utilized to gather information on cover crop adoption. The data cover twelve states in the US Corn Belt from 2005 to 2016. The county-level longitudinal data set also allows us to estimate panel data regression models (i.e., traditional linear fixed effects (FE) models and a moment-based instrumental variable (IV) FE model), which helps control for the potential endogeneity of cover crop adoption (i.e., due to time-invariant and/or time-county-varying unobservables). A variety of robustness checks using alternative specifications and estimation procedures, as well as a long-difference analysis, were also conducted to validate the strength of the results from the FE and moment-based IV-FE models.

[2] A 2019-2020 national survey conducted by SARE-CTIC indicated that 78% of self-declared cover crop users did not need to file a prevented planting claim (See SARE-CTIC (2020)). Among the 22% of surveyed cover crop users that filed a prevented planting claim, 38% agreed that prevented planting was more likely in conventionally managed, non-cover-crop fields as compared to cover-cropped fields. However, note that 55% of the cover crop users who filed a prevented planting claim also indicated that likelihood of prevented planting is roughly equal for non-cover-cropped fields and cover-cropped fields; with 9% suggesting that prevented planting is more likely in cover-cropped fields.

As alluded to above, there are already a number of studies that have suggested that planting cover crops is a "climate-smart" practice that allows farmers to be more resilient to extreme weather events (Kaye & Quemada, 2017; Aglasan et al., 2021; Hunter et al., 2021). However, to the best of our knowledge, no study has yet investigated how cover crops affect the incidence of prevented planting losses in US agriculture. Much of the prevented planting literature relates specifically to economic decisions and behavioral issues associated with the prevented planting provisions (and payments) in crop insurance. Most of these prevented planting studies have focused on the moral hazard effects of prevented planting crop insurance provisions (see Rejesus et al., 2003; Rejesus et al., 2005; Kim & Kim, 2018; Boyer and Smith 2019; Wu, Goodwin, and Coble, 2020), rather than investigating how particular practices (e.g., cover crops) influence prevented planting losses.

Hence, our main contribution is to provide empirical evidence about the relationship between the adoption of cover crops and the magnitude of prevented planting losses at the county level. This is important because the study can quantitatively validate anecdotal reports in 2019 suggesting that the incidence of prevented planting is lower for fields that have been cover cropped over time. In addition, this study complements previous agronomic studies that show that cover crops can help control the impacts of excess moisture through improved water transpiration (Menéndez et al., 2012; Volpi et al., 2017; Wang et al., 2021) and/or better water infiltration of the soil (Al-Kaisi & Helmers, 2008; Kahimba et al., 2008). Better transpiration and water infiltration results in less standing water on the soil surface during excess moisture events, and likely reduces the probability of being prevented from planting due to wet soil conditions.

The second main contribution is that we utilize a novel longitudinal county-level data set that allows us to quantitatively analyze the relationship between cover crop use and prevented planting losses over a wider geographical region (i.e., the US Corn Belt), and over a longer period of time. Much of the agronomic studies that examine how cover crops affect excess water in the soil are typically conducted only for a particular location and only for shorter time periods. Moreover, we make a contribution by being one of the first to leverage unique and innovative sources of data to analyze the resilience effects of cover crops over time. We are able to merge new publicly available RMA COL data with unique remote sensing (i.e., satellite-based) cover crop uptake data that allows us to examine the effects of cover crops on prevented planting-related crop insurance losses. Only a few studies have used the RMA COL data for economic research to date (Perry et al., 2020; Wu, Goodwin, and Coble, 2020), and studies that utilize satellite-based cover crop data have also been limited (Seifert et al., 2018; Chen et al., 2021; Connor et al., 2021).

Our empirical results indicate that cover crops help reduce the incidence of prevented planting losses in the US Corn Belt. In general, counties with higher levels of cover crop adoption rates have statistically significantly lower incidence of prevented plantings. Multi-year use of cover crops also tends to provide larger reductions in prevented planting losses. The empirical results help substantiate the anecdotal observations of 2019 that fields where cover crops have been historically used were typically not prevented from being planted to cash crops (i.e., no need to file prevented planting claims on the main cash crops), despite the excessively wet conditions that particular spring planting season. These results provide important insights about cover crop resilience that justify and inform policy discussions about how conservation programs and policies can be better supported to encourage the adoption of climate-smart practices like cover crops.

1.2 Background

1.2.1 Prevented Planting in Crop Insurance

Over the period 1988 to 1993, four ad hoc disaster payment programs were authorized to provide relief for farmers who suffered from extreme weather events (i.e., such as during the extremely wet and cool spring and growing season in 1993). Hence, to lessen the reliance on annual ad hoc disaster payments, the Federal Crop Insurance Reform Act of 1994 expanded the scope of the US crop insurance program to cover prevented planting losses as a basic component of crop insurance policies. Over time, prevented planting coverage has evolved and is now widely available for most crops insurable under the Federal crop insurance program.

Prevented planting coverage is designed to compensate farmers for the pre-planting costs incurred up to the point of not being able to plant the crop (USDA-RMA, 2020). Prevented planting payments have been adjusted over the years to align with the estimated pre-planting sunk costs by updating the prevented planting coverage factors (USDA-RMA, 2016). Prevented planting payments are triggered when the farmer fails to plant an insured crop with the proper equipment by a designated final planting date due to insured causes of loss. To be eligible for prevented planting payments, the cause of prevented planting loss must be common in the area and should have also prevented other similar producers in the immediate vicinity from planting (Rejesus et al., 2003).

If an insured farmer is prevented from planting a first insured cash crop by the final planting date, the farmer has the following four options available: (1) plant the first crop during the late planting period, (2) plant a second (alternative) crop after the late planting period, (3) leave the acreage unplanted (e.g., leave the field fallow), or (4) plant a summer cover crop after the late planting period. For the first two options (1 and 2), the producer can receive a partial prevented planting payment, while the last two options offer the full prevented planting indemnity payment.[3] Based on a report by the US Department of Agriculture (USDA) Office of the Inspector General (OIG) (USDA-OIG, 2013), about 99% of insured producers receiving a prevented planting payment left their indemnified acreage fallow (i.e., third option), which indicates a trivial incentive for farmers to use land after being compensated for prevented planting.[4]

Of the $7.6 billion in total indemnities paid out by the US crop insurance program in 2020, about 27% ($2.06 billion) was for prevented planting claims (USDA-OIG, 2013). About 10 million acres were prevented from planting in 2020, which accounts for around 22% of total acres indemnified. Overall, the amount of prevented planting acres and the amount of prevented planting payments in the US crop insurance program is a good quantitative measure of the year-to-year prevented planting levels in US agriculture. We use these crop insurance-based measures of prevented planting as our main dependent variable in the empirical analysis, though we also use an alternative Farm Service Agency (FSA) acre-based prevented planting measure as a robustness check.

1.2.2 Cover Crops and Wet Soils

The most common cause of prevented planting claims in crop insurance is excess moisture conditions (e.g., heavy rainfall, floods) (See Figure A.1). Karl et al. (2009) predict that precipitation events will change in frequency and intensity due to climate change, with a projected increase in spring precipitation, particularly in the Northeast and Midwest US. In excessively wet conditions, farmers usually decide not to plant their intended crop acreage in a timely manner

---

[3] If the farmer plants a cash crop after the final planting date, the insurance guarantee is reduced by 1 percent for each day up until 25 days and the guarantee is fixed at 55% for corn and 60% for soybeans after 25 days. Note that if the fourth option is chosen (i.e., planting a summer cover crop), it may not be harvested before November 1.

[4] Note that this raises a separate issue where the prevented planting payments may reduce incentives for summer cover crop planting (Boyer & Smith, 2019), and in turn lead to more soil degradation because of a lack of plant canopy on fallowed acres.

because wet soils make fieldwork difficult or even impossible (i.e., restricting the mobility of equipment) and, if one continues with the fieldwork in wet soils using heavy equipment, then this action may irreparably damage soil structure (e.g., soil compaction). This can then cause reduced soil health and lower cash crop productivity.

When cover crops are grown, they transpire (i.e., draw water from the soil) in excess of what would be lost through soil evaporation alone (USDA-NRCS, 2013). Cover crops can use excess soil moisture, drying the soil and improving trafficability across the field (i.e., likely decreasing the likelihood that one will be prevented from planting). Some farmers use winter cereals like wheat or rye to have established growing plants in the spring that will use water and dry surface soils and provide an adequate seedbed for the desired cash crop such as corn or soybean (USDA-NRCS, 2013). Cover crop species with strong taproots can also alleviate soil compaction issues by growing through the compaction layers, creating pathways and holes for downward water movement, ultimately enhancing soil infiltration. In addition, better soil infiltration provides more room in the soil profile, which reduces flood severity and decreases the adverse impact of excess moisture on the soil. Overall, the agronomic mechanisms discussed above lay the foundation for why cover crops have the potential to reduce excess soil moisture and can decrease the probability of farmers being prevented from planting in the spring.

1.3 Data Description

The county-level panel data set constructed for this study covers the period 2005 to 2016 and is based on information collected from a variety of sources. The main dependent variables of interest are prevented planting measures based primarily on loss data from the RMA COL database. The RMA COL database has detailed county-level information on the amount of indemnities and the amount of acres indemnified due to various causes of loss (including prevented-planting-related losses). We only consider the prevented-planting-related losses that are identified through "stage codes" in the RMA-COL database.[5] Furthermore, we limit our data to the following most commonly purchased individual crop insurance plans where the prevented planting option is available: Yield Protection (YP), Revenue Protection (RP), and Revenue Protection with Harvest Price Exclusion (RPHPE).

The main dependent variables used in the study are derived from the RMA-COL and the RMA Summary of Business (SOB) data, which can be categorized as "indemnity-based" and "acre-based" measures.[6] The "indemnity-based" dependent variable is a prevented-planting-related loss cost ratio (PP-LCR) calculated by taking the ratio of total prevented-planting-related indemnities over total liabilities. On the other hand, the "acre-based" dependent variable represents the proportion of total insured acres with prevented-planting-related indemnities, which we call the PP-Loss Acres Ratio (PP-LAR). Note that the main dependent variables used in the study are "normalized" by some factor (i.e., the denominator of the ratio) so that potential scale effects are controlled for (i.e., control for larger counties having more aggregate prevented planting claims by virtue only of its size). In our empirical analysis, we also utilize the percentage form (%) of the PP-LCR and PP-LAR variables (i.e., multiply the ratios by one hundred).

The county-level data for cover crop adoption rates (i.e., the main independent variable of interest in percent form) are drawn from the Operational Tillage Information System (OpTIS) developed by Dagan Inc.®, Applied GeoSolutions (AGS) for the period 2005-2016.[7] OpTIS

---

[5] See Appendix Table A.1 for more details about the stage codes used to identify prevented-planting-related indemnities and insured acres with prevented planting indemnities.

[6] Note that the RMA SOB data includes all insured farms, while the RMA-COL data only includes insured farms that experienced a loss. Therefore, observations with no losses were coded as having zero prevented-planting-related indemnities or acres.

[7] Dagan Inc. AGS (now called Regrow Ag), a geospatial analysis company, partnered with the Conservation Technology Information Center (CTIC) to create OpTIS and generate satellite imagery of back dated cover crop coverage going back to 2005. This effort was funded by USDA, Monsanto, John Deere, Soil Health Partnership, the Indiana Soybean Alliance and Indiana Corn Marketing Council.

produces remotely sensed (satellite-based) data on conservation practices, including the planting of winter cover crops. The OpTIS calculations are performed and validated at the farm-field scale, but the privacy of individual producers is fully protected by spatially aggregating the results to the county level or higher (Hagen et al., 2020). The county-level OpTIS cover crop adoption data utilized here cover 646 counties over 12 States in the US Corn Belt – Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, Ohio, Oklahoma, South Dakota, and Wisconsin.[8]

Validation of the OpTIS cover crop adoption data was mainly done through comparisons with photo and roadside survey information collected at the field level for several representative counties (See Hagen et al. (2020) for more details on the validation procedure). In their validation analysis of the full OpTIS data (in over 30 counties across twelve states on 1195 fields), Hagen et al. (2020, p.12-13) suggest that the remote-sensing based OpTIS data for cover crop adoption is 87.9% accurate and the false-positive rate is only at 0.03. Notwithstanding the relatively high accuracy of the OpTIS cover crop data based on the validation procedure described above, note that there are still known discrepancies between the OpTIS-estimated cover crop adoption acres vis-à-vis other aggregate cover crop data sets (i.e., like those from the Census of Agriculture (AgCensus)) (Hagen et al., 2020).[9] The differences in the cover crop adoption estimates among these two datasets likely stem from the different methods used to collect the data. For example, the AgCensus relies on surveys of the complete census of growers and likely captures their intent to grow cover crops, and whether they indeed planted cover crops at the time of the survey.

---

[8] All counties in Illinois, Indiana, and Iowa are covered in the OpTIS data, while partial sets of counties are coved in the remaining states, See Hagen et al. (2020).

[9] The county-level AgCensus and EWG cover crop adoption data were not utilized in the study because of the limited number of years these data are available. For example, the cover crop adoption data from the AgCensus are only available in 2012 and 2017. Hence, the AgCensus data do not allow for analysis over a longer period of yearly time-periods, while the OpTIS data does.

Therefore, a farmer may indicate that he/she planted (or intend to plant) cover crops in the winter, which is recorded in the AgCensus as adoption. Nonetheless, it is possible that a cold snap and wet conditions prevented the actual establishment of a canopy that can be detected by satellites. In this case, OpTIS will not register a cover crop, while the AgCensus might. In general, one cannot say that one data source is superior to another, and each (OpTIS or AgCensus) is a reasonable source of cover crop adoption data. It should also be noted that the OpTIS data now has a track record of being used in agricultural economics research published in peer-reviewed journals (Chen et al., 2021; Connor et al., 2021).

Note that a crop year in the OpTIS data set is from November 1 of the previous year through October 31 of the year when the subsequent cash crop is planted. For example, the 2015 crop year extends from November 1, 2014 through October 31, 2015. Cover crop adoption occurs in the winter months after the harvest of the row crop from the previous year and before planting the subsequent cash crop. For example, the OpTIS cover crop adoption data in the crop year 2015 reflects acres planted to cover crop detected by the satellites starting in November 2014, after the harvest of the cash crop in the Fall of 2014. We mainly utilize information on the proportion of cover crop acres planted after corn, soybeans, small grains, or other cash crops.

Lastly, in addition to data on prevented planting and cover crop adoption, we also collected weather data from the Parameter Regression Independent Slopes Model (PRISM) climate group to serve as control variables in our analysis. This data source has been utilized in a number of studies that examined a variety of climate change issues over the years (See Annan & Schenkler, 2015 for example), and is considered one of the best sources of US weather and climate-related data. Data on the following weather-related variables were collected and used in our analysis: a degree day measure below 10°C (DD10), a degree day measure between 10 and 29°C (DD1029), a degree day measure above 29°C (DD30+),[10] and the Palmer Drought Severity Index (PDSI) for wet (PDSI_W) and dry (PDSI_D) conditions.[11] Since our main dependent variable is a prevented-planting measure, the pertinent time frame for the analysis is the typical April to May cash crop planting window in the US Midwest. Thus, the weather variables are aggregated for the months of April and May. We include these weather-related variables as controls in our specification since environmental conditions in each county vary over time and may influence the amount of prevented planting each year.

Descriptive statistics for the main variables used in this study are summarized in Table 1.1. The geographical coverage of the study is depicted in Figure 1.1, and information about the PP-LCR and PP-LAR variables in the study area for the last year in our data (2016) is presented as well. In addition, the year-to-year average values of the prevented planting measures and the cover crop adoption rates can be seen in Figures A.2 and A.3, respectively. Although there is year-to-year variation, these graphs indicate that prevented planting payments have generally increased over the last decade, and cover crop adoption has an increasing trend over the last fifteen years.

1.4 Empirical Specification and Estimation Strategies

We estimate the county-level impact of cover crop adoption on prevented planting according to the following empirical specification:

$$PP_{it} = \beta_1 CC_{it} + \beta_2 W_{it} + \alpha_i + \gamma_t + \varepsilon_{it} \qquad (1)$$

[10] We use the Schlenker and Roberts (2009) approach (and their recommended cutoffs) for calculating the degree-day measures used in this study. These degree-day measures have been widely used in the climate econometrics literature (e.g., Annan & Schenkler, 2015; Burke & Emerick, 2016; Lusk et al., 2019; Wang et al., 2021a), especially when estimating the effects of rising temperatures on crop yields. This is primarily because of the ability of degree-day measures to capture non-linearities in the temperature effects.

[11] The county-level PDSI data are from the Centers for Disease Control and Prevention (CDC) National Environmental Public Health Tracking Network. PDSI values range from -10 to +10. As alluded to above, negative PDSI values reflect dryness, while positive PDSI values reflect wetness. Typically, PDSI values of -4 or below represents extreme drought, while PDSI values of 4 or above reflects an extremely wet environment (i.e., flood conditions). More details about why PDSI and the other weather covariates were chosen in the specification are discussed in the empirical specification section.

where PP_{it} is the prevented planting loss measure (i.e., PP-LCR or PP-LAR in %) for county i and year t; CC_{it} is the cover crop adoption variable (% of planted crop acres with cover crops) in county i and year t; $W_{it} = \left(W_{it}^1, \dots, W_{it}^k\right)$ is a $1 \times k$ vector of aggregate weather covariates (from April to May) that include DD10, DD1029, DD30+, PDSI_W, and PDSI_D; α_i is the county-fixed effect (i.e., to capture the unobservable time-invariant county variation); γ_t is the year (or time) fixed effects (i.e., to capture time-varying unobservables that affect all counties, such as macroeconomic shocks); β_1 and β_2 are parameters to be estimated (with β_1 as our main coefficient of interest); and ε_{it} is the idiosyncratic error term.

It is important to note here that the CC_{it} cover crop adoption variable represents cover crop adoption in the winter months prior to the occurrence of potential prevented planting claims PP_{it} (i.e., typically in the April to May window). That is, the timing is such that CC_{it} occurs before PP_{it} claims. In addition, the vector of weather covariates is included in the specification since weather conditions at planting time influence the likelihood and magnitude of prevented-planting-related losses (Chen and Miranda 2007; Kim and Kim 2018). The three "degree-day" variables are included to control for the nonlinear effects of temperatures on prevented planting. In particular, DD10 is included in the specification because colder temperatures typically influence planting decisions. Cooler soil temperatures at planting tend to adversely affect crop growth and so producers avoid planting when the soil is too cold (Elmore 2012).

Moisture levels (or measures of water availability) are also incorporated in the empirical specification since the majority of the prevented planting losses in the RMA data are due to excess moisture conditions (See Figure A1.1). We use two PDSI variables as measures of water availability, rather than quadratic functions of precipitation or rainfall levels.[12] A drought index

[12] Although we use PDSI in our main specification, note that we also conduct a robustness check below where we

like the PDSI is appropriate as a measure of water/moisture availability because its values are referenced to local climate, which allows one to calculate dryness or wetness relative to local norms (Xu et al., 2013; Kolář et al., 2014; Rejesus et al., 2015; Wang et al., 2021b). In addition, local soil attributes are partly accounted for when measuring these drought indices. Local soil type and quality are important factors in a crop's ability to handle extreme dryness or wetness. Using both the positive and negative PDSI values in our specification also adequately accounts for nonlinearities in the effects of water availability (i.e., typically reflected by having a quadratic precipitation term in previous studies).

Given the panel nature of our county-level dataset, we utilize a traditional linear panel fixed effects (FE) model to estimate equation (1). The FE model allows us to address endogeneity due to time-invariant unobservables (i.e., the so-called unobservable heterogeneity across counties). For example, the overall soil quality of the county (as proxied by the predominant soil types in the area (e.g., clay, sandy, loam, etc.)) is considered roughly time-invariant in our county-level context, and these variables may influence both cover crop use and the likelihood and magnitudes of prevented planting payments. These kinds of time-invariant unobservables are absorbed by the county-fixed-effects (α_i) included in the specification and allow for better identification of the cover crop adoption impacts. In addition, the time-fixed-effects (γ_t) control for time-varying shocks that affect all counties in the sample in a particular year the same way (e.g., unobserved macroeconomic shocks affecting all counties), and again helps in addressing potential endogeneity issues. To facilitate proper inference, we also use standard errors clustered by county to account for potential year-to-year correlations within a county.

utilize a quadratic precipitation specification.

Notwithstanding the desirable properties of the traditional linear panel FE models, one potential drawback of these approaches is the potential presence of residual endogeneity (even after controlling for unobserved heterogeneity and including a number of relevant controls in the specification). This residual endogeneity can be due to time-county-varying unobservables (i.e., time-varying county-specific unobservables) affecting both the outcome (PP_{it}) and the main independent variable of interest (CC_{it}). In this case, the typical approach is to use instrumental variable (IV)-based panel FE models (IV-FE), where the IVs are correlated with the potentially endogenous main independent variable but uncorrelated with the outcome variable (i.e., satisfying the IV exclusion restrictions). Unfortunately, we do not have any time-county-varying IVs available that strongly satisfy these so-called traditional "exclusion" restrictions. Therefore, we implement a recently developed moment-based IV approach (see Lewbel, 2012) to address the possible residual endogeneity due to correlations between time-county-varying unobservables that affect county-level prevented planting measures and county-level cover crop adoption rates.

The moment-based IV estimator utilizes heteroscedasticity in the error terms from the first-stage regressions (e.g., regression of the potentially endogenous variable on the observable covariates) to identify the coefficients of the endogenous variables in the main equation even in the absence of exclusion restrictions or valid instruments. According to Lewbel (2012), the model is identified if the error terms in the first-stage equations are heteroscedastic. That is, if a subset (or all) of the exogenous control variables is correlated with the variance of the first-stage error terms but not with the covariance between the first-stage error term and the error term in the main second-stage equation, then the subset of the exogenous covariates in mean-centered form multiplied by the residuals from the first-stage equation are valid instruments.

More formally, with equation (1) as our main estimating equation and the cover crop adoption measure (CC_{it}) as the potentially endogenous variable, the first-stage regression in the Lewbel (2012) approach can be defined as follows:

$$CC_{it} = \gamma_1 W_{it} + \alpha_i + \gamma_t + u_{it} \quad (2)$$

where γ_1 is a parameter to be estimated and u_{it} is an idiosyncratic error term for the first-stage cover crop adoption rate equation. The vector W_{it} again represents the exogenous weather variables included in the main specification. The parameters α_i and γ_t are the county and year fixed effects. In the presence of heteroscedasticity in (2) (i.e., $Cov(W_{it}, u_{it}^2) \neq 0$), Lewbel (2012) has shown that $(W_{it} - \bar{W})u_{it}$ can be used as valid IVs in a standard two-stage least squares (2SLS) procedure. We use the Breusch-Pagan test to validate the presence of heteroscedasticity in our first-stage regression (Breusch and Pagan 1979). In our empirical context, the Breusch-Pagan statistic is 2542.68 (p-value is <0.0001) for the first stage regression, which allows us to reject the null hypothesis of homoskedasticity. This means that we can use the moment-based IV approach of Lewbel (2012) as an alternative estimation procedure that can help address residual endogeneity due to time-county-varying unobservables.

Aside from the Breusch-Pagan test, we also implemented other statistical tests to assess the strength of the Lewbel (2012) IV approach in terms of properly identifying the cover crop effect. First, we use the Kleibergen-Paap rk LM test to evaluate whether IV approach used is underidentified (Kleibergen & Paap, 2006). The Kleibergen-Paap rk LM test rejects the null hypothesis that the IV model used is underidentified (i.e., see the lower panel of Tables 1.2 to 1.4 below). In addition, we use the Kleibergen-Paap rk Wald test to evaluate whether the IVs used in the estimation are weak or strong. Based on the large values of the calculated Kleibergen-Paap rk Wald statistics (i.e., again see the lower panel of Tables 1.2 to 1.4 below), we reject the null hypothesis that the instruments are weak. Lastly, we implement the Hansen J overidentification test to assess the validity of all IVs used in the Lewbel (2012) procedure (see bottom panel of Tables 1.2 to 1.4). The test statistics from the Hansen J tests reject the null hypothesis that all the IVs used in the Lewbel (2012) estimation are valid (i.e., at least one of the IVs in the estimation may not be valid). Notwithstanding this Hansen J test result, Baum and Lewbel (2019) indicate that it is possible for the Hansen J test to fail to support the validity of all IVs used, but the key assumptions for the moment-based IV model to work still hold in general (i.e., especially if the heteroscedasticity condition and the tests for IV strength still holds). With the majority of the diagnostic tests supporting the Lewbel (2012) IV approach, we believe it provides valid results and we thus proceed to utilize it in our empirics. In light of the foregoing discussion, we also implement an "external-IV-free" estimation method developed by Kiviet (2020) as another robustness check to validate results from the linear FE and moment-based IV approaches.

1.5 Results and Discussion

1.5.1 Main Empirical Specification Results

Table 1.2 presents the parameter estimates from the traditional linear panel FE and moment-based IV estimation of equation (1). Results from the two regression approaches indicate that cover crop adoption has a negative and statistically significant effect on the magnitude of prevented planting indemnities observed at the county level. Counties with higher levels of cover crop adoption are more likely to have significantly lower levels of prevented planting losses. The parameter estimates for the cover crop adoption variable are fairly similar in magnitude for both the linear FE models and the moment-based IV model, suggesting that the main unobservables are likely time-invariant and/or county-invariant such that the county and time-fixed effects in the linear panel FE model sufficiently account for endogeneity concerns. For the PP-LCR model, a one percentage point cover crop adoption increase will result in approximately a 0.04 percentage point reduction in prevented-planting-related loss cost ratios. For the PP-LAR model, we find that a one percentage point cover crop adoption increase in a county will result in approximately a 0.05 percentage point reduction in the proportion of insured acres with prevented planting losses. Hence, we find consistent empirical evidence that cover crop adoption reduces the incidence and magnitude of prevented planting losses in US agriculture. This main finding is consistent with previous agronomic studies that indicate that cover crops are better able to deal with excess soil moisture conditions (i.e., through better water absorption and improved water infiltration in the soil), and consequently reduce the likelihood and extent of prevented planting losses.

To better contextualize the magnitude of the estimated cover crop effect, we conduct a simple calculation based on 2019 crop insurance data. Total insured liabilities in 2019 were approximately $96.73 billion for the YP, RP, and RPHPE products, and total PP-related indemnities were about $4.32 billion. In this case, this means that the PP-LCR was about 4.47% in 2019. Our estimates suggest that the PP-LCR would be reduced to 4.43% (i.e., 4.47% - 0.04%) if the cover crop adoption rate had been increased by one percentage point in 2019 (i.e., about 3.9 million acre increase in cover crop use nationwide). If the liability amount remained at $96.73 billion, then the total reduction in prevented-planting-related indemnities would have been approximately $38.69 million. Hence, a relatively small percentage increase in cover crop adoption results in a large reduction in prevented-planting-related indemnities. A similar calculation for the PP-LAR results indicates that a one percentage point increase in the cover crop adoption rate will decrease the acreage with prevented planting-related losses by about 0.12 million acres. These calculations reflect a fairly substantial reduction in prevented planting losses given a relatively small increase in the US cover crop adoption rate.

With regards to the weather variables that serve as controls in our specification, the estimated county-level effects largely follow a priori expectations (Table 1.2). Prolonged colder conditions (i.e. higher April-May DD10) tend to have a positive and statistically significant effect on prevented planting losses. This is consistent with the observation that producers tend to avoid planting when the soil is too cold. In addition, we also find that the April-May PDSI_W has a positive and statistically significant effect on prevented planting losses. This supports the common observation that wetter conditions tend to be a major factor in limiting farmers' ability to plant their cash crops in a timely manner. Lastly, we find the opposite result for the April-May PDSI_D, where we find a negative and statistically significant effect, suggesting that drier conditions make it easier for farmers to plant on time (thereby reducing the incidence of prevented planting losses).

1.5.2 Robustness Checks

To validate the strength and stability of our main model results, we conduct several robustness checks that consider the following: (i) a PP-LAR measure from an alternative source (instead of the RMA-based PP-LAR measure), (ii) a more homogeneous sample that only includes prevented planting losses from RP insurance plans, (iii) an alternative two-way standard error clustering procedure by county-year, (iv) various alternative empirical specifications using different sets of weather variables as controls, (v) a specification that uses time trends instead of time-fixed effects, and (vi) a recently developed "external-IV-free" estimation procedure (called "kinky least squares" (KLS) regression) that can serve as an alternative to the Lewbel IV approach.

First, we ran a robustness check where we used an alternative PP-LAR measure based on data collected from the FSA instead of the prevented planting data from the RMA. Since the FSA does not oversee the US crop insurance program, this agency only has information on the number of acres in a county where farmers were prevented from planting. Hence, we are only able to construct an analogous PP-LAR dependent variable based on FSA data (and not an alternative

FSA-based PP-LCR measure). To assess the similarity between the FSA-based PP-LAR measure and the RMA-based PP-LAR measure, we graphed these two PP-LAR variables (as well as the raw prevented planting acres data from both sources), with the FSA data points on the x-axis and the RMA data points on the y-axis (See Figure A.4). The figure suggests that the data from the two sources are fairly similar (most points are at, or near, the 45-degree line). The strong positive correlation coefficient between the RMA-based and the FSA-based PP-LCR (R =.9472) also supports the contention that these two data sources are comparable. Results of our main regression using the FSA-based PP-LAR data are presented in Table 1.3 (Panel A). The parameter estimates from this FSA-based model are consistent with the estimates from the RMA-based model in Table 1.2 and also suggest that counties with higher cover crop adoption rates tend to have a lower proportion of planted crop acres that are prevented from planting. The magnitudes of the effects are also similar to our baseline results.

Second, we estimate the empirical specification in equation (1), but limit the analysis to only consider prevented plantings in the RP insurance plan. We do this since RP is the most widely purchased crop insurance product for the study period under consideration. Results from this RP robustness check are shown in Table A.2. Findings are consistent with our main result that cover crops tend to lower the incidence and magnitude of prevented planting losses. However, note that parameter estimates from the RP insurance robustness check are larger than the main model estimates in Table 1.2. Third, instead of one-way standard error clustering by county (as is used in our main estimation strategy), we also use a two-way standard error clustering procedure by county-year. Allowing for clustering "by year" accounts for potential spatial correlation across counties for a particular year. Results are shown in Table A.3 and inferences are the same as the main model results (i.e., negative and significant cover crop adoption effect on prevented planting losses).

Fourth, we ran a battery of robustness checks where we utilized different sets of weather variables in the empirical specification (Tables A.4 to A.7). In our main empirical specification, we utilized aggregate April-May weather variables. Thus, we conduct robustness checks where we used separate "per month" April and May weather variables in the specification (Table A.4), more aggregate April-June three-month weather variables instead of two-month April-May weather variables (Table A.5), and separate "per month" April, May, June weather variables (Table A.6). Findings from all three of these robustness checks with alternative weather specifications still support our main result that higher county-level cover crop adoption results in lower prevented planting losses (albeit the magnitudes of the cover crop coefficients tend to be slightly lower when using the three-month April to June weather variables). We also ran an alternative weather control specification where we use precipitation and precipitation squared to account for nonlinear moisture levels in the April to May months, instead of using the PDSI_W and PDSI_D measures (Table A.8). The main finding holds with this specification, where higher county-level cover crop adoption rates have a negative and statistically significant relationship with prevented planting losses. In addition, we find that precipitation tends to have a positive linear relationship with prevented planting claims given that the linear precipitation term is positive and significant while the quadratic specification term is statistically insignificant.

Fifth, we also ran a robustness check where we replaced the time-fixed effects with time trends. In some cases (especially with low sample sizes), having both county and time-fixed effects in an empirical specification absorbs a lot of the variation in the data such that "the signal" from the independent variable is effectively diluted and cannot be effectively detected statistically. Even though our baseline models with time-fixed effects still show a strong signal that cover crops have a statistically significant negative effect on prevented planting losses (and we have a fairly large sample size), we still do a robustness check to determine if results change using a specification with time trends instead of time-fixed effects. Results from this robustness check still support our main result where counties with high cover crop adoption have less prevented planting losses (Table A.7).

Lastly, we implement a recently developed KLS regression method as a robustness check.[13] This "external-IV-free" estimation procedure is an alternative to the moment-based Lewbel (2012) IV approach (which does not require external IVs as in the traditional 2SLS procedure). The KLS approach achieves identification of the estimated regression parameters by confining the admissible correlation of the regressors with the error term within plausible bounds. No excluded instruments are required. Instead, the potential bias in non-IV ordinary least squares (OLS) type procedures is analytically corrected on a grid (or range) of endogeneity correlations that the analyst assumes. This provides a set of coefficient estimates in accordance with the postulated endogeneity range. In our empirical context, we posit that the range of residual endogeneity present in our baseline specification (equation 1) would be minimal given that we already control for time-invariant and county-invariant unobservables through the county and time-fixed effects. We also argue that the endogeneity correlations are positive such that the potentially endogenous cover crop variable is positively correlated with the remaining time-county-varying unobservables in the error term (i.e., unobserved soil conservation effort that may be correlated with cover crop adoption). Hence, it is reasonable to assume that the range of endogeneity correlation in our context

[13] In the interest of brevity, we do not discuss details of the KLS regression method here. We direct the reader interested in the details of this method to following references: Kiviet (2013), Kiviet (2020), and Kripfganz and Kiviet (2021).

is only between 0.1 to 0.4. Given this correlation range, we implement the KLS procedure using our baseline specification and present results in Tables A.9 and A.10 for the PP-LCR and PP-LAR dependent variables, respectively. For this reasonable endogeneity correlation range of 0.1 to 0.4, the KLS regression results support the main conclusion that counties with higher cover crop adoption tend to have lower prevented planting losses.

1.5.3 Long Differences Analysis

The empirical specifications and estimations we used so far consider the short-term impacts of cover crop use on prevented planting risk. If the beneficial impacts of cover crops on soil health (and climate resilience) take time to materialize fully (i.e., it accumulates over time), then the estimated short-term impacts presented above might understate the long-term benefits of continuous cover crop use on prevented planting losses. To examine the longer-term effect of cover crops on prevented planting losses, we adopt a "long-differences" approach and model county-level "changes" in prevented planting losses over time as a function of long-term "changes" in cover crop adoption (and weather).[14]

We construct long-term variables of the prevented planting loss measures, cover crop use, and weather variables at two different points in time for each county, and calculate changes in the average prevented planting loss measures as a function of changes in cover crop use and changes in weather variables. More formally, consider two multiyear periods denoted τ_1 and τ_2, each spanning n years. Our approach is to sum over all the years in each period. That is, the average prevented planting loss measured in the period τ_1 is calculated as follows: $\overline{PP}_{i\tau_1} = \frac{1}{n}\sum_{t\in\tau_1} PP_{it}$.

[14] The long-differences approach has been applied in many studies that investigate the long-term impact of many kinds of right-hand-side variables. In particular, this econometric strategy has been often employed to show evidence of detrimental climate change impact (i.e., long-term weather patterns) on longer-term crop productivity (Hsiang, 2016; Burke & Emerick, 2016; Lobell & Asner, 2003).

The average cover crop use, $\overline{CC}_{i\tau_1}$, and the average weather variables, $\overline{W}_{i\tau_1}$, is calculated similarly over the same period τ_1. The equation for period τ_1 can then be expressed as:

$$\overline{PP}_{i\tau_1} = \hat{\beta}_{1\tau_1}\overline{CC}_{i\tau_1} + \hat{\beta}_{2\tau_1}\overline{W}_{i\tau_1} + \alpha_i + \bar{\varepsilon}_{i\tau_1} \quad (3)$$

Similarly, the equation for the period τ_2 can be obtained based on the calculations above. Using two equations for the two periods, we can construct a "long difference" model:

$$\left(\overline{PP}_{i\tau_2} - \overline{PP}_{i\tau_1}\right) = \hat{\beta}_1\left(\overline{CC}_{i\tau_2} - \overline{CC}_{i\tau_1}\right) + \hat{\beta}_2\left(\overline{W}_{i\tau_2} - \overline{W}_{i\tau_1}\right) + (\alpha_i - \alpha_i) + (\bar{\varepsilon}_{i\tau_2} - \bar{\varepsilon}_{i\tau_1}) \quad (4)$$

By dropping time-invariant county-fixed effects, the above equation can then be rewritten as:

$$\Delta\overline{PP}_{i\tau} = \hat{\beta}_1\Delta\overline{CC}_{i\tau} + \hat{\beta}_2\Delta\overline{W}_{i\tau} + \Delta\bar{\varepsilon}_{i\tau} \quad (5)$$

where $\hat{\beta}_1$ represents the extent to which long-term average cover crop use influences long-term average prevented planting losses.

In our long-difference approach, we split the data into two 6-year periods where the first period (τ_1) is from 2005 to 2010 and the second period (τ_2) is from 2011 to 2016. As in the traditional linear FE model, the long difference model in (5) can be estimated by OLS. Note that the first difference specification in (5) also accounts for time-invariant unobservables. Hence, the specification in (5) allows one to quantify how longer-term use of cover crops influences the longer-term prevented planting risk while accounting for time-invariant unobservables.

Results from the long-difference method are presented in Table 1.3 (Panel B). We find that counties with higher long-term use of cover crops have statistically significantly lower prevented planting risk in the long term. In addition, the magnitude of the cover crop use effect on prevented planting risk in the long-difference model is larger than the magnitudes from our short-term specification. This supports the notion that soil health benefits and climate resilience effects of cover crop use tend to accumulate over time such that longer-term use of cover crops would result in larger reductions in the likelihood and magnitude of prevented plantings losses (Boyer et al., 2018; Thompson et al., 2020). Specifically, our results suggest that the magnitude of the prevented planting resilience effect of cover crops is doubled in the long difference models (as compared to our more short-term baseline models).

1.6 Conclusion

Extreme weather events like excess rainfall or flooding in the spring months have historically been the most common reason that US row crop farmers are prevented from planting. This usually results in millions of dollars in prevented planting-related crop insurance losses. With climate change increasing the frequency and intensity of these extreme weather events, the incidence and magnitudes of prevented planting losses will also likely continue to grow in the future, unless agricultural practices can be identified to reduce the impact of extreme weather events on prevented planting. Planting cover crops is considered one such practice, especially given the anecdotal experience from the excessively wet spring of 2019 in the US Midwest. At that time, it was observed that farmers who have regularly used cover crops in their fields were generally not prevented from planting. However, there is still no systematic empirical evidence that carefully documents the relationship between cover crop use and prevented planting in US agriculture. This study aims to fill this gap in the literature and empirically examines whether planting cover crops help reduce the risk of prevented planting losses. We construct a unique county-level panel data set that merged novel satellite-based information on cover crop adoption with publicly available data on prevented planting losses and weather variables. The data used in the analysis encompasses twelve States in the Corn Belt for the period 2005-2016. Moreover, traditional linear panel FE models, a moment-based IV model, and several robustness checks (including a long-difference approach) are used in the empirical analysis.

Our empirical findings indicate that the extent of county-level cover crop use has a negative and statistically significant effect on prevented planting losses. A number of robustness checks validated this conclusion. Results from these robustness checks are consistent with the findings from the main model runs. Overall, the central finding from our study suggests that counties with higher cover crop adoption rates tend to have smaller prevented planting losses. Our results are consistent with the notion that planting cover crops can help control excess water in the soil through improved transpiration and better water infiltration. In turn, the improved soil conditions help reduce the likelihood of being prevented from planting. In addition, using a long-difference procedure, we find evidence that longer-term use of cover crops would allow for more accumulation of soil health benefits over time, and results in larger reductions in prevented planting risk in the long-term.

The core empirical finding in our study validates the anecdotal reports from 2019 in the Midwestern US where farmers with cover cropped fields were typically not prevented from planting. Thus, the results of this study advance our understanding of the relationship between cover crop use and prevented planting, and provide support for the contention that cover cropping can enhance resilience to extreme weather events. Hence, an important policy implication from this main finding is that there should be renewed outreach and information dissemination efforts that impart knowledge to stakeholders about the "prevented planting resilience benefit" of cover crops. This prevented planting benefit has not been noted as a core economic benefit of cover crop use. If farmers do not know about the potential benefits of cover crops, then this lack of knowledge can be a barrier to adoption (especially if costs of cover crop adoption are viewed as higher than its benefits without this prevented planting benefit accounted for in the decision calculus). Federal agencies, such as the Natural Resource Conservation Service (NRCS), and the Cooperative

Extension Services in each state should list "potential reduction in prevented planting risk" as an additional benefit of cover crops in their outreach materials. A coordinated outreach and education effort among different federal and state agencies, as well as agricultural stakeholder groups, can enhance awareness and understanding of the multiple benefits of cover crops – including lowering prevented planting risk – and likely encourage further adoption of this practice.

Moreover, given the potential prevented planting risk reduction benefits of cover crops (among other private and environmental benefits), it may also be prudent for NRCS to re-evaluate the prioritization, funding levels, and level of technical assistance given to cover cropping as a practice supported by the two main federal conservation payment programs for "working lands" in the US – the Environmental Quality Incentives Program (EQIP) and the Conservation Stewardship Program (CSP). Improved targeting of EQIP and CSP payments towards cover crop use may reduce the incidence of prevented plating-related crop insurance losses.

These findings also imply that it may be an opportune time to revisit and possibly adjust the prevented planting factors and guidelines in crop insurance. For example, it may be important to revisit the standard 55% (or 60%) prevented planting coverage levels in most crops, and perhaps to consider higher prevented planting coverages when cover crops are implemented. In addition, it may be time to explore how the reduction in prevented planting risk from cover crops can be quantitatively incorporated into RMA's premium rating procedures (in general, and possibly for developing prevented planting specific premiums). Because cover crops have the ability to make farms more resilient to prevented planting losses, appropriate adjustments to the crop insurance program that take this into account have the potential to lower prevented planting claims/losses, premiums, and ultimately taxpayer costs.[15]

---

[15] For example, Pandemic Cover Crop Program (PCCP) provides $5 per acre crop insurance premium discount with eligible farmers who implement cover crops in Illinois, Iowa, and Indiana. The pilot program is a positive step in

Notwithstanding the empirical contribution of the present study, it is important to acknowledge limitations and discuss promising avenues for future research. First, our study only explores the relationship between prevented planting losses and one single soil conservation practice – cover crops. To provide a better understanding on the prevented planting effects of improved soil health and soil conservation technology adoption (in general), it may be necessary to capture the simultaneous impacts of a number of other agricultural conservation practices (i.e., no-till, crop rotation, etc.) on prevented planting. Understanding the prevented planting impact of particular cover crops (e.g., legumes, brassicas, grasses), or cover crop mixtures is also an important future research direction. Second, the data used in this study was observed at the county level and only focuses on a specific region in the US. Although the geographic coverage in our study includes a major agricultural production area, individual farm-level panel data that covers a wider geographical range would likely provide more generalizable results (e.g., better external validity) and more nuanced inferences. Heterogeneity in the cover crop effect across regions and crops may also be better analyzed using farm-level data with a wider geographical scope. Lastly, our empirical analysis dealt with the issue of endogeneity due to time-county-varying unobservables by utilizing the recently developed moment-based IV model of Lewbel (2012) and the KLS regression approach. This was necessary because we were not able to find strong external IVs that would allow us to implement more traditional panel IV approaches. (e.g., 2SLS, control function approaches). If valid external instruments could be identified, it would be useful to assess whether our inferences still hold when using these traditional panel IV methods. We leave these potential extensions for future work.

terms of providing reduced premiums for cover crop users (given the evidence from this study that it reduces prevented planting risk). One policy direction that can be pursued in light of the prevented planting benefits we find in the study is making the $5 per acre cover crop premium discount a permanent part of the crop insurance program and making it available annually and for all States.

References

Aglasan, S., R.M. Rejesus, S.C. Hagen, and W. Salas. (2021). An Analysis of Crop Insurance Losses, Cover Crops, and Weather in US Crop Production. Selected paper presentation, 2021 AAEA Meetings, Austin, TX (Aug. 1-3, 2021).

Alkaisi, M., & Helmers, M. (2008). Heavy Rain, Soil Erosion and Nutrient Losses. Integrated Crop Management (ICM) Extension and Outreach Article, Iowa State University, Ames, IA.

Annan, F., & Schlenker, W. (2015). Federal crop insurance and the disincentive to adapt to extreme heat. *American Economic Review, 105*(5), 262-66.

Arbuckle, J. G., & Roesch-McNally, G. (2015). Cover crop adoption in Iowa: The role of perceived practice characteristics. *Journal of Soil and Water Conservation*, 70(6), 418-429.

Baum, C. F., & Lewbel, A. (2019). Advice on using heteroskedasticity-based identification. *The Stata Journal, 19*(4), 757-767.

Boyer, C. N., Lambert, D. M., Larson, J. A., & Tyler, D. D. (2018). Investment analysis of cover crop and no-tillage systems on Tennessee cotton. *Agronomy Journal*, 110(1), 331-338.

Boyer, C. N., & Smith, S. A. (2019). Evaluating changes to prevented planting provision on moral hazard. *Journal of Agricultural and Applied Economics, 51*(2), 315-327.

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the econometric society*, 1287-1294.

Burke, M., & Emerick, K. (2016). Adaptation to climate change: Evidence from US agriculture. *American Economic Journal: Economic Policy, 8*(3), 106-40.

Chen, B., Gramig, B. M., & Yun, S. D. (2021). Conservation tillage mitigates drought-induced soybean yield losses in the US Corn Belt. *Q Open, 1*(1), qoab007.

Chen, S. L. & Miranda, M. J. (2007). Effects of Insurance on Farmer Crop Abandonment. Selected paper presentation, 2007 AAEA Meetings, Portland, OR (July 29-Aug. 1, 2007).

Connor, L., Rejesus, R. M., & Yasar, M. (2021). Crop insurance participation and cover crop use: Evidence from Indiana county-level data. *Applied Economic Perspectives and Policy*.

Elmore, R. (2012). Influence of Soil Temperature on Corn Germination and Growth. Integrated Crop Management (ICM) Extension and Outreach Article, Iowa State University, Ames, IA.

Hagen, S. C., Delgado, G., Ingraham, P., Cooke, I., Emery, R., P Fisk, J., ... & Gustafson, D. (2020). Mapping conservation management practices and outcomes in the corn belt using the operational tillage information system (OpTIS) and the denitrification–decomposition (DNDC) model. *Land, 9*(11), 408.

Hsiang, S. (2016). Climate Econometrics. *Annual Review of Resource Economics*. 8(1): 43-75.

Hunter, M. C., Kemanian, A. R., & Mortensen, D. A. (2021). Cover crop effects on maize drought stress and yield. *Agriculture, Ecosystems & Environment*, 311, 107294.

Kahimba, F. C., Ranjan, R. S., Froese, J., Entz, M., & Nason, R. (2008). Cover crop effects on infiltration, soil temperature, and soil moisture distribution in the Canadian Prairies. *Applied engineering in agriculture, 24*(3), 321-333.

Karl, T. R., Melillo, J. M., Peterson, T. C., & Hassol, S. J. (Eds.). (2009). *Global climate change impacts in the United States*. Cambridge University Press.

Kaye, J. P., & Quemada, M. (2017). Using cover crops to mitigate and adapt to climate change. A review. *Agronomy for sustainable development, 37*(1), 1-17.

Kim, T., & Kim, M. K. (2018). Ex-post moral hazard in prevented planting. *Agricultural Economics*, *49*(6), 671-680.

Kiviet, J. F. (2013). Identification and inference in a simultaneous equation under alternative information sets and sampling schemes. *The Econometrics Journal, 16*(1), S24-S59.

Kiviet, J. F. (2020). Testing the impossible: identifying exclusion restrictions. *Journal of Econometrics, 218*(2), 294-316.

Kladivko, E. J., Kaspar, T. C., Jaynes, D. B., Malone, R. W., Singer, J., Morin, X. K., & Searchinger, T. (2014). Cover crops in the upper midwestern United States: Potential adoption and reduction of nitrate leaching in the Mississippi River Basin. *Journal of Soil and Water Conservation, 69*(4), 279-291.

Kleibergen, F., & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of econometrics, 133*(1), 97-126.

Kolář, P., Trnka, M., Brázdil, R., & Hlavinka, P. (2014). Influence of climatic factors on the low yields of spring barley and winter wheat in Southern Moravia (Czech Republic) during the 1961–2007 period. *Theoretical and applied climatology, 117*(3), 707-721.

Kripfganz, S., & Kiviet, J. F. (2021). kinkyreg: Instrument-free inference for linear regression models with endogenous regressors. *The Stata Journal, 21*(3), 772-813.

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics, 30*(1), 67-80.

Lobell, D. B., & Asner, G. P. (2003). Climate and management contributions to recent trends in US agricultural yields. Science, 299(5609), 1032-1032.

Lusk, J. L., J. Tack, and N.P. Hendricks. (2019). Heterogeneous yield impacts from adoption of genetically engineered corn and the importance of controlling for weather. In Schenkler, W. (Ed), *Agricultural Productivity and Producer Behavior* (pp. 11-40). University of Chicago Press.

Menéndez, S., Barrena, I., Setien, I., González-Murua, C., & Estavillo, J. M. (2012). Efficiency of nitrification inhibitor DMPP to reduce nitrous oxide emissions under different temperature and moisture conditions. *Soil Biology and Biochemistry*, *53*, 82-89.

Myers, R., Weber, A., Tellatin, S. (2019). Cover Crop Economics: Opportunities to Improve Your Bottom Line in Row Crops. SARE Technical Bulletin, Ag. Innovation Series. pp. 1-24.

Perry, E. D., Yu, J., & Tack, J. (2020). Using insurance data to quantify the multidimensional impacts of warming temperatures on yield risk. *Nature communications, 11*(1), 1-9.

Poeplau, C., & Don, A. (2015). Carbon sequestration in agricultural soils via cultivation of cover crops–A meta-analysis. *Agriculture, Ecosystems & Environment, 200*, 33-41.

Rejesus, R. M., Aglasan, S., Knight, L. G., Cavigelli, M. A., Dell, C. J., Lane, E. D., & Hollinger, D. Y. (2021). Economic dimensions of soil health practices that sequester carbon: Promising research directions. *Journal of Soil and Water Conservation, 76*(3), 55A-60A.

Rejesus, R. M., Coble, K. H., Miller, M. F., Boyles, R., Goodwin, B. K., & Knight, T. O. (2015). Accounting for weather probabilities in crop insurance rating. *Journal of Agricultural and Resource Economics*, 306-324.

Rejesus, R. M., Escalante, C. L., & Lovell, A. C. (2005). Share tenancy, ownership structure, and prevented planting claims in crop insurance. *American journal of agricultural economics, 87*(1), 180-193.

Rejesus, R. M., Lovell, A. C., Little, B. B., & Cross, M. H. (2003). Determinants of anomalous prevented planting claims: theory and evidence from crop insurance. *Agricultural and Resource Economics Review, 32*(2), 244-258.

Sustainable Agriculture Research and Education-Conservation Technology Information Center. (2020). *Annual Report 2019-2020: National Cover Crop Survey*. Conservation Technology Information Center Report, West Lafayette, IN.

Schlenker, W., & Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to US crop yields under climate change. *Proceedings of the National Academy of sciences, 106*(37), 15594-15598.

Schnepf, M., & Cox, C. (2006). *Environmental benefits of conservation on cropland: the status of our knowledge*. Soil and Water Conservation Society, Ankeny, IA. pp. 1-87.

Seifert, C. A., Azzari, G., & Lobell, D. B. (2018). Satellite detection of cover crops and their effects on crop yield in the Midwestern United States. *Environmental Research Letters, 13*(6), 064033.

Snapp, S. S., Swinton, S. M., Labarta, R., Mutch, D., Black, J. R., Leep, R., ... & O'neil, K. (2005). Evaluating cover crops for benefits, costs and performance within cropping system niches. *Agronomy journal, 97*(1), 322-332.

Thompson, N. M., Armstrong, S. D., Roth, R. T., Ruffatti, M. D., & Reeling, C. J. (2020). Short-run net returns to a cereal rye cover crop mix in a Midwest corn–soybean rotation. *Agronomy Journal*, 112(2), 1068-1083.

U.S. Department of Agriculture, Natural Resource Conservation Service. (2013). Cover Crops to Improve Soil Prevented Planting Fields. St. Paul, MN: USDA-NRCS. Fact Sheet. https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1142714.pdf (Accessed March 2021).

U.S. Department of Agriculture, Office of Inspector General. (2013). RMA: Controls over Prevented Planting. Washington, DC: USDA-OIG, Audit Report 05601-0001-31. https://www.usda.gov/sites/default/files/05601-0001-31.pdf (Accessed March 2021).

U.S. Department of Agriculture, Risk Management Agency. (2020). Frequently Asked Questions. https://www.rma.usda.gov/News-Room/Frequently-Asked-Questions/Prevented-Planting-Coverage (Accessed March 2021).

Volpi, I., Laville, P., Bonari, E., o di Nasso, N. N., & Bosco, S. (2017). Improving the management of mineral fertilizers for nitrous oxide mitigation: the effect of nitrogen fertilizer type, urease and nitrification inhibitors in two different textured soils. *Geoderma*, 307, 181-188.

Walthall, C. L., Hatfield, J., Backlund, P., Lengnick, L., Marshall, E., Walsh, M., ... & Ziska, L. H. (2013). *Climate change and agriculture in the United States: Effects and adaptation*. United States Department of Agriculture, Agricultural Research Services, Climate Change Program Office. USDA Technical Bulletin 1935. Washington, DC. 186 pages.

Wang, H., Beule, L., Zang, H., Pfeiffer, B., Ma, S., Karlovsky, P., & Dittert, K. (2021). The potential of ryegrass as cover crop to reduce soil N2O emissions and increase the population size of denitrifying bacteria. *European Journal of Soil Science, 72*(3), 1447-1461.

Wang, R., Rejesus, R.M. & Aglasan, S. (2021a). Warming Temperatures, Yield Risk, and Crop Insurance Participation. *European Review of Agricultural Economics. 48*(5): 1109-1131.

Wang, R., Rejesus, R.M., Tack, J.B., Aglasan, S. (2021b). Do Higher Temperatures Influence How Yields Respond to Increasing Planting Density? *Agricultural and Resource Economics Review. 50*(2): 273-295.

Wu, S., Goodwin, B. K., & Coble, K. (2020). Moral hazard and subsidized crop insurance. *Agricultural Economics, 51*(1), 131-142.

Xu, Z., Hennessy, D. A., Sardana, K., & Moschini, G. (2013). The realized yield effect of genetically engineered crops: US maize and soybean. *Crop Science, 53*(3), 735-745.

Zulauf, C., & Brown, B. (2019). Cover crops, 2017 US census of agriculture. *Farmdoc daily, 9*(135).

Figures and Tables

Table 1.1. Summary statistics of variables used in the baseline empirical models, 2005-2016 (N=7,752)

Variable	Mean	St.Dev.	Min	Max
PP-LCR (%)	1.02	2.79	0	45.34
PP-LAR (%)	1.65	3.98	0	51.72
CC (%)	2.79	5.16	0	60.30
DD10 April-May	65.47	53.61	0	380.27
DD1029 April-May	281.07	93.30	40.74	645.85
DD30+ April-May	0.00	0.00	0	0.14
PDSI_W	1.63	1.55	0	8.24
PDSI_D	0.43	0.75	0	3.99

Table 1.2. Impacts of cover crops on prevented planting: Main regression results from the linear FE and moment-based Lewbel IV models

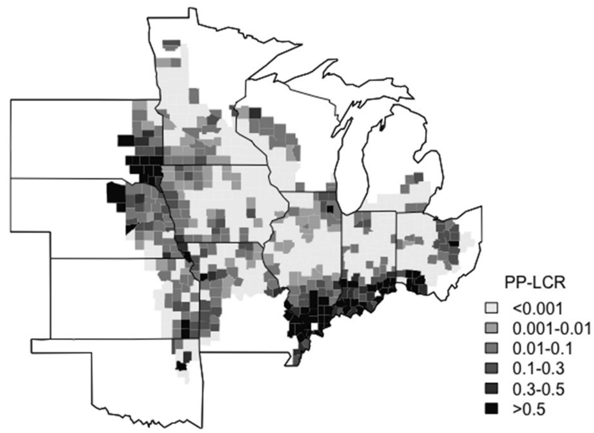
Independent Variables	Dependent Variable: PP-LCR		Dependent Variable: PP-LAR	
	Linear FE	Lewbel IV	Linear FE	Lewbel IV
CC	-0.0343***	-0.0377**	-0.0564***	-0.0480**
	(-4.39)	(-2.55)	(-4.66)	(-2.34)
April-May DD10	0.0094***	0.0094***	0.0185***	0.0186***
	(5.32)	(5.54)	(7.52)	(7.87)
April-May DD1029	0.0001	0.0001	0.0007	0.0007
	(0.12)	(0.11)	(0.39)	(0.42)
April-May DD30+	-3.3221	-3.3449	0.2602	0.3165
	(-0.33)	(-0.34)	(0.02)	(0.03)
April-May PDSI_W	0.3770***	0.3772***	0.5369***	0.5365***
	(9.12)	(9.52)	(11.09)	(11.57)
April-May PDSI_D	-0.1496***	-0.1498***	-0.2523***	-0.2518***
	(-4.23)	(-4.44)	(-5.33)	(-5.58)
No. of Obs.	7752	7752	7752	7752
Hansen J Statistic		50.969		50.678
Hansen J P-value		0.0000		0.0000
Kleibergen-Paap rk Wald F Statistic		47.785		47.785
Kleibergen-Paap rk LM Statistic		86.458		86.458
Adj R-squared	0.2649	0.0446	0.2938	0.0736

Notes: All columns include county and year fixed effects; Errors are clustered by county; t statistics are shown in parenthesis; * p<0.1, ** p<0.05, *** p<0.01.

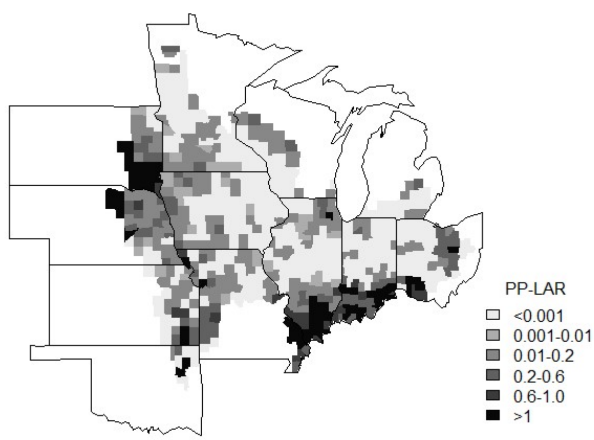
Table 1.3. Selected robustness checks: Regression results when using FSA prevented planting acres data (Panel A) and a long-differences approach (Panel B)

	---- Panel A ----		---- Panel B ----	
Independent Variables	Dependent Variable: (FSA) PP-LAR		Dependent Variable: (RMA) PP-LCR	Dependent Variable: (RMA) PP-LAR
	Linear FE	Lewbel IV	Long-differences	Long-differences
CC	-0.0576***	-0.0415*	-0.0773***	-0.1203***
	(-4.31)	(-1.84)	(-3.81)	(-3.74)
April-May DD10	0.0177***	0.0178***	0.0092***	0.0174***
	(6.98)	(7.33)	(3.59)	(5.29)
April-May DD1029	-0.0008	-0.0008	0.0071***	0.0097***
	(-0.42)	(-0.41)	(3.42)	(2.90)
April-May DD30+	-1.0935	-0.9863	3.5079	-9.0331
	(-0.09)	(-0.08)	(0.28)	(-0.60)
April-May PDSI_W	0.5351***	0.5343***	0.5241***	0.9181***
	(10.86)	(11.33)	(5.32)	(6.32)
April-May PDSI_D	-0.2970***	-0.2962***	0.1330	0.3730**
	(-5.57)	(-5.83)	(1.23)	(2.12)
No. of Obs.	7740	7740	1292	1292
Hansen J Statistic		49.485		
Hansen J P-value		0.0000		
Kleibergen-Paap rk Wald F Statistic		47.500		
Kleibergen-Paap rk LM Statistic		85.692		
Adj R-squared	0.2798	0.0519	0.1302	0.1733

Notes: All columns include county and year fixed effects; Errors are clustered by county; t statistics are shown in parenthesis; * p<0.1, ** p<0.05, *** p<0.01.



(a) Prevented-planting-related Loss Cost Ratio (PP-LCR)



(b) Prevented-planting-related Loss Acres Ratio (PP-LAR)

Figure 1.1. Maps of the PP-related measures for the study area, 2016

Chapter 2

Spatial Models for Estimating Systemic Yield Risk: Corn Yield in the Corn Belt, 1951-2019

2.1 Introduction

Federal crop insurance is one of the most prevalent farm safety net programs: In crop year 2019, more than 85% of planted acres for corn, soybeans, and wheat were insured through the federal crop insurance program (Rosch, 2021). The absence of private crop insurance is attributed to the existence of systemic and non-diversifiable risks associated with insured crop yield losses (Miranda & Glauber, 1997). The systemic risk originates from the spatial correlation in crop yields because systemic weather risk induces a high correlation among individual farm yields, disabling for insurers to pool uncorrelated risk drawn from the same distribution in crop insurance. In conjunction with asymmetric information problems, particularly adverse selection and moral hazard problems[16], systemic risk is one of the major causes of perceived crop insurance market failure (Glauber, 2004). Note that a well-known precondition of insurability is that individual risks

[16] See Glauber (2004) for a review of the early crop insurance literature that investigated asymmetric information issues. Among moral hazard-relevant literature, Goodwin et al. (2004) found the negative correlation between crop insurance participation and input usage (i.e., fertilizer and chemical usage).

are independent or at least the covariance among risks is small enough to be manageable from the insurer’s solvency perspective (Okhrin et al., 2013). The insurer’s effort to pool risks across farms could be defeated because weather conditions could cause large-scale agricultural failures. Therefore, a better understanding of spatial correlation among crop yield risks relevant to large-scale weather patterns is a key to achieving accurate farm risk management (i.e., risk-sharing arrangements through risk pool and transfer via reinsurance).

Spatial dependence refers to the unique case where the response variable or error term at each spatial location is correlated with those at other locations (Lambert et al., 2004). According to the widely accepted first law of geography developed by Waldo Tobler, *Tobler’s first law*, observations collected over regions nearer to each other tend to have similar characteristics, as compared to distant regions (Tobler, 1970; Miller, 2004). From a statistical perspective, this feature is attributed to the fact that the spatial autocorrelation between pairs of regions tends to be higher for regions that are close than for those that are remote (Kyung & Ghosh, 2009). This spatial process observed over arbitrary regions is usually modeled using autoregressive models.

Many geostatistical models can capture the spatial dependence among areal units, particularly by incorporating spatial information as discrete location information rather than as continuous variables. As a member of the family of Gaussian spatial processes, simultaneous autoregressive (SAR) models, also called spatial lag model, has been developed (Ord, 1975) and applied for the analysis of spatial data in diverse areas such as ecology (Lichstein et al., 2002; Haining, 2003; Kissling & Carl, 2008) and epidemiology studies (Clayton & Kaldor, 1987; Cressie & Chan, 1989; Hii et al., 1997; MacNab et al., 2006; Lawson, 2013). Also, SAR models can explicitly treat spatial patterns that may be present in the agricultural data while non-spatial regression techniques (e.g., ordinary least squares, OLS) cannot. There is a robust body of agricultural economic literature that has employed SAR models, especially focusing on the spatial structure of crop yields (Long, 1998; Lambert et al., 2004; Zhang et al., 2010; Seffrin et al., 2018). Long (1998) has shown the importance of spatial perspective on yield risk modeling. The paper confirms that OLS estimates without consideration of spatial dependence can underestimate standard errors and mislead R^2 using wheat yield data in Montana. Lambert et al. (2004) compare four different spatial regression methods using corn yield data in Argentina. They found several advantages of spatial autoregressive approaches that include simultaneous autoregressive (SAR) and spatial error (SEM) models. For example, SAR can work for a small number of observations than other spatial regression methods. Multiple studies compared OLS regression models and spatial regression models including the SAR model for predicting corn yield (Zhang et al., 2010; Seffrin et al., 2018). Their findings show that the spatial autoregressive model is recommended because of its significant improvement in model performance over the OLS model: (1) higher R^2 (2) lower Akaike's information criterion (AIC), Bayesian information criteria (BIC), and standard errors.

In addition to studies on spatial dependence of crop yield, multiple agricultural research use SAR models. The forecasting power of the wheat quality (measured by the level of protein content), as well as yield, can be enhanced by adding a spatial lag effect (Lee et al., 2013). Seo (2008) measures the impact of climate change on farm-level land values in South America using OLS, SAR, and SEM model specifications. The paper identifies and underscores the importance to model spatial dependence using SAR and SEM when response variables tend to be spatially correlated.

Among numerous factors affecting agricultural productivity, two components are spatially dependent: weather and agronomic characteristics (Zhou et al., 2009; Chaney et al., 2015, Ramcharan et al., 2018; Shang et al., 2019). The spatial dependence in weather is distinctively direction-dependent (Anisotropy). For example, two areas located on the same latitude are more likely to experience similar temperatures than those located in the same longitude. Along with growing concerns about climate change, many studies discover that climate change may alter the spatiotemporal patterns of weather, and in turn soil characteristics (Houghton et al., 2001; McCarthy et al., 2001; Ziegler et al., 2003). In this context, understanding and accurately estimating spatial dependence associated with crop yields can help estimate the impact of climate change on the agriculture and insurance sector.

In previous studies, spatial dependence structure is assumed to be isotropic–they are only subject to distances or adjacency between the observations. However, spatial processes are often characterized by irregular patterns including directional dependence structures. Ignoring the anisotropic dependence of crop yield across regions, and as a result, may explain the spatial dependence of crop yield partially. There have been limited attempts to accommodate the anisotropic feature of spatial dependence. In Kyung & Ghosh (2009), the directional CAR (DCAR) model was proposed to capture the directional spatial dependence. Their findings suggest that the DCAR model outperforms the regular CAR model. Based on two subsets of neighborhoods (i.e., north-south and east-west sub-neighborhoods), two-directional weight matrices are constructed. Considering that the weather pattern varies in more than two directions, DCAR may also explain partial spatial autocorrelation. Merk and Otto (2020) propose time-varying spatial weight matrices that are parameterized with respect to the wind direction and speed to model spatio-temporal spillover effects of air pollutants. But, to the best of our knowledge, there have been no studies that address four-directional spatial autocorrelation in agricultural studies where weather information is playing an increasingly important role in determining the success of agricultural production.

To overcome the limitation of previously suggested SAR models that estimate parameters of isotropic covariance function (White & Ghosh, 2009), we model the distribution of crop yield using an extension of SAR where spatial dependence of crop yield varies depending on four directions, North, South, East, and West. This extension is called a *Four-directional SAR (*FDSAR) model. The proposed model can capture the characteristic of spatial dependence of crop yield across regions varying with not only distance but also four cardinal directions. In the FDSAR model, the spatial correlations depending on distance and direction are incorporated to explain the directional spatial dependence of crop yield. We present the estimation method for the parameters based on Maximum Likelihood (ML) methods which is the most common method used to fit SAR models (Ord, 1975; Smirnov, 2005; Kyung & Ghosh, 2014). The ML methods are used to identify the optimal parametric distributions and relevant conditioning factors that can be used to quantify crop yield risks.

To test the proposed model, county-level corn yield data in the Corn Belt from the period of 1951 to 2019 is investigated. This study can contribute to more abundant literature in terms of a spatial autoregressive model in addition to helping identify a more accurate spatial correlation of crop yield. Also, the result can measure how much the dependence structure of yield risks varies as a function of two geographic natures: distance and direction. It is important to point out that the proposed model in this study can also be applied to measuring other risks as long as they are weather-relevant.

Through our statistical method development to measure crop yield risks associated with anisotropic weather characteristics, the contribution of this work is twofold. First, the heterogeneous nature of crop yield risks across farms can be better identified which offers insights into the design of actuarially fair crop insurance policies. A better understanding of yield risks can be applied to the problem confronting the USDA Risk Management Agency's (RMA) role in setting actuarially fair rates, which depends on their ability to estimate yield risks for an insured unit. USDA is required to make crop insurance available to all eligible producers, regardless of the existence of historical yield records. When the observed historical loss experience is not available, RMA could utilize the more accurate simulation-based model that accounts for the spatial correlation of yield risks based on our suggested approach.

The second main contribution is that our proposed model can be employed in many studies involving weather (or climate) risks where the directional spatial structure is carefully investigated. With our suggested model, the directional adjustments within the SAR framework allow the anisotropy parameters to capture the directional effects.

2.2 Spatial Models

The statistical models for data collected at spatial location (spatial data) can be divided into three classes: (1) geostatistical models for point-level data, (2) lattice models, also called areal models for regionally aggregated data, (3) spatial point pattern analysis to deal with the data where the locations themselves are random variables (Cressie, 1993). County-level crop yield data is characterized as lattice data that is often observed at every site at which it occurs (discrete spatial process), rather than at all points within a region such as weather (continuous process). To deal with the discrete spatial process, there are two fundamentally different ways to define the neighbor structure. One way is to treat lattice data as if it is observed on continuous indexing set by assigning the summary for the whole region to the centroid. Another way of modeling spatial structure is to embrace the discrete index nature of lattice data by defining neighbor structure in accordance with the shape of the lattice, for instance, whether they share a common boundary (Wall, 2004). In this paper, the first method is applied to investigate the spatial characteristics of crop yield.

In the statistical analysis of autocorrelated data, autoregressive models are usually employed in temporal and/or spatial statistics to account for temporal or spatial dependence among observations. In the presence of spatial dependence among dependent variables (invalid independence assumption among observations), OLS estimates may be biased, inconsistent, or inefficient (Anselin & Bera, 1998). The data property, spatial dependence, necessitates the application of the spatial statistical model. (Anselin, 1989). Spatial autocorrelation models can measure the degree of spatial dependency that a phenomenon of interest has (Cliff & Ord, 1981).

When dealing with space, the spatial correlation depending on distance is usually considered at first because the property of random variables at pairs of locations within a certain range tends to be more correlated with than out of range (Tobler, 1970). However, the deterministic distance threshold to define a neighborhood generally affect the estimated spatial dependence. For example, White and Ghosh (2009) proposed Stochastic Neighborhood CAR (SNCAR) to relax the deterministic definition of a neighborhood within a CAR model. By allowing neighborhood structure to be selected based on the unknown parameter(s), SNCAR outperforms other spatial models (traditional CAR and exponential models). The results confirm the importance of neighborhood selection.

In addition, the distance alone may not a good indicator to measure the spatial autocorrelation of crop yield. Considering the close link between weather, soil condition, and crop yields (Wheeler et al., 1996; Batts et al., 1997; Lizana & Calderini, 2013; Kawasaki et al., 2016) and spatial anisotropy in weather variation, the combination of direction and distance can be a better indicator to explain the spatial structure of crop yield.

2.2.1 SAR Models

We begin the development of our model by reviewing the conventional SAR model and comparing the conventional SAR and our proposed SAR, FDSAR. Among spatial econometric model specifications, the SAR model has been employed that explain spatial variability in crop yield risk through the inclusion of endogenous interaction term.

Let $\{X(A_i): A_i \in (A_1, \ldots, A_n)\}$ be a Gaussian random process where spatial unit set $\{A_1, \ldots, A_n\}$ forms a lattice of A. That is, spatial unit set forms a lattice of A if the spatial unit set is a simple partition of A, $A_1 \cup A_2 \cup \ldots \cup A_n = A$ and $A_i \cap A_j = \emptyset, \forall i \neq j$. SAR can model the Gaussian random process. Let's begin with the first-order autoregressive model. For each subregion, a variable of interest, $Y_i = Y(A_i)$ and set of $z < n$ explanatory variables, $X_i = X(A_i) = (X(A_{i1}), \ldots, X(A_{iz}))'$, is formulated to specify spatial autoregression:

$$Y_i = X_i'\beta + \sum_{j=1}^{n} b_{ij}(Y_j - X_j'\beta) + \epsilon_i \qquad (1)$$

Where $\varepsilon_i \sim \mathrm{N}(0, \sigma_i^2)$ is an error term, $\beta = (\beta_1, \ldots, \beta_Z)' \in \mathbb{R}^z$ is the vector of unknown regression parameters; let $\mathrm{E}(Y_i) = X_i'\beta = \mu_i$, B=$(b_{ij})_{n\times n}$, $and\ D = diag(\sigma_1^2, \ldots, \sigma_n^2)$. This model is called 'simultaneous' autoregressive (SAR) because the random disturbance terms, ε_i will be correlated with $\{X(A_j): j \neq i\}$. If *n* is finite, the joint distribution of $Y(A) = (Y_1, \ldots, Y_n)'$ is:

$$Y(A) \sim N(\mu, (I - B)^{-1} D (I - B')^{-1}) \qquad (2)$$

where μ is the mean vector, $(\mu_1, \mu_2, \ldots, \mu_n)'$ with the element μ_i representing mean at region *i* (equivalently, mean yield at county *i*), I is the n-dimensional identity matrix, $B = \rho W$ with ρ an unknown spatial parameter describing the relative strength of spatial dependence, and $W = (\omega_{ij})_{n\times n}$ a known weight (neighborhood) matrix representing the spatial structure of data. Under the given weight matrix, dependence varies depending on distances between two areal units which can be called a unidirectional weight matrix.

However, dependence can vary depending on the relative location of two areal units, direction. For example, latitude has a larger effect on climate than longitude because latitude usually controls solar energy that a specific location takes. Assume there are three cities (A, B, C) that have the same inter-distances, 400 miles. If the latitude of A and C are identical, it is more likely that those two cities have similar spatial information correlated with climate (e.g., temperature) than A and B whose latitudes are different. Note that temperature is inversely related to latitude: as latitude increases, the temperature falls, and vice versa. Also, latitude influences the distribution of precipitation.

Our suggested SAR model, four-directional SAR (FDSAR) is more desirable to identify direction-relevant areal information dependence than conventional SAR because FDSAR considers the anisotropic pattern that dependence contains. In FDSAR, n-dimensional identity matrix B is:

$$B = \rho W = \sum_{\mathrm{k}=1}^{\mathrm{k}} \rho_k W_k \tag{3}$$

where ρ is an unknown parameter describing the relative strength of spatial dependence, and $W_k = (\omega_{ij})_{n \times n}$ is a known weight matrix representing the spatial structure of data, and k is the number of the neighborhood group. If neighbor counties are divided into four groups (Northern, Southern, Western, and Eastern), k is 4. If k=1 which means dependence variation depending on direction is omitted, the model reduces to the conventional SAR (i.e., SAR model for the isotropic spatial process): B will be a single weight matrix (W) with one parameter, ρ.

2.2.2 The Choice of Weight Matrix

This section illustrates how neighborhood measure is set up to define the off-diagonal elements of the weight matrix, $W = (\omega_{ij}: i, j = 1, \ldots, n)$. W is an n x n matrix with zero diagonals ($\omega_{ii} = 0$) by convention. As long as the resulting matrix $(I - \rho W)$ is positive definite, the matrix W can be defined in various ways depending upon data characteristics and/or research purposes: weights based on the boundary, distance, or combined distance-boundary (Banerjee et al., 2014). The spatial parameter $\rho = 0$ corresponds to the 'iid' case: $\mathrm{Y_i} - X_i'\beta \sim N(0, \sigma^2)$

For simpler illustration and notation simplicity, we assume that S_i are sub-regions in a two-dimensional space, i.e., $S_i \subseteq \mathbb{R}^2, \forall i$. First, we define neighborhood structure that depends on the combination of distance and direction between centroids for any pair of sub-regions, rather than distance only or border sharing. Let $s_i = (s_{1i}, s_{2i})$ represent a coordinate of the sub-region's centroid S_i, where s_{1i} corresponds to the longitude (horizontal coordinate) and s_{2i} corresponds to the latitude (vertical coordinate).

To parameterize the weight matrix, we specify neighbors based on proximity calculated using two auxiliary data: distance and angle among county centroids. From the coordinate of centroids, geographic distance and angle are calculated to determine neighborhood structure. First, geographic distance between a pair of county centroids on a plane[17] identifies neighborhood structure. If the intra-point distance for the centroid of i and j county, $d_{ij} = d(S_i, S_j) = ||S_i - S_j||$[18] is less than the distance threshold (d_u), both become one of each other's neighbor counties. In addition, directional correlation is also mattered because spatial dependence of yield is attributed to the anisotropic pattern of weather and soil. The angle in radians between S_i and S_j is defined as the following (Figure 2.1):

$$\alpha_{ij} = \alpha(S_i, S_j) = f(x) = \begin{cases} \left| tan^{-1}\left(\frac{s_{2j} - s_{2i}}{s_{1j} - s_{1i}}\right) \right| & if \ s_{2j} \geq s_{2i} \\ -\left| \pi - tan^{-1}\left(\frac{s_{2j} - s_{2i}}{s_{1j} - s_{1i}}\right) \right| & if \ s_{2j} < s_{2i} \end{cases} \tag{4}$$

[17] Among various methods to measure the distance on a plane, distance between two points on the World Geodetic System (WGS) ellipsoid, standard for use in satellite navigation including GPS.
[18] $|| \cdot ||$ denotes the Euclidean distance in Euclidean space.

Let N_i^k represent a set of neighborhood indices for the ith region S_i based on distance threshold to determine whether they are neighbors of S_i or not. The subset of the neighborhood is clustered based on different k directions. If k=1, it becomes back to the general (omnidirectional) SAR model. In this paper, two-directional SAR and four-directional SAR models are investigated. In the case of the four-directional SAR model, the directions of neighbors are clustered into four groups: North(N_i^{N}), East (N_i^{E}), West (N_i^{W}), South(N_i^{S}) for each S_i such that $N_i = N_i^{\mathrm{N}} \cup N_i^{\mathrm{E}} \cup N_i^{\mathrm{W}} \cup N_i^{\mathrm{S}}$:

$$N_i^{\mathrm{N}} = \{j : j \in N_i, \frac{\pi}{4} < \alpha_{ij} < \frac{3\pi}{4}\}, \tag{5}$$

$$N_i^{\mathrm{S}} = \left\{j : j \in N_i, \frac{5\pi}{4} < \alpha_{ij} \leq \frac{7\pi}{4}\right\},$$

$$N_i^{\mathrm{W}} = \left\{j : j \in N_i, \frac{3\pi}{4} < \alpha_{ij} \leq \frac{5\pi}{4}\right\},$$

$$N_i^{\mathrm{E}} = \left\{j : j \in N_i, \frac{7\pi}{4} < \alpha_{ij}\ or\ \alpha_{ij} \leq \frac{\pi}{4}\right\}$$

Take Dallas County in Iowa as an example. Assume the distance threshold to define neighborhood structure is 200 miles. Then, Dallas county's neighbor counties can be grouped into four sections as Figure 2.2 illustrates: North and South neighbor county group (Light grey shaded area) and West and East neighbor county group (Dark grey shaded area). In Figure 2.2, Greene County is in the North region of S_{Dallas} because angle $\alpha(S_{Dallas}, S_{Greene})$ is in $(\frac{\pi}{4}, \frac{3\pi}{4}]$ and distance $|S_{Dallas}, S_{Greene}|$ is less than the distance threshold (200 miles). Likewise, Madison is included in the Southern neighbor, Shelby in the Western, and Jasper in the Eastern.

Based on subsets of the neighborhoods, four-directional spatial weight matrices can be constructed. Each nonnegative matrix, $W = (\omega_{ij} : i, j = 1, \ldots, n)$, is a possible spatial weight matrix summarizing spatial relations between n spatial units (counties). Each spatial weight (ω_{ij})

reflects the spatial influence of county j on county i. And zeros on the diagonal of W imply that there are no self-relationships ($\omega_{ii} = 0, \forall i$) for individual county i. Since the number of neighbor counties varies depending on a geographical characteristic such as their adjacent county sizes, a normalized weight matrix is employed to remove dependence on extraneous scale factors. Among normalizations of spatial weights, a row normalized weight matrix is utilized so that each row sums to one. Note that the ith row of Wrepresents spatial weights affecting county i. If the weights in each row are normalized to have a unit sum, it produces a row normalization of W. The normalized weight (ω_{ij}) can be interpreted as the fraction of all spatial influence of county j on county i. The specification of a weight matrix simply with 0 and 1's is not consistent in the case where the number of neighbors varies (Wall, 2004). The row standardized weight matrix corresponds with an understanding of the process that has larger uncertainty when certain areal units are farther from other units in the process (Ver Hoef, et al., 2018):

$$\sum_{j=1}^{n} \omega_{ij} = 1, i = 1, \dots, n \tag{6}$$

Since neighbor groups are divided into four-directional groups in FDSAR, four weight matrices are defined:

weight matrix of Northern (W^N), Southern (W^S), Western (W^W), and Eastern (W^E) neighborhood. Equivalently,

$$W^k = (\omega_{ij}^k : i, j = 1, \dots, n \; and \; k = N, S, W, E)$$

$$f(x) = \begin{cases} \frac{\omega_{ij}}{\sum_j \omega_{ij}}, \; if \; j \in N_i^k \\ 0, \; if \; j = i \\ 0, \; otherwise \end{cases} \quad where \; k = N, S, W, E \tag{7}$$

In the same manner, an omnidirectional weight matrix (W^{NSWE}) for the conventional SAR model (for an isotropic spatial process) and two-directional weight matrix (W^{NS}, W^{WE}) for two-directional SAR or Directional SAR (DSAR) can be generated. In the next section, we compare the performance of FDSAR with the conventional SAR and DSAR models.

As mentioned earlier, one subtlety that is often overlooked in SAR model specification is the way to define neighborhood structure (White & Ghosh, 2009). The neighborhood is commonly identified by, for example, the adjacency of areal units (Ozaki et al., 2008) or distance among areal units (Kissling & Carl, 2008). However, a certain degree of uncertainty is also inherent in the definition used to define areal units within or outside a respective neighborhood. In this analysis, we employ various thresholds that determine neighbor structure to investigate how spatial dependence varies depending on the definition of neighborhood. To account for the uncertainty associated with the definition of neighborhood structure, various combinations of distance and direction between centroids for any pair of sub-regions are utilized. In terms of distance applied, various distances are employed from 80 miles to 400 miles as a threshold to determine neighbor. Note that it almost covers from the 5th percentile (101.83 miles) to the 50th percentile (397.13 miles) of distance among 919 counties in the Corn Belt (For details, see the appendix).

2.3 Application to the US Corn yield data

2.3.1 Data Description

Agriculture production is exposed to pronounced systemic yield risks caused by climatic conditions. The existence of systemic risk has been a major roadblock to the development of private crop insurance markets (Miranda & Glauber, 1997). Therefore, modeling the spatial pattern of crop yield risks can help understand and quantify the systemic crop yield risk which mainly stems from the impact of geographically large-scale weather patterns. A better understanding of systemic risks enables insurers to pool crop loss risks across farms more efficiently so that crop insurers do not have to pass the cost of bearing the systemic risk onto farmers. Therefore, farmers can manage crop yield risk better by purchasing crop insurance. The existence of systemic yield risks among farms is our motivation to propose the four-directional SAR. Particularly county-level corn yield data[19] is employed to assess the performance of lattice FDSAR as compared to the SAR and DSAR. Our analysis focuses on corn production of 917 counties in the Corn Belt over 69 years (1951-2019). The Corn Belt[20] is a region of the Midwestern United States that has dominated corn production in the United States: As of 2020, the region accounts for 83% of total corn production in bushels (USDA-NASS).

To estimate yield risk, the "two-stage" method usually is employed: detrend time-series yield at the first stage and then estimate yield distribution based on detrended yield as "observed" data at the second stage (Miranda & Glauber, 1997; Atwood et al., 2003; Harri, Coble, Ker & Goodwin, 2011 among others). Models for deterministic trends included linear, polynomial, and spline models. In this study, a linear model is employed. (For details, see the appendix) Residuals from the deterministic trend are treated as observed yields for density estimation. Normalization is a common method to correct yields for heteroskedasticity.

We also use county-level weather information driven from the PRISM[21] data set. In particular, we collected and merged the following weather variables: Heating degree days (HDD), Growing degree days (GDD), Cooling degree days (CDD), precipitation ($Prec$), precipitation-squared ($Prec^2$) during the growing season (March to September). Since the ideal temperature range for a crop to plant and grow is 10-29ºC for corn and 10-30ºC for soybeans (Annan & Schlenker, 2015), three types of degree days are included to quantify temperature conditions

[19] The corn yield data (Corn, Grain-Yield, Measured in Busher/Acre) is obtained from the National Agricultural Statistics Service (NASS) of the United States Department of Agriculture (USDA)
[20] The Corn Belt states are the following: Iowa, Illinois, Nebraska, Minnesota, Indiana, Kansas, South Dakota, Ohio, Missouri, and Wisconsin.
[21] Parameter-elevation Relationships on Independent Slopes Model

during the growing season (Lusk et al., 2018; Tack et al., 2018). First, HDD is a measure of how cold the temperature is. One HDD is given for each degree the daily mean temperature is below 10 degrees Celsius. For example, a day with a mean temperature of 2°C has 8 HDD. In addition to HDD, extreme hot stress can be captured by cooling degree days (CDD) measuring how much cooling is required during days. CDD is measured with 30°C as a threshold. To estimate the growth and development of corn crops, GDD is employed which measures the extent of favorable temperature conditions. To adequately account for nonlinearities in the effects of moisture levels on crop yield risk, $Prec$ and $Prec^2$are employed. Figure 2.3 depicts the distribution of weather characteristics (i.e., July temperature and precipitation) and corn yield in 2019. In addition to the spatial lag term that all SAR models have, a one-year lagged yield is used to explain temporal dependence. All three SAR models (conventional SAR, DSAR, FDSAR) include spatial and temporal dependent components[22].

2.4 Performance of FDSAR model

2.4.1 Out-of-sample performance of the FDSAR model

To measure the performance of the FDSAR models as compared to conventional SAR and DSAR, county-level detrended corn yield data in the Corn Belt are investigated. From 917 counties of corn yield data (from 1951 to 2019), ten data sets are randomly selected and those data sets are roughly equal-sized and their average sample size is 91.7 (917/10). To separate train and test groups of yield, spatially separated blocks are generated such that 10-fold cross-validation can be appropriately used. (For details, see the appendix). For each analysis, one fold (test fold) is reserved as the external fold and the remaining nine folds (training fold) are used to fit a given

[22] For details, see the appendix.

model. After model calibration, the test fold is employed to evaluate the model prediction. In total, ten model runs are conducted with each fold acting as the external folder once. To retain the evaluation score, root mean squared error (RMSE) is used. If y^w denotes the vector of withheld values (test fold) and y^p denotes the corresponding predicted values, we measure the quality of estimator average RMSE given by $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y^w - y^p)^2}$ where n is the number of withheld values. A correctly specified regression model yields unbiased regression coefficients and unbiased predictions of the response variable.

Figure 2.4 illustrates the results using average RMSE as the performance measure. At smaller distances (less than 20 percentile of inter-distances) the anisotropic characteristics of yield risks are apparent: FDSAR gives better predictions than traditional SAR and DSAR. At larger distances, however, such a pattern is not shown due to the decline in spatial dependence with increasing distance which supports *Tobler's first law*. We cautiously note that if we keep increasing inter-distances as a threshold to define neighborhoods, the number of observations available within a sub-neighborhood increases. However, the spatial dependence shrinks and the number of parameters to be estimated increases. Thus, we need to restrict the distance by introducing some form of a penalty term and/or using some form of criterion to pick the inter-distance threshold.

Results in Figure 2.4 also suggest that, in general, the high prediction quality is shown when a short inter-distance threshold is applied no matter what kind of models (i.e., FDSAR, DSAR, SAR) is applied. Those results confirm that estimated spatial dependence can vary by direction and distance.

2.4.2 Results based on spatial analysis of Crop yield

With three distinctive SAR models, we estimate the spatial parameter(s) in addition to one year-lagged detrended yield and weather variables (i.e., three types of degree days and two types of precipitation) to explain the spatial and temporal dependence of detrended crop yield. We fit the SAR model with unidirectional spatial effects, the DSAR model with two-directional spatial effects, and the FDSAR model with four-directional spatial effects.

Table 2.1 reports the estimation of spatial dependence parameters. In the FDSAR, four-directional spatial dependence coefficients decrease as the inter-distance threshold to determine the neighborhood structure increases which is consistent with the first law of geography. In the DSAR, ρ^{NS} and ρ^{WE} generally follows the pattern while, ρ^{WE} increases as the distance increase at higher distances (i.e., over 340 miles). In the SAR, the relationship between unidirectional spatial dependence coefficients and inter-distance is less clear. The negative relation has been shown until 260 miles while the positive relation from 280 miles to 400 miles. Overall, directional spatial dependence coefficients (either in DSAR or FDSAR) are greater than unidirectional spatial dependence coefficients in SAR at shorter distances (below 200 miles). In DSAR models, the North-South spatial dependence parameter is stronger than East-West spatial dependence. Regardless of the neighborhood inter-distance criteria in FDSAR, the North spatial dependence parameter (ρ^N) is strongest followed by the East spatial dependence parameter (ρ^E). Those results are counter-intuitive that West-East directional weather patterns would affect crop yield and provide further research questions.

Adjusted R^2 and coefficients of explanatory variables and are reported in Table 2.2. In terms of adjusted R^2, FDSAR generally outperforms two SAR models, especially at shorter distances. The conventional SAR model captures stronger temporal dependence than directional SAR which includes DSAR and FDSAR. The smaller temporal dependence in FDSAR might be explained by the fact that directional SAR, especially FDSAR captures the largest spatial dependences in total. Therefore, the partial spatial dependence might be captured by temporal dependence. It also provides another research question. The coefficients of weather variables support the notion that favorable weather conditions during the growing season (captured by GDD and $PRCP$) enable farmers to increase their production while extreme weather conditions (capture by $HDD, CDD, PRCP^2$) disable.

2.5 Conclusion

The prosperity of private crop insurance is often impeded by the existence of systemic risk in agricultural production (Duncan & Myers, 2000) and this study focuses on the spatial and temporal dependence associated with crop yield to provide a better understanding of risks associated with crop yield. In contrast to the classical SAR model relying on an isotropic weight matrix, we investigated the maximum likelihood estimation of corn yield risk distribution in the context of (two-directional and) four-directional spatial dependence. As a consequence, the dependence structure can be explained better by four-directional spatial dependence (North, South, West, and East). Our approach is especially suited to handling risks associated with directional-spatial dependence (e.g., temperature, precipitation, wind patterns). In case of corn yield risks in 1951-2019, distinctive multi-directional spatial dependence has been found which the classic SAR model cannot capture.

Our extended SAR model, the FDSAR model allows the spatial parameters to change over direction to evaluate the directional effect of systemic risk. The existence of spatial dependence variation by direction justifies the existence of systemic risks in agricultural production and crop insurance reinsured by the federal crop insurance corporation. Our major finding, distinctive directional effect in the crop yield risk, could enable insurers to diversify crop yield risk better by targeting exposure to its directional risk factors such as weather patterns. A better understanding of these spatial systemic risks offers the potential to improve not only crop insurance (and reinsurance) but also risk management scheme relevant to directional risk factors.

These findings imply that it may be an opportune time to revisit actuarially fair crop insurance rates and redesign reinsurance contracts between spatially correlated regions. For example, individual historical yield data (at least four years) is required to calculate the normal actual production history (APH) yield. If no acceptable production records are available, USDA substitutes a county-specific transition, T-yield, for the missing periods. T-yields are based on the 10-year average yield for the entire county. Farmers without such records are assigned 65% of the T-yield as their yield. In addition to county data, neighboring county data can be used to better estimate historical crop yield risks. Appropriate adjustments to the crop insurance program that take this into account have the potential to manage production risk accurately so that farmers, insurers, and ultimately taxpayers can benefit.

Notwithstanding the empirical contribution of the present study, it is important to acknowledge limitations and discuss promising avenues for future research. First, the data used in this study was observed at the county level in the Corn Belt of the United States. Although the geographic coverage in our study includes a major corn production area that accounts for more than 85 percent of corn produced in the United States, individual farm-level panel data that covers a wider geographical range would likely provide more generalizable results (e.g., better external validity) and more nuanced inferences. Second, our study identifies the existence of directional spatial dependence only using SAR model specification. One of the potential extensions for future work is to investigate systemic crop yield risks using other spatial regression models such as spatial error model (SEM) or conditional autoregressive (CAR) to see how spatial dependence varies by direction.

References

Annan, F., & Schlenker, W. (2015). Federal crop insurance and the disincentive to adapt to extreme heat. *American Economic Review*, *105*(5), 262-66.

Atwood, J., Shaik, S., & Watts, M. (2003). Are crop yields normally distributed? A reexamination. *American Journal of Agricultural Economics*, 85(4), 888-901.

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data*. CRC press.

Batts, G. R., Morison, J. I. L., Ellis, R. H., Hadley, P., & Wheeler, T. R. (1997). Effects of CO2 and temperature on growth and yield of crops of winter wheat over four seasons. *European Journal of Agronomy*, *7*(1-3), 43-52.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, *36*(2), 192-225.

Chaney, N. W., Roundy, J. K., Herrera-Estrada, J. E., & Wood, E. F. (2015). High-resolution modeling of the spatial heterogeneity of soil moisture: Applications in network design. *Water Resources Research*, *51*(1), 619-638.

Clayton, D., & Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, 671-681.

Cressie, N., & Chan, N. H. (1989). Spatial modeling of regional variables. *Journal of the American Statistical Association*, *84*(406), 393-401.

Ghosh, S. K., & Kyung, M. (2009). Bayesian inference for directional conditionally autoregressive models. *Bayesian Analysis*, *4*(4), 675-706.

Glauber, J. W. (2004). Crop insurance reconsidered. *American Journal of Agricultural Economics*, *86*(5), 1179-1195.

Goodwin, B. K., Vandeveer, M. L., & Deal, J. L. (2004). An empirical analysis of acreage effects of participation in the federal crop insurance program. *American Journal of Agricultural Economics*, *86*(4), 1058-1077.

Haining, R. (2003). *Spatial data analysis: theory and practice*. Cambridge university press.

Harri, A., Coble, K. H., Ker, A. P., & Goodwin, B. J. (2011). Relaxing heteroscedasticity assumptions in area-yield crop insurance rating. *American Journal of Agricultural Economics*, *93*(3), 707-717.

Hii, J. L. K., Smith, T., Mai, A., Mellor, S., Lewis, D., Alexander, N., & Alpers, M. P. (1997). Spatial and temporal variation in abundance of Anopheles (Diptera: Culicidae) in a malaria endemic area in Papua New Guinea. *Journal of medical entomology*, *34*(2), 193-205.

Houghton, J. T., Ding, Y. D. J. G., Griggs, D. J., Noguer, M., van der Linden, P. J., Dai, X., ... & Johnson, C. A. (2001). *Climate change 2001: the scientific basis*. The Press Syndicate of the University of Cambridge.

Kawasaki, K., & Uchida, S. (2016). Quality Matters more than quantity: asymmetric temperature effects on crop yield and quality grade. *American Journal of Agricultural Economics*, *98*(4), 1195-1209.

Kissling, W. D., & Carl, G. (2008). Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology and Biogeography*, *17*(1), 59-71.

Lambert, D. M., Lowenberg-Deboer, J., & Bongiovanni, R. (2004). A comparison of four spatial regression models for yield monitor data: A case study from Argentina. *Precision Agriculture*, 5(6), 579-600.

Lawson, A. B. (2013). *Statistical methods in spatial epidemiology*. John Wiley & Sons.

Lee, B. H., Kenkel, P., & Brorsen, B. W. (2013). Pre-harvest forecasting of county wheat yield and wheat quality using weather information. *Agricultural and forest meteorology*, *168*, 26-35.

Lee, D., & Mitchell, R. (2012). Boundary detection in disease mapping studies. *Biostatistics, 13*(3), 415-426.

Lichstein, J. W., Simons, T. R., Shriner, S. A., & Franzreb, K. E. (2002). Spatial autocorrelation and autoregressive models in ecology. *Ecological monographs*, *72*(3), 445-463.

Lizana, X. C., & Calderini, D. F. (2013). Yield and grain quality of wheat in response to increased temperatures at key periods for grain number and grain weight determination: considerations for the climatic change scenarios of Chile. *The Journal of Agricultural Science*, *151*(2), 209.

Long, D. S. (1998). Spatial autoregression modeling of site-specific wheat yield. *Geoderma*, *85*(2-3), 181-197.

Lusk, J. L., Tack, J., & Hendricks, N. P. (2018). Heterogeneous yield impacts from adoption of genetically engineered corn and the importance of controlling for weather. *Agric. Product. Prod. Behav.*

MacNab, Y. C. (2003). Hierarchical Bayesian modeling of spatially correlated health service outcome and utilization rates. *Biometrics*, *59*(2), 305-315.

McCarthy, J. J., Canziani, O. F., Leary, N. A., Dokken, D. J., & White, K. S. (Eds.). (2001). *Climate change 2001: impacts, adaptation, and vulnerability: contribution of Working Group II to the third assessment report of the Intergovernmental Panel on Climate Change* (Vol. 2). Cambridge University Press.

Merk, M. S., & Otto, P. (2020). Estimation of Anisotropic, Time-Varying Spatial Spillovers of Fine Particulate Matter Due to Wind Direction. *Geographical Analysis*, *52*(2), 254-277.

Miller, H. J. (2004). Tobler's first law and spatial analysis. *Annals of the association of American geographers*, *94*(2), 284-289.

Miranda, M. J., & Glauber, J. W. (1997). Systemic risk, reinsurance, and the failure of crop insurance markets. *American journal of agricultural economics*, *79*(1), 206-215.

Okhrin, O., Odening, M., & Xu, W. (2013). Systemic weather risk and crop insurance: the case of China. *Journal of Risk and Insurance*, *80*(2), 351-372.

Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70(349), 120-126.

Ozaki, V. A., Ghosh, S. K., Goodwin, B. K., & Shirota, R. (2008). Spatio-temporal modeling of agricultural yield data with an application to pricing crop insurance contracts. *American journal of agricultural economics*, *90*(4), 951-961.

Ramcharan, A., Hengl, T., Nauman, T., Brungard, C., Waltman, S., Wills, S., & Thompson, J. (2018). Soil property and class maps of the conterminous United States at 100-meter spatial resolution. *Soil Science Society of America Journal*, *82*(1), 186-201.

Rosch, S. (2021). Federal Crop insurance: A Primer. *Congressional Research Service (CRS)*. https://crsreports.congress.gov/product/pdf/download/R/R46686/R46686.pdf/

Seffrin, R., Araújo, E. C. D., & Bazzi, C. L. (2018). Regression models for prediction of corn yield in the state of Paraná (Brazil) from 2012 to 2014. *Acta Scientiarum. Agronomy*, *40*.

Shang, H., Xu, M., Zhao, F., & Tijjani, S. B. (2019). Spatial and Temporal Variations in Precipitation Amount, Frequency, Intensity, and Persistence in China, 1973–2016. *Journal of Hydrometeorology*, *20*(11), 2215-2227.

Smirnov, O. A. (2005). Computation of the information matrix for models with spatial interaction on a lattice. *Journal of Computational and Graphical Statistics*, 14(4), 910-927.

Ver Hoef, J. M., Hanks, E. M., & Hooten, M. B. (2018). On the relationship between conditional (CAR) and simultaneous (SAR) autoregressive models. *Spatial statistics*, *25*, 68-85.

Ver Hoef, J. M., Peterson, E. E., Hooten, M. B., Hanks, E. M., & Fortin, M. J. (2018). Spatial autoregressive models for statistical inference from ecological data. *Ecological Monographs*, *88*(1), 36-59.

Wheeler, T. R., Batts, G. R., Ellis, R. H., Hadley, P., & Morison, J. I. L. (1996). Growth and yield of winter wheat (Triticum aestivum) crops in response to CO2 and temperature. *The Journal of Agricultural Science*, *127*(1), 37-48.

White, G., & Ghosh, S. K. (2009). A stochastic neighborhood conditional autoregressive model for spatial data. *Computational statistics & data analysis*, *53*(8), 3033-3046.

Tack, J., Coble, K., & Barnett, B. (2018). Warming temperatures will likely induce higher premium rates and government outlays for the US crop insurance program. *Agricultural economics*, 49(5), 635-647.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic geography*, *46*(sup1), 234-240.

Zhang, L., Lei, L., & Yan, D. (2010). Comparison of two regression models for predicting crop yield. In *2010 IEEE International Geoscience and Remote Sensing Symposium* (pp. 1521-1524). Ieee.

Zhou, L., Dai, A., Dai, Y., Vose, R. S., Zou, C. Z., Tian, Y., & Chen, H. (2009). Spatial dependence of diurnal temperature range trends on precipitation from 1950 to 2004. *Climate Dynamics*, *32*(2), 429-440.

Ziegler, A. D., Sheffield, J., Maurer, E. P., Nijssen, B., Wood, E. F., & Lettenmaier, D. P. (2003). Detection of intensification in global-and continental-scale hydrological cycles: Temporal scale of evaluation. *Journal of Climate*, *16*(3), 535-547.

Figures and Tables

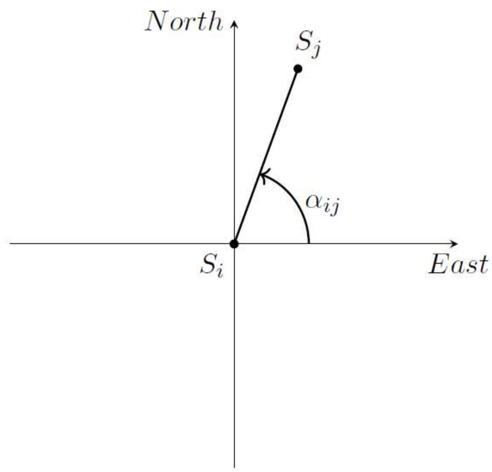


Figure 2.1. Directional neighborhood

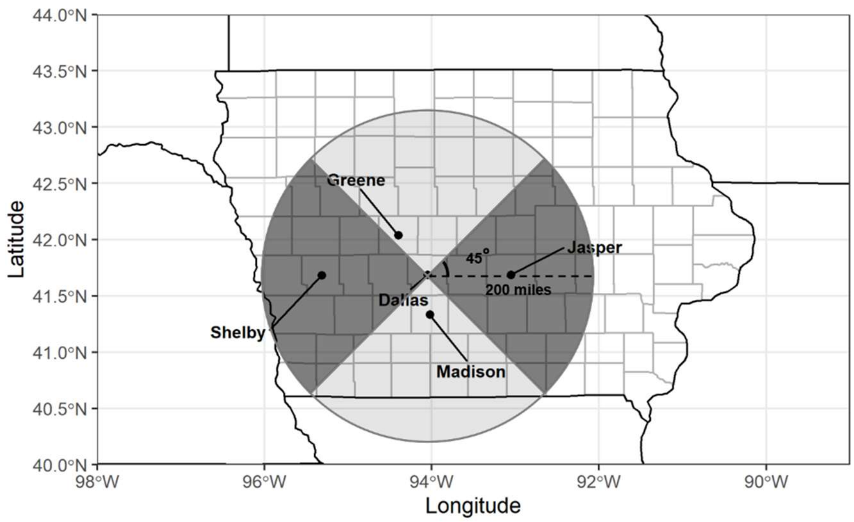


Figure 2.2. Example of neighbor structure for Dallas County, IA

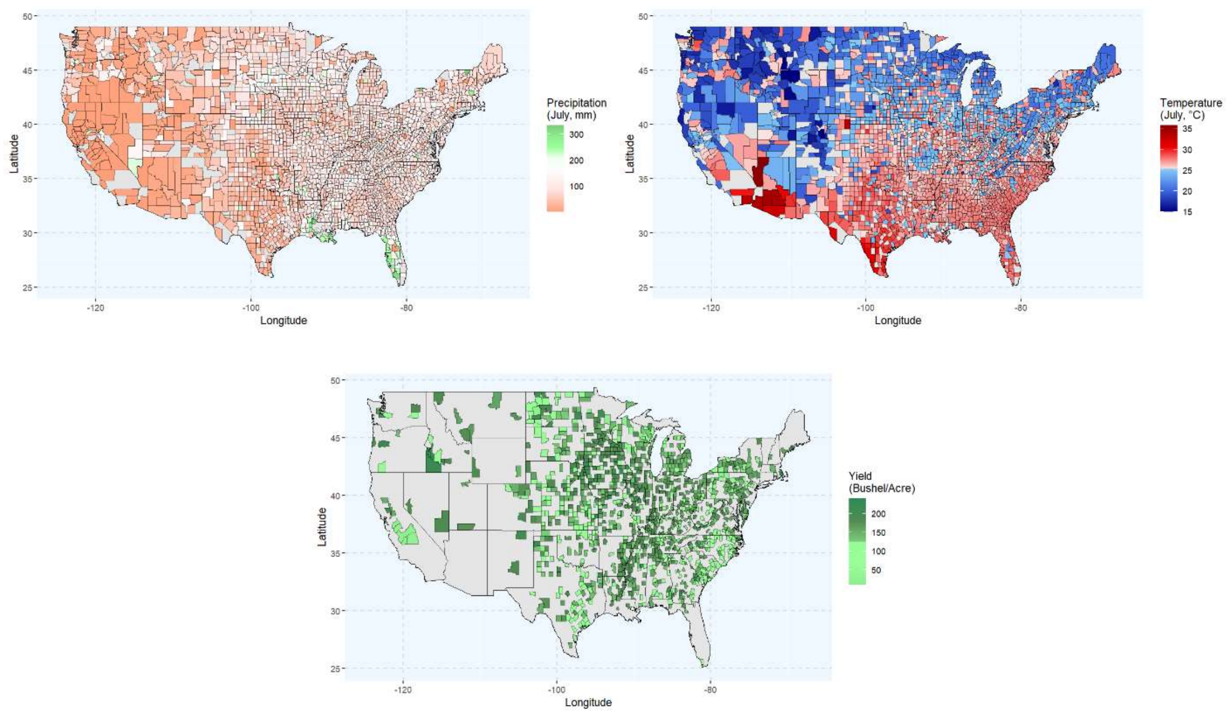


Figure 2.3. Spatial pattern: July precipitation, July temperature, Corn yield

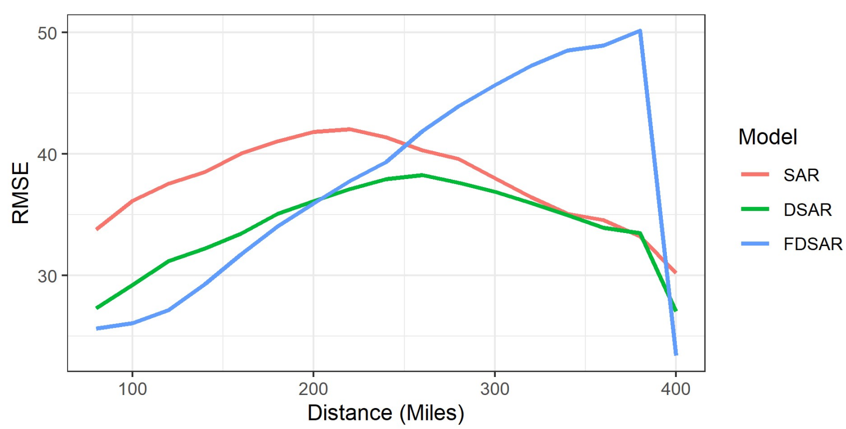


Figure 2.4. RMSE results: SAR, DSAR, FDSAR

Table 2.1. Regression results: Spatial dependence
(Dependent variable: Detrended corn yield in the Corn Belt, 1951~2019)

Model	SAR	DSAR		FDSAR			
Spatial dependence	ρ^{NEWS}	ρ^{NS}	ρ^{WE}	ρ^{N}	ρ^{S}	ρ^{W}	ρ^{E}
80 miles	0.052	0.091	0.066	0.132	0.105	0.126	0.110
100 miles	0.032	0.058	0.041	0.090	0.057	0.083	0.068
120 miles	0.021	0.038	0.027	0.058	0.042	0.052	0.042
140 miles	0.015	0.026	0.020	0.045	0.033	0.036	0.029
160 miles	0.011	0.019	0.015	0.034	0.027	0.027	0.022
180 miles	0.009	0.016	0.011	0.027	0.020	0.020	0.018
200 miles	0.008	0.014	0.008	0.022	0.015	0.014	0.014
220 miles	0.007	0.013	0.007	0.018	0.012	0.011	0.012
240 miles	0.006	0.013	0.006	0.016	0.010	0.009	0.010
260 miles	0.006	0.012	0.005	0.014	0.009	0.008	0.008
280 miles	0.007	0.012	0.005	0.013	0.007	0.007	0.006
300 miles	0.008	0.012	0.005	0.011	0.007	0.007	0.005
320 miles	0.013	0.012	0.005	0.009	0.006	0.006	0.005
340 miles	0.016	0.013	0.006	0.008	0.006	0.006	0.004
360 miles	0.021	0.012	0.008	0.008	0.005	0.005	0.004
380 miles	0.023	0.011	0.013	0.007	0.005	0.005	0.004
400 miles	0.030	0.010	0.017	0.007	0.005	0.006	0.005

Notes: **ρ^{NEWS} is the spatial dependence parameter in SAR model.**
ρ^{NS}, ρ^{WE} are the spatial dependence parameters in the North-South and West-East direction in DSAR model, respectively.
$\rho^{N}, \rho^{S}, \rho^{W}, \rho^{E}$ are the spatial dependence parameters in the North, South, West, and East direction in FDSAR model, respectively.

Table 2.2. Regression results: Coefficients and adjusted R-squared
(Dependent variable: Detrended corn yield in the Corn Belt, 1951~2019)

Distance	**SAR**							**DSAR**							**FDSAR**						
Estimate	YL1	CDD	HDD	GDD	PRCP	$PRCP^2$	Adj.R^2	YL1	CDD	HDD	GDD	PRCP	$PRCP^2$	Adj.R^2	YL1	CDD	HDD	GDD	PRCP	$PRCP^2$	Adj.R^2
80	0.15*	-0.02*	-0.38*	0.01*	0.01*	-0.00*	0.30	0.12*	-0.01*	-0.26*	0.01*	0.00*	-0.00*	0.34	0.07*	-0.00*	-0.14*	0.00*	-0.00*	0.00	0.40
100	0.17*	-0.02*	-0.45*	0.01*	0.01*	-0.00*	0.28	0.14*	-0.01*	-0.35*	0.01*	0.01*	-0.00*	0.31	0.09*	-0.01*	-0.20*	0.00*	-0.00	-0.00	0.33
120	0.17*	-0.02*	-0.5*	0.01*	0.01*	-0.00*	0.26	0.16*	-0.01*	-0.43*	0.01*	0.01*	-0.00*	0.29	0.10*	-0.01*	-0.27*	0.01*	0.00	-0.00	0.29
140	0.18*	-0.02*	-0.54*	0.01*	0.01*	-0.00*	0.26	0.17*	-0.02*	-0.48*	0.01*	0.01*	-0.00*	0.26	0.12*	-0.01*	-0.35*	0.01*	0.00	-0.00	0.28
160	0.18*	-0.02*	-0.58*	0.02*	0.01*	-0.00*	0.25	0.17*	-0.02*	-0.5*	0.01*	0.01*	-0.00*	0.26	0.13*	-0.01*	-0.43*	0.01*	0.00	-0.00	0.26
180	0.19*	-0.02*	-0.64*	0.02*	0.01*	-0.00*	0.25	0.17*	-0.02*	-0.54*	0.01*	0.01*	-0.00*	0.25	0.14*	-0.01*	-0.49*	0.01*	0.00*	-0.00*	0.24
200	0.19*	-0.02*	-0.7*	0.02*	0.01*	-0.00*	0.24	0.18*	-0.02*	-0.57*	0.01*	0.01*	-0.00*	0.26	0.15*	-0.01*	-0.53*	0.01*	0.01*	-0.00*	0.23
220	0.2*	-0.02*	-0.75*	0.02*	0.01*	-0.00*	0.24	0.18*	-0.02*	-0.58*	0.01*	0.01*	-0.00*	0.26	0.16*	-0.02*	-0.54*	0.01*	0.01*	-0.00*	0.22
240	0.2*	-0.02*	-0.79*	0.02*	0.01*	-0.00*	0.23	0.18*	-0.02*	-0.62*	0.02*	0.01*	-0.00*	0.26	0.16*	-0.02*	-0.53*	0.01*	0.01*	-0.00*	0.22
260	0.2*	-0.02*	-0.82*	0.02*	0.01*	-0.00*	0.22	0.19*	-0.02*	-0.66*	0.02*	0.01*	-0.00*	0.26	0.17*	-0.02*	-0.53*	0.02*	0.01*	-0.00*	0.23
280	0.2*	-0.02*	-0.84*	0.02*	0.01*	-0.00*	0.22	0.19*	-0.02*	-0.71*	0.02*	0.01*	-0.00*	0.26	0.17*	-0.02*	-0.52*	0.02*	0.01*	-0.00*	0.24
300	0.2*	-0.02*	-0.83*	0.02*	0.01*	-0.00*	0.23	0.19*	-0.02*	-0.75*	0.02*	0.01*	-0.00*	0.25	0.17*	-0.02*	-0.52*	0.02*	0.01*	-0.00*	0.25
320	0.2*	-0.02*	-0.8*	0.02*	0.01*	-0.00*	0.23	0.19*	-0.02*	-0.75*	0.01*	0.01*	-0.00*	0.25	0.18*	-0.02*	-0.52*	0.02*	0.01*	-0.00*	0.26
340	0.19*	-0.02*	-0.79*	0.01*	0.01*	-0.00*	0.24	0.19*	-0.02*	-0.76*	0.01*	0.01*	-0.00*	0.25	0.18*	-0.02*	-0.54*	0.02*	0.01*	-0.00*	0.26
360	0.19*	-0.02*	-0.8*	0.01*	0.01*	-0.00*	0.23	0.19*	-0.02*	-0.76*	0.01*	0.01*	-0.00*	0.25	0.18*	-0.02*	-0.57*	0.02*	0.01*	-0.00*	0.26
380	0.19*	-0.02*	-0.81*	0.01*	0.01*	-0.00*	0.24	0.19*	-0.02*	-0.76*	0.01*	0.01*	-0.00*	0.25	0.18*	-0.02*	-0.6*	0.02*	0.01*	-0.00*	0.27
400	0.19*	-0.02*	-0.8*	0.01*	0.01*	-0.00*	0.30	0.18*	-0.02*	-0.76*	0.01*	0.01*	-0.00*	0.25	0.19*	-0.02*	-0.62*	0.02*	0.01*	-0.00*	0.27

Notes: **$YL1$ is the coefficient of one-year lagged yield. Adj.R^2 represents adjusted R-squared; Asterisks indicate statistical significance at the α=0.10 or smaller level.**

Chapter 3

Exploring a New Dimension of Agri-food Sector Safety: The Role of Agri-food Production and Processing on Foodborne Illnesses

3.1 Introduction

The Centers for Disease Control and Prevention (CDC) estimates that there are approximately 48 million episodes of foodborne illness, 128,000 hospitalizations, and 3,000 deaths in the United States (Scallan et al., 2011) every year. The incidence of foodborne illnesses leads to a significant financial burden (Scharff, 2011; Hoffmann et al., 2012). Anyone can be susceptible to foodborne pathogens, but specific groups of people are more likely to get sick easily and experience more severe symptoms. People at risk include children, older adults, pregnant women, and other people with weakened immune systems (CDC, 2019; Lund and O'Brien, 2011; Simon et al., 2015). While healthy people usually experience mild symptoms and recover well from foodborne illness, the consequences in those vulnerable populations tend to be more serious and such individuals may become ill from infection caused by lower numbers of a pathogen (Farber et al., 1996; Lund, 2016). The susceptibility to foodborne pathogens in the elderly (those age 65 and older) is largely due to their decreased immune function (Smith, 2017). They also may experience underlying medical conditions (e.g, chronic diseases) and/or undergo major surgeries, which can exacerbate susceptibility to foodborne illnesses and trigger more severe complications from these (Smith, 1998; Buzby, 2002; Montecino-Rodriguez, 2013).

From the occupational perspective, an unsafe work environment can lead to immune system deficiencies (Rein, 1992; Schenker et al., 1998; Ye et al., 2013). Farmworkers are often exposed to hazardous chemicals such as pesticides, disinfectants, and air pollutants (CDC, 2018), and their exposure level varies depending on the work environment. For example, farmworkers in large animal confinement buildings are more likely to be exposed to a higher level of those hazardous substances, including chemicals and air pollutants (Gold et al., 2007; Basinas et al., 2015; Parks et al., 2016). As one of the most heavily regulated industries, the food processing sector is subject to more stringent hygiene and health regulations by federal and state regulatory agencies than is the case for the food production sector.

In addition to the occupational health impact of agricultural production, the health conditions of rural residents can be affected by agricultural activity through the proximity of their residences to agricultural production. Gunier et al. (2011) find that agricultural pesticide use near residences is a significant determinant of concentrations of pesticides in household dust. Farmworkers, particularly migrant farmworkers, may transport pesticides into their home as residues on clothing or shoes (Bradman et al., 2007; Freeman et al., 2004; McCauley et al., 2001). Exposure to agricultural chemicals often leads to a variety of immediate or long-term health problems. Ward et al. (2006) find that there is a significantly positive correlation between the proximity to crop fields and detections and concentration rates for agricultural herbicides in Iowa. Their result indicates that the odds of detecting agriculture herbicides increase as the acreage of nearby crop fields increases. Von Ehrenstein et al. (2019) found a positive correlation between residential health problems and agricultural chemical exposure. Prenatal exposure to ambient pesticides from the nearby agricultural region during pregnancy could increase the possibility of a baby having a compromised immune system. Caballero et al. (2018) studied the association between residential exposure to agricultural chemicals and premature mortality from Parkinson's disease. They found that individuals exposed to land-use associated with herbicide in Washington State had 33% higher odds of premature mortality than those that were not exposed.

Over the past few decades, the United States Department of Agriculture (USDA) and the Food and Drug Administration (FDA) in addition to numerous state and local agencies have implemented food safety regulations and enforcement mechanisms to prevent foodborne illness. However, numerous foodborne disease outbreaks continue to occur every year in the United States. In light of a large number of such illnesses, several studies have investigated the economic burden of foodborne illnesses. Based on a 2012 estimation of the incidence of foodborne illness from the CDC, Hoffmann et al. (2012) and Scharff (2012) estimate the aggregated annual cost of foodborne illness in the United States to be as high as $14.1 billion to $16.3 billion[23], respectively. The cost of illness includes treatment costs, the value of time lost due to illness, and the willingness to pay to prevent death (Hoffman and Anekwe, 2013). In addition to these health impacts, these cases present significant financial, brand, reputation, and other risks to the private sector businesses who are, or who are suspected to be, at fault in providing this unsafe food.

---

[23] The aggregated annual cost difference is originated from the different number of pathogens and valuation methods that two studies have applied.

The broad objective of this study is to investigate the linkages between agricultural production and the incidence and nature of foodborne illnesses. Specifically, this paper addresses the question of how the characteristics of agricultural production affect the risks associated with foodborne illnesses. To identify contributing factors to foodborne illness outbreaks, we construct a unique county-level panel data set that includes rich information on foodborne illness outbreaks, characteristics of agricultural production, and health factors relevant to the length and quality of life (i.e., health behaviors, accessibility to clinical care, and economic factors). We also consider the relevance of several demographic characteristics (age and race).

Based on the unpublished CDC's National Outbreak Reporting System (NORS) data, we measure four dimensions of risk associated with food-borne illness—the occurrence of foodborne illnesses, doctor visits, emergency room visits, and hospitalizations. We intend to capture a different level of severity of foodborne illness. Our main independent variables to characterize agriculture production are derived from the U.S. Bureau of Economic Analysis (BEA) farm income and expense data. The county-level regional data are disaggregated into different income and expense categories: cash receipts from livestock and crops, production costs (e.g., feed, seed, fertilizer, labor), and farm income. By capturing the characteristics of the input costs pattern (e.g. proportion of farm income, livestock sales, and labor costs), we indirectly measure (i) the relative size of the livestock sector, (ii) the size of the processing sector in the food industry, and (iii) the size of the agriculture sector. The panel data cover 50 states and Washington D.E. in the US from 2009 through 2018.

Maximum-likelihood methods are employed to identify the optimal parametric distributions and relevant conditioning factors that can be used to quantify risk. To determine county-level factors associated with foodborne illness outbreaks, we estimate Cragg's double-hurdle (DH) model (Cragg, 1971) model. Given the fact that the number of reported foodborne illness outbreaks is frequently zero, the double-hurdle specifications address the preponderance of zeros problem. A variety of robustness checks using different measures of the risks associated with foodborne illness outbreaks are also conducted to validate the core result.

A robust body of literature has dealt with factors affecting the occurrence of foodborne illnesses. Many studies investigated various agents that contaminate food: micro-pathogens (bacteria, viruses, and parasites), chemicals, and toxins (Scallan et al.,2011; Geissler et al., 2017; Marsh et al., 2018 among others). Scallan et al. (2011) revealed the main etiology is norovirus, which accounts for 58% of foodborne illnesses while nontyphoidal salmonella spp. is the leading cause of foodborne-related severe outcomes (hospitalizations and deaths), 35% and 28% respectively.

The transmission method and food vehicle is another area of research on foodborne illnesses (Harvey et al., 2016; Barrett et al., 2017; Chai et al., 2017; Whitham et al., 2021). A number of studies have documented how fresh or minimally processed vegetables can play an important role in food-borne illness outbreaks (Francis et al., 1999; Sivapalasingam et al., 2004; Lynch et al., 2009). Particularly, fresh produce including fruits and vegetables are identified as novel foodborne outbreak-associated food vehicles which have not been implicated in past outbreaks (Whitham et al., 2021). Unlike foods of animal origin (e.g., meat, poultry, and dairy products), fresh produce is typically consumed raw without cooking which is an important step to kill off most harmful bacteria.

The setting of exposure is also extensively investigated (Gould et al., 2013; Angelo et al., 2017; Marlow et al., 2017). Bellemare and Nguyen (2018) studied the positive relationship between the number of farmers' markets and the number of foodborne illness outbreaks. Marlow et al. (2017) investigated the epidemiology of foodborne outbreaks in correctional institutions. They found that the annual median number of outbreak-relevant illnesses per 100,000 population in correctional institutions is higher than the number in other places, which represents a disproportionate number of outbreak-associated foodborne illnesses.

While there are already a large number of studies investigating foodborne, no study has illustrated how agriculture production-induced immune system conditions are related to the risks associated with foodborne illnesses. The main contribution of this paper is to provide empirical evidence on the correlation between agriculture production and the magnitude of risks associated with foodborne illness outbreaks at the county level. To the best of our knowledge, this is the first study to focus on the influence of host factors (e.g., immunodeficiency) affected by agricultural production on the risks of foodborne illness and immune status in an agricultural setting. Since the proportion of disease transmitted by food differs by host factor, especially the immune system (Scallan et al., 2011), this distinction can better quantify the risks associated with food-related illness. Another contribution of our analysis is the use of extensive data that allows us to conduct an empirical analysis over wider geographical regions (50 states and Washington D.C.), and over a longer time period (2009~2018). In addition, a better understanding of the factors associated with widespread food-related illnesses is critical to the development of effective management and risk mitigation mechanisms. Considering the growing interest in food liability insurance, the findings of this study can also provide a better understanding of the factors associated with disease outbreaks to better inform both the insurer and the insured. Finally, the results offer insights into how to best allocate public healthcare resources efficiently by estimating the frequency of foodborne outbreaks based on the characteristics of residents.

Our empirical results reveal a statistically significant relationship between foodborne illness risks and levels of various agricultural activities (i.e., size of livestock farming, size of the food processing sector, size of agriculture sector). The likelihood of foodborne illness outbreaks increases as the scale of livestock farming increases, as the size of the food processing sector decreases, and as the size of the overall agricultural sector increases. The results of this analysis have implications for assessing food safety risks. We conclude with a discussion of public policy issues associated with food-borne disease outbreaks.

3.2 Background

This study's conceptual framework focuses on various metrics of the intensity and scale of different agricultural work environments as reflected in input costs and sales patterns. A large body of immunology studies confirms that agricultural work environments with health hazards (e.g., intensive livestock farming and producer sector with a low hygiene standard) can weaken the immune system that farmworkers and neighboring residents have. Therefore, an accurate indicator of the work environment based on agricultural input and sales patterns is key to measuring the immune system characteristics of individuals in communities (defined as counties) and, as a result, the likelihood of foodborne illnesses. In this analysis, we integrate the findings of immunological studies with insights into the economic analysis (risks associated with foodborne illnesses) so that we can properly identify the linkages among agricultural production, the host factor (i.e., the status of the immune system), and in turn the likelihood and incidence of foodborne illnesses. Figure 3.1 visually presents our research approach.

3.2.1 Livestock Farming

Agriculture ranks among the most hazardous industries. According to the Bureau of Labor Statistics (BLS) Survey of Occupational Injuries and Illnesses, an estimated 1,900 farmworkers experienced an illness in 2015 (31.8 per 10,000 full-time workers), which is higher than other hazardous occupations (Chari et al., 2018). Safety and health issues related to agricultural operations are well documented (Brackbill et al., 1994; Beaumont et al., 1995; Von Essen et al., 1998; Frank et al., 2004; Schenker et al., 2009; Nordgren and Bailey, 2016 among others). Particularly, livestock farmworkers are exposed to numerous health, biological, and respiratory hazards. The trend of increasing concentration in the confined livestock industry with fewer and larger operations generates multiple pressing public health issues associated with concentrated animal feeding operations (CAFO) including manure management that contains nutrients, pathogens, growth hormones, antibiotics, and chemicals (Burkholder et al., 2007; Hribar, 2010).

Livestock farming produces a variety of dust, vapors, and fumes (e.g., organic dust from livestock barns and confinements, pesticides) such that livestock and poultry production workers have an increased risk for respiratory diseases (Beaumont et al., 1995; McClendon et al., 2015; Nordgren and Bailey, 2016). For example, organic dust–an aggregate of air-suspended particles sourced from plants and animals–is a major air pollutant within intensive livestock farming workplaces (Basinas et al., 2015). In concentrated animal feeding operations, livestock workers are exposed to organic dust that contains bacterial products in addition to toxic gases such as ammonia (Linaker and Smedley, 2002; Nordgren and Bailey, 2016). Repetitive organic dust exposures can cause various chronic respiratory diseases and reductions in lung function, especially in large animal farming environments (Donham et al., 1995; Von Essen et al., 1998; May et al., 2012; Poole and Romberger, 2012). Bioaerosols inhaled by livestock farmers are rich in hazardous substances including endotoxins, which have been associated with both acute and chronic illnesses. Donham et al. (1995) discovered a strong correlation between dust exposure in swine confinement facilities and reduced lung function of swine producers.

Intensive animal husbandry also leads farm employees to be exposed to many types of antibiotics, agricultural chemicals (such as pesticides), and animal diseases transmittable to humans. Although the occurrence of antibiotic exposure is possibly natural in some cases, intensive animal farming is a primary contributor to the increased environmental burden of antibiotic-resistant genes (Neyra et al., 2012; Hille et al., 2017; Li et al., 2019; Ma et al., 2019). Neyra et al. (2012) discovered that exposure to pathogens and antimicrobial-resistant bacteria could increase the disease risk for workers and, indirectly, for their families and communities.

Farming is one of the few industries in which household members often provide labor and live on the premises which may expose them to health, biological, and respiratory hazards (CDC National Institute for Occupational Safety and Health (NIOSH)). Studies have shown that family members can serve as reservoirs for antibiotic-resistant bacteria and endotoxins and that transmission can occur between family members including young children (Thorne et al., 2005; Huijsdens et al., 2006; Rafee et al., 2012). Merchant et al. (2005) found a high prevalence of asthma among children living on farms that raise swine.

The presence of livestock farms nearby residential areas can also influence the health of residents. Several studies found that intensive livestock farm emissions such as high levels of organic dust may affect not only the respiratory health of farmers but also the health of neighboring residents (Schulze et al., 2006; Borlée et al., 2017; van Dijk et al., 2017). Borlée et al. (2017) found that lung disease patients living near farms had increased respiratory symptoms. The findings of these studies are consistent with the argument that continuous exposure to concentrated animal feeding operations can dampen innate immune responses (Sahlander et al., 2012).

Also, the prevalence of antibiotic resistance in animal farms and the surrounding environment has been extensively reported. (Alam and Zurek, 2004; Schmid et al., 2013; Hille et al., 2017; Markland et al., 2019). Hille et al. (2017) found that less intensive livestock farming combined with better hygiene in beef cattle lowers the risks associated with antibiotic resistance including antimicrobial resistance in one of the major foodborne pathogens in the United States, *E. coli*. Resistance to antibiotics could weaken the immune system (Bhaskaran et al., 2018) and make treatments less effective. Consequently, intensive livestock farming could increase the odds that neighboring residents (in addition to agricultural workers) are exposed to animal diseases transmittable to humans, which thereby adversely impacts the community's residents' immune systems.

Environmental exposure from concentrated animal feeding operations also affects residential health through two pathways: emissions into water and air. Air emissions from livestock farming including endotoxins, particulate substances, organic dust, and ammonia could hurt residents' health (Mitloehner and Schenker, 2007). There is a large body of literature that has addressed community health issues, particularly in regard to air emissions from livestock farming[25] (Elliott et al., 2004; Mirabelli et al., 2006; Radon et al., 2007; Schinasi et al., 2011; Borlée et al., 2015; Douglas et al., 2018, among others). For example, De Rooij et al. (2019) concludes that exposure to particulate matter in livestock farm emissions is associated with respiratory health effects in neighboring residents. In addition to exposure to airborne emissions, pollution from livestock farms through animal wastes can harm both surface and groundwater, which could lead

[25] Since 1969, air emissions from confined animal feeding operations have been studied extensively. There are multiple studies that provide a review of literature on the community health concerns of livestock farming. Donham, K. J. (2010) conducted a literature review with a focus on pork production. Also, Casey et al., (2015) reviewed the relevant literature since 2000 and identified four health outcomes (respiratory disorders, antibiotic resistance, zoonotic disease (infection disease that is transmitted from animal to humans), stress). Rooij et al. (2019) present an extensive literature review on the respiratory health problems of residents in the vicinity of livestock farming.

to public health problems (Soupir et al., 2006; Koike et al., 2007; Sapkota et al., 2007; Fahrenfeld et al., 2014; Marti et al., 2014). Contaminants from animal wastes can enter the environment through leakage from manure lagoons or atmospheric deposition (Burkholder et al., 2007). Also, bacterial pathogens carried by flies and rodents are one vector through which livestock farming can lead to adverse health effects in nearby communities (Graham et al., 2009; Ahmad et al., 2011; Van de Giessen et al., 2009). To measure the proportion of livestock farming in a county, we utilize the proportion of livestock sales to total agricultural sales (livestock sales /total agriculture sales).

3.2.2 Food Process Sector

As a diverse industry that includes multiple occupational and environmental exposures and widely varying labor efforts (Kirkhorn and Garry, 2000), the food industry consists of four sectors: the farm service sector; the producer sector; the processor sector; and the marketer sector. From an epidemiological and industrial hygiene perspective, the processing sector is more likely to maintain the hygiene level of their workplace than is the case for the primary producer.

Hazard analysis and critical control point (HACCP) is a management system in which food safety is addressed by monitoring the total food system, from harvesting to consumption, to reduce the incidence of foodborne illness (Ropkins and Beck, 2000). As a proactive alternative to end-point testing, which does not effectively ensure food safety, federal regulations covering the meat, poultry, fruit juice, and seafood processing industries require the implementation of HACCP in the United States. Every processing plant must establish and carry out its own HACCP plan to address all hazards associated with its processed products and the effectiveness of the plan will be continuously verified and inspected by the USDA Food Safety Inspection Service (FSIS) (USDA-FSIS, 1996). To effectively identify, evaluate, and control food safety hazards, HACCP has the seven principles: (1) conduct a hazard analysis, (2) determine the critical control points (CCP), (3) establish critical limits, (4) establish monitoring procedures, (5) establish corrective actions, (6) establish verification procedures, and (7) establish record-keeping and documentation procedures. CCP is a step at which biological/chemical/physical control can be applied and is essential to eliminate or reduce a food safety hazard to an acceptable level (FDA, 2017).

Another centerpiece of HACCP is the sanitation standard operating procedures (SSOPs). Every plant must adopt and implement its own plan to meet its daily sanitation responsibilities such that adulteration of the product will not occur. While SSOPs vary depending on the specific circumstances of products, production processes, and facilities of the establishment, SSOPs have two main parts: pre-operational (e..g., cleaning equipment and facilities) and operational sanitation (e.g., cleaning and sanitizing employee's hands and equipment). Putting into place more stringent hygiene standards in the workplace can the health risks of workers in the processing sector due to the high level of occupational hygiene in their workplace.

While the FDA and USDA require mandatory HACCP for meat, poultry, seafood, and fruit juice, the HACCP is not yet incorporated into the agricultural production sector. While HACCP is being suggested for the agricultural production sector, the implementation of HACCP on the farm has challenges. For instance, on-farm implementation requires that scientifically documented steps and preventive measures exist that can be applied at known critical control points. However, potential biological hazards on the farm typically do not have enough critical control points. If HACCP outlaws the presence of bacteria, viruses, and parasites on the farm, this rule will be hard to enforce (Cullor, 1997). Several factors create this difficulty (de W Blackburn and McClure, 2009): (1) pathogens can never be entirely eliminated from a farm; and (2) detecting pathogens is often difficult because even perfectly healthy animals may be disease carriers.

In 1998, the FDA issued guidance to outline Good Agricultural Practices (GAP) to address microbial food safety hazards which can be implemented on farms in the production of fresh produce (FDA, USDA, and CDC, 1998). However, the lack of incentives and/or regulations (GAP is strictly voluntary) may limit the acceptance of management practices intended to minimize foodborne disease (Diez-Gonzalez and Mukherjee, 2009).

In the agriculture industry, the producer sector and processor sector have distinctive levels of labor intensity. Nolte and Ostermeier (2017) found that the processor sector is labor-intensive compared to the producer sector using wage data. Morris et al. (2009) also confirmed that commercial agriculture can generate a large number of jobs in off-farm operations, such as processing and packaging. We measure labor intensity through the labor cost share (farm labor expenses/total production costs).

3.2.3 Agriculture Sector

As alluded to above, agriculture ranks among the most hazardous industries. According to BLS data, the agriculture sector (including agriculture, forestry, fishing, and hunting) continues to have the highest incidence of nonfatal occupational injuries and illness, 5.3 per 100,000 full-time workers in 2018. Particularly, farmers suffer from exposure to toxic chemicals. Approximately 85% of herbicide usage, 36% of insecticides, and 84% of fungicides in the United States occur in agriculture (Atwood and Paisley-Jones, 2017). As such, agricultural workers are at greater risk of chemical exposure than non-agricultural workers. Calvert et al. (2008) found that the incidence rate for acute pesticide poisoning is 53.6 per 100,000 among agricultural workers while only 1.4 per 100,000 among non-agricultural workers. Therefore, we hypothesize that heavy exposure to hazardous chemicals leads to a higher probability of foodborne outbreaks among farmworkers due to a potentially compromised immune response. To account for the hypothesis that agriculture workers are less likely to maintain good health conditions, we employed agriculture's share of the overall county economy (farm income/total personal income).

3.3 Data Description

The county-level panel data set constructed for this analysis cover the period 2009 to 2018 and are based on information collected from different sources. The main dependent variables are metrics of the risks associated with food-related illnesses based on the CDC National Outbreak Reporting System dataset (NORS) database. Since 1971 CDC has collected various types of reported outbreaks. (e.g., Foodborne, Waterborne, Animal Contact, Environmental, Person to Person) According to CDC, a foodborne outbreak is defined as, "an incident in which two or more persons experience a similar illness resulting from the ingestion of a common food." In the CDC outbreak data, we measure foodborne-relevant risks in four ways: (1) the total number of foodborne primary cases, including lab-confirmed and probable cases, based on the outbreak-specific definition (*Illnesses*), (2) the number of foodborne primary cases who visited a health care provider (*Doctorvisits*), (3) the number of foodborne primary cases who visited an emergency room (*ERvisits*), and (4) the number of foodborne primary cases who were hospitalized (*Hospitalizations*). In terms of the primary mode of transmission, this analysis focuses on foodborne outbreaks which have "Food" as a primary mode of transmission. And this analysis includes all types of etiology including unknown etiology.

From the CDCNORS dataset, our analysis utilizes 5,012 reported foodborne outbreaks in the United States. from 2009 through 2018. Outbreaks can be divided into two groups geographically: single-county outbreaks and multicounty outbreaks. In single-county outbreaks, exposures to the source of the outbreak (e.g., food, water, animal) occur in a single county. However, exposures sometimes occur in multiple counties. We omit multicounty outbreaks, which account for 3.95% of total valid samples (206 of 5,218 foodborne outbreaks).[27] We then aggregate single-county outbreak cases by year and county. The main dependent variables used in this study are "normalized" by the population size. Rather than counts of incidences (e.g., the number of illnesses), their ratio (e.g., the number of illnesses per 100 thousand population) is more relevant to the model because the ratio accounts for the scale effect. Although the CDC NORS data are comprehensive, there are some limitations. Only a small proportion of foodborne illnesses reported to the CDC are associated with outbreaks: Many foodborne illnesses that are reported to CDC are not confirmed by laboratory testing. Also, Outbreaks are likely underreported because reporting of outbreaks to the CDC is voluntary. For example, in CDC NORS, only 6% of Salmonella illnesses were part of a recognized outbreak in 2006 (Gould, et al., 2013).

To investigate how the characteristics of agricultural operations affect risks associated with foodborne-related illnesses, the county-level BEA personal income data are employed. Specifically, the following input costs and sales are used: livestock sales /total farm sales (%, *gLivestock*), hired farm labor costs/total farm costs (%, *gProcessing*), and farm income/total income (%, *gFarmincome*) (Table 1). In measuring the risks associated with agricultural production, one must determine the level at which impacts will be measured. In our case, we consider "neighbor" effects intended to capture the increased risk due to nearby agriculture production. The neighborhood is commonly identified by, for example, the adjacency of areal units (Ozaki et al., 2008), the distance among areal units (Kissling and Carl, 2008), or a fixed spaced

[27] This omission of about 4% of the data naturally raises concerns of selection bias. The small proportion of data omitted minimizes this concern and approaches to including multi-county outbreaks remains a topic for future research. Standard approaches to dealing with selection bias are complicated by the fact that the relevant counties are not identified when the outbreaks are multi-county.

grid (Goodwin and Piggott, 2009). In this analysis, we employ geographical adjacency to quantify the neighbor effect.[28]

Regarding control variables in our analysis, several measurements are utilized: the number of primary care physicians, measures of food insecurity, income, unemployment, demographic characteristics (i.e., age structure, ethnicity), and fresh food expenditures. Ensuring access to quality health care is essential for preventive and primary care to maintain health (Baicker and Chandra, 2004). Having a primary care provider is associated with a higher likelihood of appropriate care which should result in better health outcomes (Clancy et al., 2012). To measure the accessibility of primary care physicians, the ratio of the number of primary care physicians relative to the total county population (*Physician*) is employed.

According to the USDA, food insecurity is defined as "the disruption of food intake or eating patterns because of lack of money and other resources" (Nord et al., 2005). A lack of consistent access to enough food is related to negative health outcomes such as obesity and premature mortality (Dhurandhar, 2016). We also collected a food insecurity index drawn from "Feeding America," a United States-based nonprofit organization. Feeding America first published the "Map the Meal Gap" project in 2011 to learn more about the face of hunger (i.e., food insecurity) at the local level (County Health Rankings, n.d.). The food insecurity index is measured by those members of the population that lack access to enough food over the total county population. This measure also addresses the ability of individuals and families to provide balanced meals, including fruits and vegetables, further addressing barriers to healthy eating.

[28] In particular, explanatory factors are measured using the average value of the variable for the county of interest and all contiguous counties. This is intended to capture the fact that individuals in a county may be closer to the centroid of an adjacent county that to the county in which they reside.

In addition to direct health factors (i.e., clinical care and health behaviors), social and economic factors such as income and employment can influence health outcomes. An extensive literature, summarized by Mode, Evans, and Zonderman (2016), has investigated the influence of economic status on health and mortality. In the United States, more than 1 in 10 people still live in poverty (Semega et al., 2020) and a steady income allows people to maintain their health needs. Unemployed individuals are 54% more likely to be in poor or fair health than individuals who are employed, and are more likely to suffer from increased stress, high blood pressure, heart disease, and depression (Ross and Mirowsky, 1995; Braveman et al., 2011). There is a large number of studies showing that unemployment could lead to an increase in unhealthy behaviors (e.g., alcohol, tobacco consumption, diet, exercise, and other health-related behaviors), which in turn may lead to a substantial increase in risk for disease or mortality (Dooley et al., 1996; Wanberg et al., 2002; McKee-Ryan et al., 2005). Low-wage workers are less likely to have health insurance and access to preventive health care and are more likely to be exposed to hazardous work environments (Collins et al., 2003). In contrast, high-wage earners with steady employment are more likely to be healthy because employment provides stable income and benefits that can help employees and their families meet health care needs (e.g., health insurance, paid sick leave). We employ two variables to address the impact of economic status on health conditions: median household income to measure neighborhood economic status (*Income*); and the unemployment rate (*Unemployed*). Note that the median household income is a well-known indicator of income and poverty that can often compromise physical and mental health (Galea et al., 2011). By including the median income and unemployment rate as control variables, we account for economic and social determinants of health and quality of life.

To account for demographic characteristics, age structure and ethnic composition, that is populations of Hispanic (%, $Hispanic$) and non-Hispanic black(%, $Black$) are utilized drawn from Census, leaving “non-black” as the default ethnic category. As mentioned above, the elderly do not respond to the immune challenges as strongly as the young (Montecino-Rodriguez, 2013). To account for the age structure of a population, the population of age 65 and older (%, $Over65$) is used. Food consumption patterns are often influenced by demographic characteristics (Gilbert and Khokhar, 2008; Franco et al., 2009).

The consumption of fresh produce continues to rise in recent years due to the perceived potential health benefits of fresh produce in the United States (Pollack, 2001; Sela and Manulis-Sasson, 2015). According to the USDA, per capita consumption of fresh vegetables in the United States grew vigorously from 154 pounds in 1970 to 202 pounds in 2017, while per capita consumption of processed vegetables increased steadily from 182 pounds in 1970 to 201 pounds in 2017. Since fresh produce is generally consumed raw, the risk of contamination by foodborne pathogens is higher than food that has been cooked (Yeni et al., 2016). Contamination of fresh vegetables with pathogens may occur from farm to fork: through contaminated water, soil, and manure during the preharvest phase or/and water or cross-contamination (such as equipment, surface, handlers) during the postharvest phase. Among foodborne outbreaks with confirmed food vehicle and etiology in the United State from 2004 to 2010, 9.2% were attributed to fresh produce (CDC, 2017). To illustrate patterns of fresh vegetable consumption, fresh vegetable expenditure share is introduced which is given by the ratio of fresh vegetable expenditure to total expenditure (%, $Freshfood$) from the U.S. Bureau of Labor Statistics Consumer Expenditure Surveys (BLS-CE) program. The BLS-CE data are collected in two surveys: the interview survey and the diary survey to measure expenditures, income, and demographic characteristics of consumers in the

United States. Descriptive statistics for the main variables used in this study are summarized in Table 3.1.

3.4 Model Empirical Specification and Estimation Strategy

To test the conceptual predictions (or equivalently estimate the impact of the agricultural work environment on risks associated with foodborne outbreaks), the following empirical specification is utilized:

$$FoodRisk_{it} = \beta_1 Main_{it} + \beta_2 X_{it} + \gamma_t + \varepsilon_{it} \tag{1}$$

where $FoodRisk_{it}$ is the foodborne outbreak-relevant risk measure (i.e., $Illnesses_{it}$, $Doctorvisits_{it}$, $ERvisits_{it}$, or $Hospitalizations_{it}$) for county i in year t; $Main_{it}$ is a 1x3 vector of main independent variables, the three measurements of agriculture production: the ratio of labor costs over total agricultural costs ($gProcessing_{it}$), the ratio of livestock sales over total agricultural sales ($gLivestock_{it}$), and the ratio of farm income over total income ($gFarmIncome_{it}$); $X_{it} = \left(X_{it}^1, \ldots, X_{it}^k\right)$ is a 1×k vector of time-varying covariates that includes: $Physician$, $Foodinsecure$, $Income$, $Unemployed$, $Over65$, $Hispanic$, $Black,\ Freshfood$; γ_t is the year (or time) fixed effects to capture county-invariant annual characteristics that affect food-related disease occurrence (such as the impact of the FDA Food Safety Modernization Act (FSMA) in 2012); β_1 and β_2 are the parameters to be estimated; and ε_{it} is the idiosyncratic random error term.

Our main dependent variable, the measure of foodborne outbreaks within a county, is the normalized count of primary foodborne illnesses. This suggests that count-data models, such as Poisson and Negative Binomial Model, may be appropriate. However, standard count-data model specifications may not be appropriate in cases where a significant number of the counts are zero, as is the case in our application. The proportions of "zero" count observations are 89.4% in

Illnesses, 94.5% in *Doctorvisits*, 95.4 % in *ERvisits*, and 96.2 % in *Hospitalizations*. Figure 3.2 contains histograms of our dependent variables. Given the nature of our data, with a large frequency of extreme observations, the application of a specific distribution that standard county-data models assume may result in critical misspecification since the distributional aspect of the observed count data (i.e., *illnesses, Doctorvisits, ERvisits, Hospitalizations*) may result in a significant understatement of the probability of "zero" foodborne outbreaks counts.

Given the nature of our data with a large frequency of extreme observations, we utilize Cragg's DH model which was initially proposed by Cragg (1971) to address the issues associated with excessive zeros or other truncation issues. DH model combines two models (hurdles):

1. The selection Model: binary model (i.e., probit model) to predict zeros;
2. The outcome Model: Truncated (at zero) normal model to predict the nonzero dependent variable.

The selection model (probit model) determines the boundary points of the dependent variable. It judges whether the hurdle can be cleared (dependent variable>0) or not (dependent variable=0). The outcome model (linear/exponential regression model with (truncated) normal distribution) determines how explanatory factors affect the nonbounded values of the dependent variable (y>0). Therefore, the second model determines the value of the outcome conditional on having cleared the hurdle. For example, one of our dependent variables is the number of outbreaks in county *i* at year *t*. The hurdle model can be characterized by the relationship:

$$y_{it} = s_{it} h_{it}^{*} \tag{2}$$

where y_{it} is the observed number of outbreaks at county i at year *t*. s_{it} is a selection variable and h_{it}^{*} is a continuous latent variable. The selection variable (s_{it}) is 1 if the dependent variable is not bounded and 0 otherwise. If the lower limit that binds the response variable y_{it} is zero, the selection model is given by

$$s_{it} = \begin{cases} 1 & if\ z_{it}\gamma + \varepsilon_{it} > 0 \\ 0 & otherwise \end{cases} \tag{3}$$

where z_{it} is a vector of explanatory variables with a vector of their coefficients γ and ε_{it} is a standard normal error term.

Thus, the continuous latent variable h^*_{it} is observed only if $s_{it} = 1$. We allow the outcome model to be the exponential model:

$$\text{Exponential model: } h^*_{it} = exp(x_{it}\beta + e_{it}) \tag{4}$$

where x_{it} is a vector of explanatory variables to predict non-zero dependent variables with a vector of related coefficients β and e_{it} is a random error term. For the exponential model, the error term has a normal distribution. In the DH model, it does not necessary for the explanatory variables for the two models z_{it} and x_{it} to be the same. This is a key distinction between Cragg's DH model and Tobit's (corner-solution) model. The key limitation of the Tobit model is that the probability of a positive value and the actual value are determined by the same underlying process and the same parameters (Burke, 2009). In the DH model, identical covariates are used in both the selection and main outcome models. To assess the goodness of fit of a standard Tobit and Cragg's double-hurdle model, the likelihood-ratio test has been conducted. The null hypothesis, that is Tobit is a better fitting model, can be rejected (χ^2(1)= 14752.30, *p*<.01.)

3.5 Results and Discussion

Table 3.2 shows the parameter estimates from the DH estimation of equation (1). Note that the probit selection mechanism models the likelihood of the occurrence of one or more of the relevant indicators of food-borne illnesses while the model of the risks applies to the number of illnesses (or other measures of food-borne risks) for the subsample of counties having one or more occurrences of the relevant illness measure. The results are largely consistent across the alternative measures of food-borne illnesses as well as for the discrete selection (probit) model and the continuous level of infections.

The result indicates that the signs of coefficients for the intensity of livestock production (*gLivestock)* and agricultural production (*gFarmincome)* are both positive and have statistically significant effects on the risks associated with foodborne outbreaks at the county level. Counties with more relative animal production are more likely to have a significantly higher level of risk associated with food-borne illnesses. Greater prominence of the overall agriculture industry also tends to heighten the risk. Those results are consistent with a priori expectations. The results suggest that counties with more animal production, and hence more animal wastes and associated hazards, are more likely to have a greater prominence of food-borne diseases. Likewise, counties having a greater prominence of agricultural production in terms of overall income are more likely to realize food-borne illnesses. In contrast, those counties having significant agricultural processing activities appear to be less likely to suffer from food-borne illnesses. This may reflect the fact that food and agricultural product processing, as an industry, is likely to be associated with strong hygienic and worker safety customs and practices.

The demographic patterns associated with the age distribution of a county's population also appear to be significantly related to the incidence of food-borne illnesses. Older individuals tend to have a weakened immune response to foodborne pathogens and thus may be more sensitive to foodborne-relevant risks. This is confirmed in our empirical results, where a greater proportion of the population that is over age 65 tends to be associated with greater food-borne illness risks. Likewise, the elderly also tend to be vulnerable to a wide range of health concerns and thus may be more sensitive to disease exposure of any type, including food-borne contaminants.

Risks appear to differ across different ethnic groups. Relative to whites and Asians, a greater proportion of blacks and Hispanics in a county tends to be associated with fewer reported illnesses and doctor and hospital visits. This may reflect poorer access to health care services and facilities and a greater reliance on self-treatment among these diverse populations. Illnesses are unlikely to be captured in the data if they are self-treated and if doctors and hospitals are not involved in the treatment. Limited access to insurance benefits and health services has long been recognized as an important factor affecting the health and well-being of minorities.

The number of physicians in a county appears to be negatively associated with the risks of illnesses. This is surprising on one hand in that more doctors certainly suggest greater access to health care services and thus more reporting of illnesses. However, standards of health care in counties tend to be associated with the presence of medical facilities and doctors. Greater access to physicians may also be associated with better hygiene and preventative care measures.

One would expect that income would play an important role in the risks of contracting food-borne illnesses. Higher incomes tend to afford accessibility to a wider and potentially healthier range of food sources. Income does not appear to be statistically significant in determining the level of exposure, but it is an important determining factor in the occurrence of an outbreak, as represented in the probit equation. A higher average income in a county is associated with a smaller risk of food-borne illness outbreaks.

To validate the strength and stability of our main model result, we conduct several robustness checks using different indicators of foodborne-relevant risks (*Doctorvisits, ERvisits, Hospitalizations*). By utilizing indicators that measure different risk severity levels, we investigate how agriculture production influences foodborne-associated risks at the different severity levels.

Overall, all results support our main model result: the signs, magnitudes, and significances of most regressor variables are restored.

3.6 Conclusion

The CDC estimates that 1 in 6 Americans experience foodborne illness with diverse symptoms ranging from mild discomfort to intense illness. The estimated economic burden is tens of billions of dollars. While federal/state/local-level efforts have been made to prevent foodborne diseases, the incidence of foodborne illnesses has not declined for many years. In light of the increase in fresh vegetables and fruit in addition to the fast-growing meal-kit industry, the economic losses associated with foodborne illness will also likely continue to grow in the future unless effective food-safety risk management measures can be identified and implemented. The cornerstone to addressing these risks is an understanding of factors related to widespread foodborne illnesses. Host factors are thought to be important in determining the disease outcome following a foodborne infection. However, there is no empirical evidence that carefully documents the relationship between foodborne illness and host factors from the occupational perspective. To fill this gap in the literature, this study empirically examines economic and demographic factors related to both the probability and intensity of a food-borne illness outbreak. We construct a comprehensive county-level data set that include unpublished CDCNORS data and variant health-relevant data. To deal with systematic measurement errors leading to attenuation bias, the double-hurdle model is employed.

Our empirical findings reveal a number of factors that are associated with illness risks. Higher incomes are associated with a lower probability of illness outbreaks. Black and Hispanic populations appear to be less vulnerable to *reported* illnesses, though it is unclear whether this reflects lower risk or poorer access to medical care. Our results indicate that agriculture work environment conditions have a statistically significant effect on foodborne-relevant risks. The results demonstrate that there is a correlation between occupational hygiene in the food industry and the magnitude of foodborne outbreaks at different levels which is an unexplored channel of food safety impacts. Our contribution is that the core empirical finding in our study advances our understanding of linkages between agricultural production and the incidence and nature of foodborne illnesses.

Hence, an important policy implication from the main findings is that there should be information dissemination efforts that impart knowledge to stakeholders (not only farmworkers but also nearby residents) about "the importance of occupational hygiene". The benefit of good occupational hygiene to maintain public health has not been noticed. Federal/state/local agencies also should list "potential reduction in foodborne illnesses of workers and nearby residents" as an additional benefit of good occupational hygiene in their outreach materials. This effort can enhance awareness and understanding of the benefit and likely encourage agricultural stakeholders to improve hygienic environments in their workplace.

This paper improves on previous analysis of how foodborne illness incidence is influenced by the host factor (immune system) which is limitedly investigated as a factor of foodborne illness. An understanding of the factors associated with widespread food-related illnesses is critical to the development of effective management and risk mitigation mechanisms. The issue of insurance protection from the economic losses that may be associated with outbreaks has become increasingly important. The restaurant and prepared food sectors have an obvious risk of liability for alleged outbreak-related illnesses. However, there are other less obvious but equally important businesses that face the risk of an outbreak that could be tied to their farm, farmers' market, truck patch market, and similar such business. Insurers have an essential need to understand factors associated with disease outbreaks. Our findings offer insight into the food product liability insurance market, such as its potential and economic benefits.

Our analysis has several limitations. First, CDCNORS data provide limited information in terms of foodborne illness incidence. Among an estimated 48 million foodborne illnesses, only a small proportion of illnesses are reported and identified as being associated with foodborne contaminants. Likewise, the definition of the foodborne outbreak that the CDC applies makes it harder to distinguish between "true" zero illness due to the nonexistence of foodborne illness and "reported" zero illness due to the nonexistence of reported and confirmed foodborne illness.

References

Ahmad, A., Ghosh, A., Schal, C., and Zurek, L. (2011). Insects in confined swine operations carry a large antibiotic resistant and potentially virulent enterococcal community. *BMC microbiology*, *11*(1), 1-13.

Alam, M. J., and Zurek, L. (2004). Association of Escherichia coli O157: H7 with houseflies on a cattle farm. *Applied and environmental microbiology*, *70*(12), 7578-7580.

Angelo, K. M., Nisler, A. L., Hall, A. J., Brown, L. G., and Gould, L. H. (2017). Epidemiology of restaurant-associated foodborne disease outbreaks, United States, 1998–2013. *Epidemiology and Infection*, *145*(3), 523-534.

Atwood, D., and Paisley-Jones, C. (2017). Pesticides industry sales and usage: 2008–2012 market estimates. *US Environmental Protection Agency, Washington, DC*, *20460*.

Baicker, K., and Chandra, A. (2004). Medicare Spending, The Physician Workforce, And Beneficiaries' Quality of Care: Areas with a high concentration of specialists also show higher spending and less use of high-quality, effective care. *Health Affairs*, *23*(Suppl1), W4-184.

Barrett, K. A., Nakao, J. H., Taylor, E. V., Eggers, C., and Gould, L. H. (2017). Fish-associated foodborne disease outbreaks: United States, 1998–2015. *Foodborne Pathogens and Disease*, *14*(9), 537-543.

Basinas, I., Sigsgaard, T., Kromhout, H., Heederik, D., Wouters, I. M., and Schlünssen, V. (2015). A comprehensive review of levels and determinants of personal exposure to dust and endotoxin in livestock farming. *Journal of exposure science and environmental epidemiology*, *25*(2), 123-137.

Beaumont, J. J., Goldsmith, D. F., Morrin, L. A., and Schenker, M. B. (1995). Mortality in agricultural workers after compensation claims for respiratory disease, pesticide illness, and injury. *Journal of occupational and environmental medicine*, 160-169.

Bellemare, M. F., and Nguyen, N. (2018). Farmers markets and Food-Borne illness. *American Journal of Agricultural Economics*, *100*(3), 676-690.

Bhaskaran, N., Quigley, C., Paw, C., Butala, S., Schneider, E., and Pandiyan, P. (2018). Role of short chain fatty acids in controlling tregs and immunopathology during mucosal infection. *Frontiers in microbiology*, 1995.

Borlée, F., Yzermans, C. J., Aalders, B., Rooijackers, J., Krop, E., Maassen, C. B., ... and Smit, L. A. (2017). Air pollution from livestock farms is associated with airway obstruction in neighboring residents. *American journal of respiratory and critical care medicine*, *196*(9), 1152-1161.

Borlée, F., Yzermans, C. J., van Dijk, C. E., Heederik, D., and Smit, L. A. (2015). Increased respiratory symptoms in COPD patients living in the vicinity of livestock farms. *European Respiratory Journal*, *46*(6), 1605-1614.

Brackbill, R. M., Cameron, L. L., and Behrens, V. (1994). Prevalence of chronic diseases and impairments among US farmers, 1986–1990. *American journal of epidemiology*, *139*(11), 1055-1065.

Bradman, A., Whitaker, D., Quiros, L., Castorina, R., Henn, B. C., Nishioka, M., ... and Eskenazi, B. (2007). Pesticides and their metabolites in the homes and urine of farmworker children living in the Salinas Valley, CA. *Journal of exposure science and environmental epidemiology*, *17*(4), 331-349.

Braveman, P., Egerter, S., and Williams, D. R. (2011). The social determinants of health: coming of age. *Annual review of public health*, *32*, 381-398.

Burkholder, J., Libra, B., Weyer, P., Heathcote, S., Kolpin, D., Thorne, P. S., and Wichman, M. (2007). Impacts of waste from concentrated animal feeding operations on water quality. *Environmental health perspectives*, *115*(2), 308-312.

Buzby, J. C. (2002). Older adults at risk of complications from microbial foodborne illness. *Food Review/National Food Review*, *25*(1482-2016-121529), 30-35.

Caballero, M., Amiri, S., Denney, J. T., Monsivais, P., Hystad, P., and Amram, O. (2018). Estimated residential exposure to agricultural chemicals and premature mortality by Parkinson's disease in Washington state. *International journal of environmental research and public health*, *15*(12), 2885.

Calvert, G. M., Karnik, J., Mehler, L., Beckman, J., Morrissey, B., Sievert, J., ... and Moraga-McHaley, S. (2008). Acute pesticide poisoning among agricultural workers in the United States, 1998–2005. *American journal of industrial medicine*, *51*(12), 883-898.

Centers for Disease Control and Prevention. (2018, March 26). Veterinary Safety and Health. https://www.cdc.gov/niosh/topics/veterinary/chemical.html

Centers for Disease Control and Prevention. (2019, January 24). People With a Higher Risk of Food Poisoning. https://www.cdc.gov/foodsafety/people-at-risk-food-poisoning.html

Chai, S. J., Cole, D., Nisler, A., and Mahon, B. E. (2017). Poultry: The most common food in outbreaks with known pathogens, United States, 1998–2012. *Epidemiology and Infection*, *145*(2), 316-325.

Chari, R., Kress, A. M., and Madrigano, J. (2018). Injury and illness surveillance of US agricultural workers: assessment of recommendations and actions. *Rand health quarterly*, *8*(2).

Clancy C., Munier W., Brady J., Moy, E., Chaves, K., Freeman, W., and Bonnett, D. (2012) 2012 National Healthcare Quality Report. Rockville, MD: *Agency for Healthcare Research and Quality (AHRQ)*.

Collins, S. R., Schoen, C., Colasanto, D., and Downey, D. A. (2003). On the Edge: Low-Wage Workers and Their Health Insurance Coverage. *The Commonwealth Fund, New York, NY*.

County Health Rankings. (n.d.). County Health Rankings. Retrieved Feb 1, 2022, from https://www.countyhealthrankings.org/explore-health-rankings/measures-data-sources/county-health-rankings-model/

Cullor, J. S. (1997). HACCP (Hazard Analysis Critical Control Points): is it coming to the dairy?. *Journal of Dairy Science*, *80*(12), 3449-3452.

De Rooij, M. M., Smit, L. A., Erbrink, H. J., Hagenaars, T. J., Hoek, G., Ogink, N. W., ... and Wouters, I. M. (2019). Endotoxin and particulate matter emitted by livestock farms and respiratory health effects in neighboring residents. *Environment international*, *132*, 105009.

De W Blackburn, C., and McClure, P. J. (Eds.). (2009). *Foodborne pathogens: hazards, risk analysis and control*. Elsevier.

Dhurandhar, E. J. (2016). The food-insecurity obesity paradox: A resource scarcity hypothesis. *Physiology and behavior, 162*, 88-92.

Diez-Gonzalez, F., and Mukherjee, A. (2009). Produce safety in organic vs. conventional crops. *Microbial Safety of Fresh Produce*, 81-99.

Donham, K. J. (2010). Community and occupational health concerns in pork production: A review. *Journal of animal science*, *88*(suppl_13), E102-E111.

Donham, K. J., Reynolds, S. J., Whitten, P., Merchant, J. A., Burmeister, L., and Popendorf, W. J. (1995). Respiratory dysfunction in swine production facility workers: Dose-response relationships of environmental exposures and pulmonary function. *American journal of industrial medicine*, *27*(3), 405-418.

Dooley, D., Fielding, J., and Levi, L. (1996). Health and unemployment. *Annual review of public health*, *17*(1), 449-465.

Douglas, P., Robertson, S., Gay, R., Hansell, A. L., and Gant, T. W. (2018). A systematic review of the public health risks of bioaerosols from intensive farming. *International journal of hygiene and environmental health*, *221*(2), 134-173.

Elliott, L., Yeatts, K., and Loomis, D. (2004). Ecological associations between asthma prevalence and potential exposure to farming. *European Respiratory Journal*, *24*(6), 938-941.

Fahrenfeld, N., Knowlton, K., Krometis, L. A., Hession, W. C., Xia, K., Lipscomb, E., ... and Pruden, A. (2014). Effect of manure application on abundance of antibiotic resistance genes and their attenuation rates in soil: field-scale mass balance approach. *Environmental science and technology*, *48*(5), 2643-2650.

Farber, J. M., Ross, W. H., and Harwig, J. (1996). Health risk assessment of Listeria monocytogenes in Canada. *International journal of food microbiology*, *30*(1-2), 145-156.

Food and Drug Administration, U.S. Department of Agriculture and Centers for Disease Control and Prevention. (1998). Guidance for Industry: Guide to Minimize Microbial Food Safety Hazard for Fresh Fruits and Vegetables. Food and Drug Administration, Washington, D.C. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-industry-guide-minimize-microbial-food-safety-hazards-fresh-fruits-and-vegetables

Food and Drug Administration. (2017). HACCP principles and application guidelines. Hazard Analysis Critical Control Point, US Food and Drug Administration, Silver Spring, MD.

Franco, M., Diez-Roux, A. V., Nettleton, J. A., Lazo, M., Brancati, F., Caballero, B., ... and Moore, L. V. (2009). Availability of healthy foods and dietary patterns: the Multi-Ethnic Study of Atherosclerosis. *The American journal of clinical nutrition*, *89*(3), 897-904.

Frank, A. L., McKnight, R., Kirkhorn, S. R., and Gunderson, P. (2004). Issues of agricultural safety and health. *Annu. Rev. Public Health*, *25*, 225-245.

Freeman, N. C., Shalat, S. L., Black, K., Jimenez, M., Donnelly, K. C., Calvin, A., and Ramirez, J. (2004). Seasonal pesticide use in a rural community on the US/Mexico border. Journal of Exposure Science and Environmental Epidemiology, 14(6), 473-478.

Galea, S., Tracy, M., Hoggatt, K. J., DiMaggio, C., and Karpati, A. (2011). Estimated deaths attributable to social factors in the United States. *American journal of public health*, *101*(8), 1456-1465.

Geissler, A. L., Bustos Carrillo, F., Swanson, K., Patrick, M. E., Fullerton, K. E., Bennett, C., ... and Mahon, B. E. (2017). Increasing Campylobacter infections, outbreaks, and antimicrobial resistance in the United States, 2004–2012. *Clinical Infectious Diseases*, *65*(10), 1624-1631.

Gilbert, P. A., and Khokhar, S. (2008). Changing dietary habits of ethnic groups in Europe and implications for health. *Nutrition reviews*, *66*(4), 203-215.

Gold, L. S., Ward, M. H., Dosemeci, M., and Roos, A. D. (2007). Systemic autoimmune disease mortality and occupational exposures. *Arthritis and Rheumatism: Official Journal of the American College of Rheumatology*, *56*(10), 3189-3201.

Goodwin, B. K., and Piggott, N. E. (2009). Spatiotemporal modeling of Asian citrus canker risks: implications for insurance and indemnification fund models. *American journal of agricultural economics*, *91*(4), 1038-1055.

Gould, L. H., Rosenblum, I., Nicholas, D., Phan, Q., and Jones, T. F. (2013). Contributing factors in restaurant-associated foodborne disease outbreaks, FoodNet sites, 2006 and 2007. *Journal of food protection*, *76*(11), 1824-1828.

Graham, J. P., Price, L. B., Evans, S. L., Graczyk, T. K., and Silbergeld, E. K. (2009). Antibiotic resistant enterococci and staphylococci isolated from flies collected near confined poultry feeding operations. *Science of the total environment*, *407*(8), 2701-2710.

Gunier, R. B., Ward, M. H., Airola, M., Bell, E. M., Colt, J., Nishioka, M., ... and Nuckols, J. R. (2011). Determinants of agricultural pesticide concentrations in carpet dust. *Environmental health perspectives*, *119*(7), 970-976.

Harvey, R., Zakhour, C. M., and Gould, L. H. (2016). Foodborne disease outbreaks associated with organic foods in the United States. *Journal of food protection*, *79*(11), 1953-1958.

Hille, K., Ruddat, I., Schmid, A., Hering, J., Hartmann, M., von Münchhausen, C., ... and Kreienbrock, L. (2017). Cefotaxime-resistant E. coli in dairy and beef cattle farms—Joint analyses of two cross-sectional investigations in Germany. *Preventive veterinary medicine*, *142*, 39-45.

Hoffmann, S., Batz, M. B., and Morris, J. G. (2012). Annual cost of illness and quality-adjusted life year losses in the United States due to 14 foodborne pathogens. *Journal of food protection*, *75*(7), 1292-1302.

Hribar, C. (2010). Understanding concentrated animal feeding operations and their impact on communities. *National Association of Local Boards of Health , Bowling Green, OH. Environmental Health pub*

Huijsdens, X. W., van Santen-Verheuvel, M. G., Spalburg, E., Heck, M. E., Pluister, G. N., Eijkelkamp, B. A., ... and Wannet, W. J. (2006). Multiple cases of familial transmission of community-acquired methicillin-resistant Staphylococcus aureus. *Journal of clinical microbiology*, *44*(8), 2994-2996.

Kirkhorn, S. R., and Garry, V. F. (2000). Agricultural lung diseases. *Environmental health perspectives, 108*(suppl 4), 705-712.

Kissling, W. D., and Carl, G. (2008). Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology and Biogeography*, *17*(1), 59-71.

Koike, S., Krapac, I. G., Oliver, H. D., Yannarell, A. C., Chee-Sanford, J. C., Aminov, R. I., and Mackie, R. I. (2007). Monitoring and source tracking of tetracycline resistance genes in lagoons and groundwater adjacent to swine production facilities over a 3-year period. *Applied and environmental microbiology*, *73*(15), 4813-4823.

Li, Y., Liao, H., and Yao, H. (2019). Prevalence of antibiotic resistance genes in air-conditioning systems in hospitals, farms, and residences. *International journal of environmental research and public health*, *16*(5), 683.

Linaker, C., and Smedley, J. (2002). Respiratory illness in agricultural workers. *Occupational medicine*, *52*(8), 451-459.

Lund, B. M. (2016). Microbiological safety of food, particularly for vulnerable people. *Journal of Family Medicine and Disease Prevention*, *2*, 035.

Lund, B. M., and O'Brien, S. J. (2011). The occurrence and prevention of foodborne disease in vulnerable people. *Foodborne pathogens and disease*, *8*(9), 961-973.

Lynch, M. F., Tauxe, R. V., and Hedberg, C. W. (2009). The growing burden of foodborne outbreaks due to contaminated fresh produce: risks and opportunities. *Epidemiology and Infection*, *137*(3), 307-315.

Ma, Z., Lee, S., and Jeong, K. C. (2019). Mitigating antibiotic resistance at the livestock-environment interface: a review. *Journal of microbiology and biotechnology*, *29*(11), 1683-1692.

Markland, S., Weppelmann, T. A., Ma, Z., Lee, S., Mir, R. A., Teng, L., ... and Jeong, K. C. (2019). High prevalence of cefotaxime resistant bacteria in grazing beef cattle: a cross sectional study. *Frontiers in microbiology*, *10*, 176.

Marlow, M. A., Luna-Gierke, R. E., Griffin, P. M., and Vieira, A. R. (2017). Foodborne disease outbreaks in correctional institutions—United States, 1998–2014. *American journal of public health*, *107*(7), 1150-1156.

Marsh, Z., Shah, M. P., Wikswo, M. E., Barclay, L., Kisselburgh, H., Kambhampati, A., ... and Hall, A. J. (2018). Epidemiology of foodborne norovirus outbreaks–United States, 2009–2015. *Food Safety*, *6*(2), 58-66.

Marti, R., Tien, Y. C., Murray, R., Scott, A., Sabourin, L., and Topp, E. (2014). Safely coupling livestock and crop production systems: how rapidly do antibiotic resistance genes dissipate in soil following a commercial application of swine or dairy manure?. *Applied and environmental microbiology*, *80*(10), 3258-3265.

May, S., Romberger, D. J., and Poole, J. A. (2012). Respiratory health effects of large animal farming environments. *Journal of Toxicology and Environmental Health, Part B*, *15*(8), 524-541.

McCauley, L. A., Lasarev, M. R., Higgins, G., Rothlein, J., Muniz, J., Ebbert, C., and Phillips, J. (2001). Work characteristics and pesticide exposures among migrant agricultural families: a community-based research approach. *Environmental health perspectives*, *109*(5), 533-538.

McClendon, C. J., Gerald, C. L., and Waterman, J. T. (2015). Farm animal models of organic dust exposure and toxicity: Insights and implications for respiratory health. *Current opinion in allergy and clinical immunology*, *15*(2), 137.

McKee-Ryan, F., Song, Z., Wanberg, C. R., and Kinicki, A. J. (2005). Psychological and physical well-being during unemployment: a meta-analytic study. *Journal of applied psychology*, *90*(1), 53.

Merchant, J. A., Naleway, A. L., Svendsen, E. R., Kelly, K. M., Burmeister, L. F., Stromquist, A. M., ... and Chrischilles, E. A. (2005). Asthma and farm exposures in a cohort of rural Iowa children. *Environmental health perspectives*, *113*(3), 350-356.

Mirabelli, M. C., Wing, S., Marshall, S. W., and Wilcosky, T. C. (2006). Asthma symptoms among adolescents who attend public schools that are located near confined swine feeding operations. *Pediatrics*, *118*(1), e66-e75.

Mitloehner, F. M., and Schenker, M. B. (2007). Environmental exposure and health effects from concentrated animal feeding operations. *Epidemiology*, *18*(3), 309-311.

Montecino-Rodriguez, E., Berent-Maoz, B., and Dorshkind, K. (2013). Causes, consequences, and reversal of immune system aging. *The Journal of clinical investigation*, *123*(3), 958-965.

Morris, M. L., Binswanger-Mikhize, H. P., and Byerlee, D. (2009). *Awakening Africa's sleeping giant: prospects for commercial agriculture in the Guinea Savannah Zone and beyond*. World Bank Publications.

Neyra, R. C., Vegosen, L., Davis, M. F., Price, L., and Silbergeld, E. K. (2012). Antimicrobial-resistant bacteria: an unrecognized work-related risk in food animal production. *Safety and health at work*, *3*(2), 85-91.

Nolte, K., and Ostermeier, M. (2017). Labour market effects of large-scale agricultural investment: conceptual considerations and estimated employment effects. *World Development*, *98*, 430-446.

Nord, M., Andrews, M., and S, Carlson. (2005). Measuring Food Security in the United States: Household food security in the United States, 2005. USDA Economic Research Service, Washington, Report No. ERR-29.

Nordgren, T. M., and Bailey, K. L. (2016). Pulmonary health effects of agriculture. *Current opinion in pulmonary medicine*, *22*(2), 144.

Ozaki, V. A., Ghosh, S. K., Goodwin, B. K., and Shirota, R. (2008). Spatio-temporal modeling of agricultural yield data with an application to pricing crop insurance contracts. *American journal of agricultural economics*, *90*(4), 951-961.

Parks, C. G., Hoppin, J. A., De Roos, A. J., Costenbader, K. H., Alavanja, M. C., and Sandler, D. P. (2016). Rheumatoid arthritis in agricultural health study spouses: associations with pesticides and other farm exposures. *Environmental health perspectives*, *124*(11), 1728-1734.

Pollack, S. L. (2001). Consumer demand for fruit and vegetables: the US example. *Changing structure of global food consumption and trade*, *6*, 49-54.

Poole, J. A., and Romberger, D. J. (2012). Immunological and inflammatory responses to organic dust in agriculture. *Current opinion in allergy and clinical immunology*, *12*(2), 126.

Radon, K., Schulze, A., Ehrenstein, V., van Strien, R. T., Praml, G., and Nowak, D. (2007). Environmental exposure to confined animal feeding operations and respiratory health of neighboring residents. *Epidemiology*, 300-308.

Rafee, Y., Abdel-Haq, N., Asmar, B., Salimnia, T., Pharm, C. V., Pharm, M. J. R., and Amjad, M. (2012). Increased prevalence of methicillin-resistant Staphylococcus aureus nasal colonization in household contacts of children with community acquired disease. *BMC infectious diseases*, *12*(1), 1-7.

Rein, B. K. (1992). Health hazards in agriculture—an emerging issue. Farm safety fact sheet. Washington, DC: National Agriculture Safety Database [NASD]. US Department of Agriculture Extension Service.

Ropkins, K., and Beck, A. J. (2000). Evaluation of worldwide approaches to the use of HACCP to control food safety. *Trends in Food Science and Technology*, *11*(1), 10-21.

Ross, C. E., and Mirowsky, J. (1995). Does employment affect health?. *Journal of Health and social Behavior*, 230-243.

Sahlander, K., Larsson, K., and Palmberg, L. (2012). Daily exposure to dust alters innate immunity. *PLoS One*, *7*(2), e31646.

Sapkota, A. R., Curriero, F. C., Gibson, K. E., and Schwab, K. J. (2007). Antibiotic-resistant enterococci and fecal indicators in surface water and groundwater impacted by a concentrated swine feeding operation. *Environmental Health Perspectives*, *115*(7), 1040-1045.

Scallan, E., Hoekstra, R. M., Angulo, F. J., Tauxe, R. V., Widdowson, M. A., Roy, S. L., ... and Griffin, P. M. (2011). Foodborne illness acquired in the United States—major pathogens. *Emerging infectious diseases*, *17*(1), 7.

Scharff, R. L. (2012). Economic burden from health losses due to foodborne illness in the United States. *Journal of food protection*, *75*(1), 123-131.

Schenker, M. B., Christiani, D., Cormier, Y., Dimich-Ward, H., Doekes, G., Dosman, J., ... and Chan-Yeung, M. (1998). Respiratory health hazards in agriculture. *American journal of respiratory and critical care medicine*, *158*(5), S1-S76.

Schenker, M. B., Pinkerton, K. E., Mitchell, D., Vallyathan, V., Elvine-Kreis, B., and Green, F. H. (2009). Pneumoconiosis from agricultural dust exposure among young California farmworkers. *Environmental health perspectives*, *117(6)*, 988-994.

Schinasi, L., Horton, R. A., Guidry, V. T., Wing, S., Marshall, S. W., and Morland, K. B. (2011). Air pollution, lung function, and physical symptoms in communities near concentrated swine feeding operations. *Epidemiology (Cambridge, Mass.)*, *22*(2), 208.

Schmid, A., Hörmansdorfer, S., Messelhäusser, U., Käsbohrer, A., Sauter-Louis, C., and Mansfeld, R. (2013). Prevalence of extended-spectrum β-lactamase-producing Escherichia coli on Bavarian dairy and beef cattle farms. *Applied and environmental microbiology*, *79*(9), 3027-3032.

Schulze, A., van Strien, R., Ehrenstein, V., Schierl, R., Kuchenhoff, H., and Radon, K. (2006). Ambient endotoxin level in an area with intensive livestock production. *Annals of Agricultural and Environmental Medicine*, *13*(1), 87-91.

Sela, S., and Manulis-Sasson, S. (2015). What else can we do to mitigate contamination of fresh produce by foodborne pathogens?. *Microbial Biotechnology*, *8*(1), 29.

Semega, J., Kollar, M., Shrider, E. Creamer, J. (2020). Income and Poverty in the United States. Retrieved from https://www.census.gov/content/dam/Census/library/publications/2020/demo/p60-270.pdf

Simon, A. K., Hollander, G. A., and McMichael, A. (2015). Evolution of the immune system in humans from infancy to old age. *Proceedings of the Royal Society B: Biological Sciences*, 282(1821), 20143085.

Smith, J. L. (1998). Foodborne illness in the elderly. *Journal of food protection, 61(9)*, 1229-1239.

Smith, J. L. (2017). Infectious dose and an aging population: Susceptibility of the aged to foodborne pathogens. In *Foodborne pathogens* (pp. 451-468). Springer, Cham.

Soupir, M. L., Mostaghimi, S., Yagow, E. R., Hagedorn, C., and Vaughan, D. H. (2006). Transport of fecal bacteria from poultry litter and cattle manures applied to pastureland. *Water, Air, and Soil Pollution*, *169*(1), 125-136.

Thorne, P. S., Kulhánková, K., Yin, M., Cohn, R., Arbes Jr, S. J., and Zeldin, D. C. (2005). Endotoxin exposure is a risk factor for asthma: the national survey of endotoxin in United States housing. *American journal of respiratory and critical care medicine*, *172*(11), 1371-1377.

U.S. Department of Agriculture- Food Safety and Inspection Service (1996). Pathogen reduction: Hazard Analysis and Critical Control Point (HACCP) Systems-Final Rule. Federal Register pp. 38806-38989, 1996

Van de Giessen, A. W., van Santen-Verheuvel, M. G., Hengeveld, P. D., Bosch, T., Broens, E. M., and Reusken, C. B. E. M. (2009). Occurrence of methicillin-resistant Staphylococcus aureus in rats living on pig farms. *Preventive veterinary medicine*, *91*(2-4), 270-273.

Van Dijk, C. E., Zock, J. P., Baliatsas, C., Smit, L. A., Borlée, F., Spreeuwenberg, P., ... and Yzermans, C. J. (2017). Health conditions in rural areas with high livestock density: Analysis of seven consecutive years. *Environmental Pollution*, *222*, 374-382.

Von Ehrenstein, O. S., Ling, C., Cui, X., Cockburn, M., Park, A. S., Yu, F., ... and Ritz, B. (2019). Prenatal and infant exposure to ambient pesticides and autism spectrum disorder in children: population based case-control study. *bmj*, *364*.

Von Essen, S. G., and McCurdy, S. A. (1998). Health and safety risks in production agriculture. *Western Journal of Medicine*, *169*(4), 214.

Wanberg, C. R., Kammeyer-Mueller, J. D., and Shi, K. (2002). Job loss and the experience of unemployment: International research and perspectives.

Ward, M. H., Lubin, J., Giglierano, J., Colt, J. S., Wolter, C., Bekiroglu, N., ... and Nuckols, J. R. (2006). Proximity to crops and residential exposure to agricultural herbicides in Iowa. *Environmental health perspectives*, *114*(6), 893-897.

Whitham, H. K., Sundararaman, P., Dewey-Mattia, D., Manikonda, K., Marshall, K. E., Griffin, P. M., ... and Crowe, S. J. (2021). Novel Outbreak-Associated Food Vehicles, United States. *Emerging Infectious Diseases*, *27*(10), 2554.

Ye, M., Beach, J., Martin, J. W., and Senthilselvan, A. (2013). Occupational pesticide exposures and respiratory health. *International Journal of Environmental Research and Public Health*, *10*(12), 6442-6471.

Yeni, F., Yavaş, S., Alpas, H. A. M. I., and Soyer, Y. E. S. I. M. (2016). Most common foodborne pathogens and mycotoxins on fresh produce: a review of recent outbreaks. *Critical reviews in food science and nutrition*, *56*(9), 1532-1544.

Figures and Tables

Table 3.1. Summary statistics of variables used in the baseline empirical models, 2009~2018 (N=23,745)

Variable	Definition	Mean	Std.Dev	Min	Max
Illnesses	Number of foodborne primary cases /100 thousand population	4.72	86.62	0.00	10855.95
Doctorvisits	Number of primary cases visiting a health care provider /100 thousand population	0.63	16.83	0.00	1773.94
ERvisits	Number of primary cases visiting an emergency room /100 thousand population	0.26	4.70	0.00	320.22
Hospitalizations	Number of primary cases being hospitalized /100 thousand population	0.14	2.73	0.00	223.98
gLivestock	Livestock sales/Total sales, geographically aggregated	0.49	0.22	0.00	0.99
gProcessing	Hired farm labor expenses /total costs, geographically aggregated	0.10	0.07	0.02	0.43
gFarmincome	Farm income/Total income, geographically aggregated	0.03	0.05	-0.05	0.49
Physician	Primary care providers other than physicians/100 thousand population	54.40	32.95	0.00	598.80
Foodinsecure	Population who lacks adequate access to food (%)	14.44	3.87	3.40	37.90
Income	Per capita income, thousand 2010 USD	0.36	0.10	0.15	1.66
Unemployed	Population (age 16 and older) unemployed and looking for work (%)	6.85	3.02	1.54	29.70
Over65	Population of age 65 and older (%)	17.21	4.44	2.62	57.60
Hispanic	Population of Hispanic (%)	9.31	13.66	0.00	97.25
Black	Population of non-Hispanic black (%)	8.82	13.74	0.00	85.58
Freshfood	Fresh vegetables expenditure /Total expenditure (%)	0.60	0.14	0.13	1.27

Table 3.2. Empirical Estimates of Model of Food-Borne Illness Risks: Double Hurdle Model

Parameter	Dependent variable: *Illnesses* Estimate	t value	Dependent variable: *Doctorvisits* Estimate	t value	Dependent variable: *ERvisits* Estimate	t value	Dependent variable: *Hospitalizations* Estimate	t value
Outcome Model: Model of Positive Dependent Variable								
gLivestock	0.72	3.74***	0.68	2.29**	0.89	3.08***	0.87	3.06***
gProcessing	-4.45	-7.93***	-4.64	-6.15***	-4.59	-5.25***	-4.74	-5.99***
gFarmincome	11.19	8.57***	11.02	6.24***	10.97	5.35***	12.25	7.23***
Physician	-0.00	-2.87***	-0.01	-3.92***	-0.01	-5.74***	-0.01	-5.01***
Foodinsecure	0.01	0.70	0.01	0.30	0.00	0.05	-0.01	-0.31
Income	-0.42	-0.71	-0.73	-1.16	-0.49	-0.92	-0.32	-0.48
Unemployed	0.04	1.63	0.02	0.51	0.03	0.64	0.08	2.18**
Over65	0.05	4.37***	0.06	3.51***	0.06	3.73***	0.06	4.27***
Hispanic	-0.02	-5.82***	-0.02	-3.87***	-0.03	-4.35***	-0.04	-6.13***
Black	-0.02	-5.20***	-0.02	-3.83***	-0.02	-2.83***	-0.01	-2.26**
Freshfood	0.20	1.01	-0.24	-0.69	0.40	1.14	0.21	0.62
Intercept	2.51	5.79***	1.69	2.87***	0.82	1.44	-0.10	-0.16
Selection Model: Probit								
gLivestock	0.55	5.58***	0.46	4.40***	0.37	3.36***	0.37	3.44***
gProcessing	-3.70	-10.05***	-3.51	-8.99***	-3.10	-7.79***	-2.75	-7.06***
gFarmincome	3.05	6.35***	2.00	3.76***	3.39	4.96***	2.43	3.52***
Physician	-0.01	-5.73***	-0.01	-5.23***	-0.01	-5.04***	-0.01	-5.08***
Foodinsecure	0.04	4.64***	0.03	2.89***	0.02	2.30**	0.02	2.05**
Income	-0.90	-2.82***	-0.89	-3.18***	-0.90	-2.86***	-1.04	-3.95***
Unemployed	0.00	0.59	0.01	0.59	0.02	1.84*	-0.00	-0.19
Over65	0.04	7.35***	0.04	7.15***	0.03	6.03***	0.04	6.01***
Hispanic	0.00	1.38	0.00	0.61	-0.00	-0.07	0.00	0.17
Black	-0.01	-3.76***	-0.01	-2.95***	-0.01	-4.90***	-0.01	-3.30***
Freshfood	-0.14	-1.50	-0.13	-1.12	-0.02	-0.13	0.02	0.16
Intercept	0.96	4.69***	1.50	7.31***	1.62	7.50***	1.78	8.91***
No. of Obs.		23745		23745		23745		23745
Pseudo R^2		0.13		0.18		0.18		0.20
Log Likelihood		-14236.14		-5715.30		-4609.22		-3648.20

Notes: All columns include year fixed effects; Errors are clustered by county, and t statistics are shown in parenthesis. For the double-hurdle model, coefficients of explanatory variables are reported instead of marginal effects in the main section; * p<0.1, ** p<0.05, *** p<0.01.

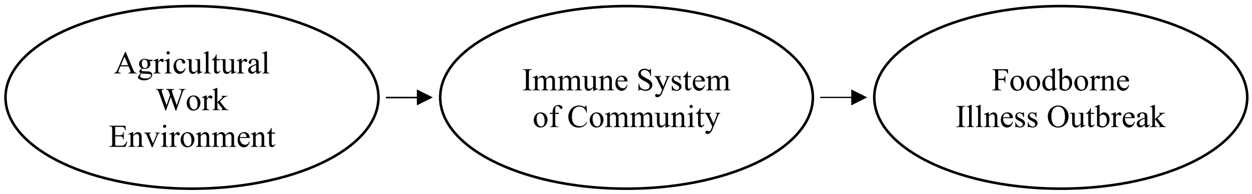
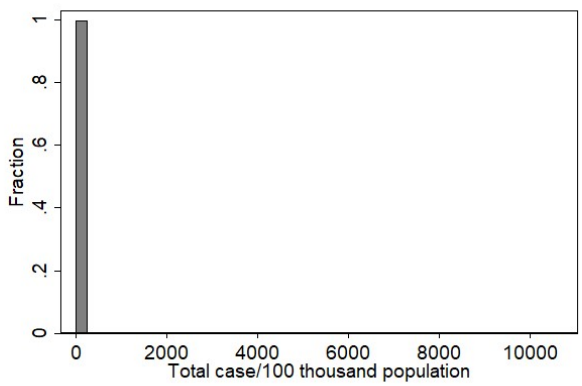
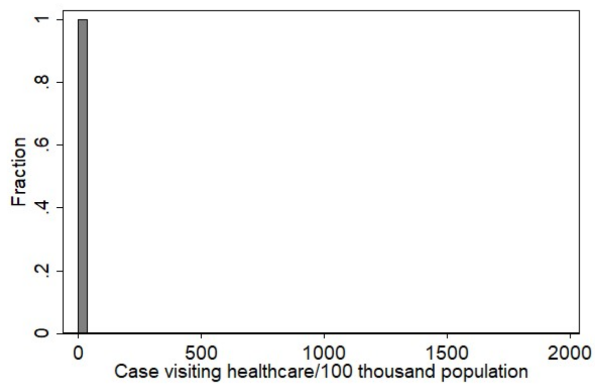


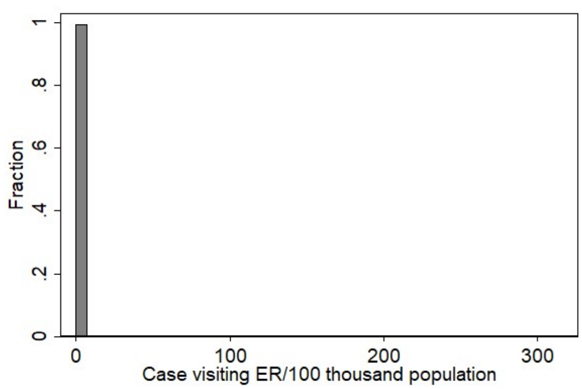
Figure 3.1. Diagram of conceptual framework



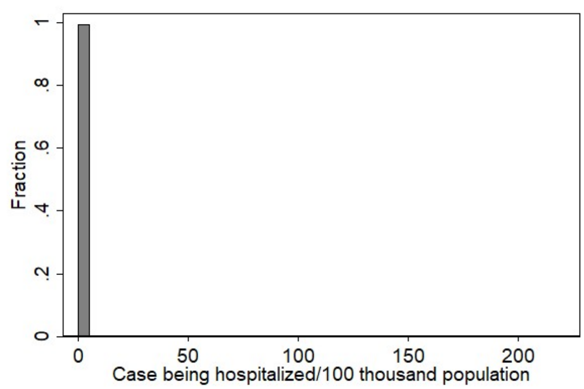
(a) Count Fraction: *Illnesses*



(b) Count Fraction: *Doctorvisits*



(c) Count Fraction: *ERvisits*



(d) Count Fraction: *Hospitalizations*

Figure 3.2. The preponderance of zeros problem

APPENDICES

Appendix A

Appendix for Chapter 1

Table A.1. Prevent planting stage codes and crop insurance plans included in the study

Type	Name	Description
Stage Code	H3	Prevented Planting Option 3 - Harvested Acres
	NP	Prevented Planting - Insured Crop NOT Planted
	P1	Prevented Planting Option 1
	P2	Prevented Planting Option 2 (Unplanted Acres-98)
	P3	Prevented Planting Option 3
	P4	Prevented Planting Option 4
	PF	Prevented Planting - Unplanted Acreage with 5% buyup option
	PL	Preliminary Loss Payment (GRP): Prevented Planting Endorsement
Insurance Plan	YP	Yield Protection
	RP	Revenue Protection
	RPHPE	Revenue Protection with Harvest Price Exclusion
	RP& RPHPE	Revenue Plans (RP, RPHPE)
	RP(HPE)&YP	3 Main Plans (RP, RPHPE, YP)

Notes: Crop insurance plans have changed their names, merged by other plan(s), or became obsolete over time. These no longer valid insurance plans are regrouped as the following: Actual Production History (APH)→YP; Crop Revenue Coverage (CRC), Income Protection (IP)→RP; Revenue Assurance (RA)→RPHPE.

Table A.2. Robustness check: Regression results using RP insurance plans only

Independent Variables	Dependent Variable: PP-LCR		Dependent Variable: PP-LAR	
	Linear FE	Lewbel IV	Linear FE	Lewbel IV
CC	-0.044147***	-0.051806***	-0.070852***	-0.090312***
	(-4.26)	(-2.59)	(-4.96)	(-3.25)
April-May DD10	0.010723***	0.010676***	0.020108***	0.019987***
	(5.21)	(5.37)	(7.15)	(7.37)
April-May DD1029	-0.001211	-0.001235	-0.002601	-0.002662
	(-0.81)	(-0.88)	(-1.33)	(-1.43)
April-May DD30+	-1.3e+00	-1.4e+00	3.936532	3.808168
	(-0.14)	(-0.15)	(0.42)	(0.42)
April-May PDSI_W	0.395342***	0.395723***	0.534843***	0.535810***
	(7.62)	(7.94)	(8.72)	(9.10)
April-May PDSI_D	-0.085055*	-0.085468*	-0.127925**	-0.128973**
	(-1.72)	(-1.83)	(-1.98)	(-2.10)
No. of Obs.	7748	7748	7748	7748
Hansen J Statistic		36.709		38.662
Hansen J P-value		0.0014		0.0007
Kleibergen-Paap rk Wald F Statistic		47.704		47.704
Kleibergen-Paap rk LM Statistic		86.571		86.571
Adj R-squared	0.2118	-0.0008	0.2490	0.0214

Notes: All columns include county and year fixed effects; Errors are clustered by county; t statistics are shown in parenthesis; * p<0.1, ** p<0.05, *** p<0.01.

Table A.3. Robustness check: Regression results using two-way standard error clustering

Independent Variables	Dependent Variable: PP-LCR		Dependent Variable: PP-LAR	
	Linear FE	Lewbel IV	Linear FE	Lewbel IV
CC	-0.034277***	-0.037688**	-0.056392***	-0.047964**
	(-4.18)	(-2.52)	(-4.94)	(-2.39)
April-May DD10	0.009388***	0.009367***	0.018519***	0.018571***
	(5.46)	(5.70)	(7.65)	(8.02)
April-May DD1029	0.000137	0.000126	0.000661	0.000687
	(0.12)	(0.12)	(0.41)	(0.45)
April-May DD30+	-3.3e+00	-3.3e+00	0.260204	0.316473
	(-0.33)	(-0.35)	(0.02)	(0.03)
April-May PDSI_W	0.377019***	0.377188***	0.536918***	0.536500***
	(13.00)	(13.57)	(14.63)	(15.27)
April-May PDSI_D	-0.149633***	-0.149822***	-0.252274***	-0.251805***
	(-3.80)	(-3.99)	(-4.48)	(-4.68)
No. of Obs.	7752	7752	7752	7752
Hansen J Statistic		42.265		42.176
Hansen J P-value		0.0002		0.0002
Kleibergen-Paap rk Wald F Statistic		32.914		32.914
Kleibergen-Paap rk LM Statistic		69.488		69.488
Adj R-squared	0.2649	0.0446	0.2938	0.0736

Notes: All columns include county and year fixed effects; Errors are clustered by county; t statistics are shown in parenthesis; * p<0.1, ** p<0.05, *** p<0.01.

Table A.4. Robustness check: Regression results with separate April and May weather variables

Independent Variables	Dependent Variable: PP-LCR		Dependent Variable: PP-LAR	
	Linear FE	Lewbel IV	Linear FE	Lewbel IV
CC	-0.030831***	-0.035844**	-0.049355***	-0.041357**
	(-4.03)	(-2.48)	(-4.21)	(-2.06)
April DD10	0.006407***	0.006360***	0.018432***	0.018507***
	(3.46)	(3.58)	(7.08)	(7.35)
April DD1029	0.004027*	0.004027*	0.002116	0.002116
	(1.67)	(1.75)	(0.58)	(0.61)
April DD30+	0.000000	0.000000	0.000000	0.000000
	(.)	(.)	(.)	(.)
April PDSI_W	-0.319826***	-0.318352***	-0.442950***	-0.445302***
	(-7.17)	(-7.40)	(-7.51)	(-7.80)
April PDSI_D	-0.302119***	-0.302575***	-0.466968***	-0.466240***
	(-5.86)	(-6.17)	(-6.10)	(-6.39)
May DD10	0.020569***	0.020713***	-0.002130	-0.002359
	(4.12)	(4.30)	(-0.27)	(-0.31)
May DD1029	0.000386	0.000357	0.001581	0.001628
	(0.22)	(0.21)	(0.71)	(0.77)
May DD30+	4.526198	4.466291	1.1e+01	1.1e+01
	(0.63)	(0.65)	(1.37)	(1.44)
May PDSI_W	0.694588***	0.693327***	0.981236***	0.983247***
	(13.52)	(14.16)	(14.59)	(15.26)
May PDSI_D	0.117143**	0.117451**	0.172789**	0.172297**
	(2.10)	(2.21)	(2.30)	(2.41)
No. of Obs.	7752	7752	7752	7752
Hansen J Statistic		51.002		49.395
Hansen J P-value		0.0001		0.0002
Kleibergen-Paap rk Wald F Statistic		45.053		45.053
Kleibergen-Paap rk LM Statistic		101.153		101.153
Adj R-squared	0.2812	0.0659	0.3082	0.0924

Notes: All columns include county and year fixed effects; Errors are clustered by county; t statistics are shown in parenthesis; * p<0.1, ** p<0.05, *** p<0.01.

Table A.5. Robustness check: Regression results using April to June weather variables

Independent Variables	Dependent Variable: PP-LCR		Dependent Variable: PP-LAR	
	Linear FE	Lewbel IV	Linear FE	Lewbel IV
CC	-0.026322***	-0.023371	-0.041561***	-0.024746
	(-3.34)	(-1.57)	(-3.46)	(-1.19)
April-Jun DD10	0.010926***	0.010956***	0.021507***	0.021675***
	(5.90)	(6.16)	(8.49)	(8.86)
April-Jun DD1029	0.004031***	0.004051***	0.007195***	0.007308***
	(3.51)	(3.72)	(4.54)	(4.84)
April-Jun DD30+	-0.286624***	-0.289022***	-0.548882***	-0.562552***
	(-5.76)	(-5.83)	(-6.62)	(-6.85)
April-Jun PDSI_W	0.454486***	0.454349***	0.640932***	0.640151***
	(10.65)	(11.11)	(12.58)	(13.12)
April-Jun PDSI_D	-0.127379***	-0.127508***	-0.243606***	-0.244339***
	(-3.32)	(-3.47)	(-4.53)	(-4.74)
No. of Obs.	7752	7752	7752	7752
Hansen J Statistic		38.103		33.789
Hansen J P-value		0.0009		0.0036
Kleibergen-Paap rk Wald F Statistic		51.510		51.510
Kleibergen-Paap rk LM Statistic		82.334		82.334
Adj R-squared	0.2781	0.0618	0.3102	0.0948

Notes: All columns include county and year fixed effects; Errors are clustered by county; t statistics are shown in parenthesis; * p<0.1, ** p<0.05, *** p<0.01.

Table A.6. Robustness check: Regression results with separate April, May, and June weather variables

Independent Variables	Dependent Variable: PP-LCR		Dependent Variable: PP-LAR	
	Linear FE	Lewbel IV	Linear FE	Lewbel IV
CC	-0.022374***	-0.025973**	-0.035453***	-0.029060
	(-2.90)	(-1.98)	(-3.02)	(-1.62)
April DD10	0.006613***	0.006575***	0.019484***	0.019551***
	(3.58)	(3.71)	(7.40)	(7.70)
April DD1029	0.002730	0.002707	-0.000074	-0.000032
	(1.26)	(1.30)	(-0.02)	(-0.01)
April DD30+	0.000000	0.000000	0.000000	0.000000
	(.)	(.)	(.)	(.)
April PDSI_W	-0.391996***	-0.390958***	-0.542951***	-0.544795***
	(-8.13)	(-8.45)	(-8.50)	(-8.86)
April PDSI_D	-0.381706***	-0.381251***	-0.621514***	-0.622321***
	(-7.01)	(-7.30)	(-7.44)	(-7.77)
May DD10	0.032593***	0.032607***	0.017474**	0.017449**
	(6.10)	(6.38)	(2.04)	(2.13)
May DD1029	-0.000412	-0.000439	0.000533	0.000580
	(-0.22)	(-0.25)	(0.23)	(0.27)
May DD30+	6.026419	5.976574	1.3e+01	1.3e+01
	(0.77)	(0.80)	(1.30)	(1.36)
May PDSI_W	0.306382***	0.305643***	0.438128***	0.439442***
	(4.93)	(5.19)	(5.70)	(6.03)
May PDSI_D	0.111477*	0.110971*	0.281145***	0.282044***
	(1.67)	(1.73)	(2.91)	(3.05)
Jun DD10	0.176692	0.177310	0.195759	0.194662
	(1.32)	(1.38)	(1.34)	(1.40)
Jun DD1029	0.013359***	0.013337***	0.022007***	0.022047***
	(6.27)	(6.59)	(7.69)	(8.11)
Jun DD30+	-0.321398***	-0.318550***	-0.521443***	-0.526500***
	(-6.16)	(-6.22)	(-6.27)	(-6.46)
Jun PDSI_W	0.492658***	0.492559***	0.681892***	0.682068***
	(8.27)	(8.65)	(8.85)	(9.26)
Jun PDSI_D	0.085559*	0.085862*	0.015597	0.015058
	(1.84)	(1.94)	(0.23)	(0.23)
No. of Obs.	7752	7752	7752	7752
Hansen J Statistic		51.002		49.395
Hansen J P-value		0.0001		0.0002
Kleibergen-Paap rk Wald F Statistic		45.510		45.510
Kleibergen-Paap rk LM Statistic		114.65		114.65

Adj R-squared	0.3019	0.0927	0.3308	0.1221

Notes: All columns include county and year fixed effects; Errors are clustered by county; t statistics are shown in parenthesis; * p<0.1, ** p<0.05, *** p<0.01.

Table A.7. Robustness check: Regression results using a time trend

Independent Variables	Dependent Variable: PP-LCR		Dependent Variable: PP-LAR	
	Linear FE	Lewbel IV	Linear FE	Lewbel IV
CC	-0.0670***	-0.2043***	-0.1092***	-0.2959***
	(-7.46)	(-6.80)	(-7.82)	(-7.03)
April-May DD10	0.0009	0.0006	0.0023	0.0019
	(0.71)	(0.48)	(1.35)	(1.15)
April-May DD1029	0.0013**	0.0006	0.0027***	0.0018**
	(2.46)	(1.02)	(3.88)	(2.29)
April-May DD30+	-7.2973	-6.8260	-7.8990	-7.2580
	(-0.62)	(-0.62)	(-0.54)	(-0.54)
April-May PDSI_W	0.3313***	0.3308***	0.4375***	0.4369***
	(8.65)	(8.85)	(9.76)	(9.94)
April-May PDSI_D	-0.2368***	-0.2107***	-0.4170***	-0.3815***
	(-6.18)	(-5.66)	(-7.67)	(-7.13)
Year	0.0459***	0.0734***	0.0655***	0.1029***
	(5.53)	(6.59)	(5.29)	(6.13)
No. of Obs.	7752	7752	7752	7752
Hansen J Statistic		30.639		39.078
Hansen J P-value		0.0000		0.0000
Kleibergen-Paap rk Wald F Statistic		42.544		42.544
Kleibergen-Paap rk LM Statistic		57.134		57.134
Adj R-squared	0.2207	-0.0581	0.2298	-0.0520

Notes: All columns include county and year fixed effects; Errors are clustered by county; t statistics are shown in parenthesis; * p<0.1, ** p<0.05, *** p<0.01.

Table A.8. Robustness check: Regression results using precipitation as a measure of moisture levels

Independent Variables	Dependent Variable: PP-LCR		Dependent Variable: PP-LAR	
	Linear FE	Lewbel IV	Linear FE	Lewbel IV
CC	-0.023163***	-0.012639	-0.041629***	-0.015593
	(-2.79)	(-0.85)	(-3.27)	(-0.76)
April-May DD10	0.010582***	0.010660***	0.019836***	0.020027***
	(5.73)	(6.02)	(7.70)	(8.11)
April-May DD1029	-0.001096	-0.001066	-0.001214	-0.001140
	(-0.89)	(-0.91)	(-0.68)	(-0.67)
April-May DD30+	0.964296	1.026553	6.249220	6.403239
	(0.10)	(0.12)	(0.65)	(0.70)
April-May Prec	0.005393**	0.005329**	0.007675***	0.007516***
	(2.35)	(2.44)	(2.70)	(2.78)
April-May Prec_sq	0.000014	0.000014	0.000016	0.000016
	(1.42)	(1.53)	(1.36)	(1.50)
No. of Obs.	7752	7752	7752	7752
Hansen J Statistic		50.425		59.426
Hansen J P-value		0.0000		0.0000
Kleibergen-Paap rk Wald F Statistic		47.808		47.808
Kleibergen-Paap rk LM Statistic		81.310		81.310
Adj R-squared	0.2806	0.0649	0.3012	0.0826

Notes: All columns include county and year fixed effects; Errors are clustered by county; t statistics are shown in parenthesis; * p<0.1, ** p<0.05, *** p<0.01.

Table A.9. Robustness check: Regression results using kinky least squares regression with PP-LCR as a dependent variable

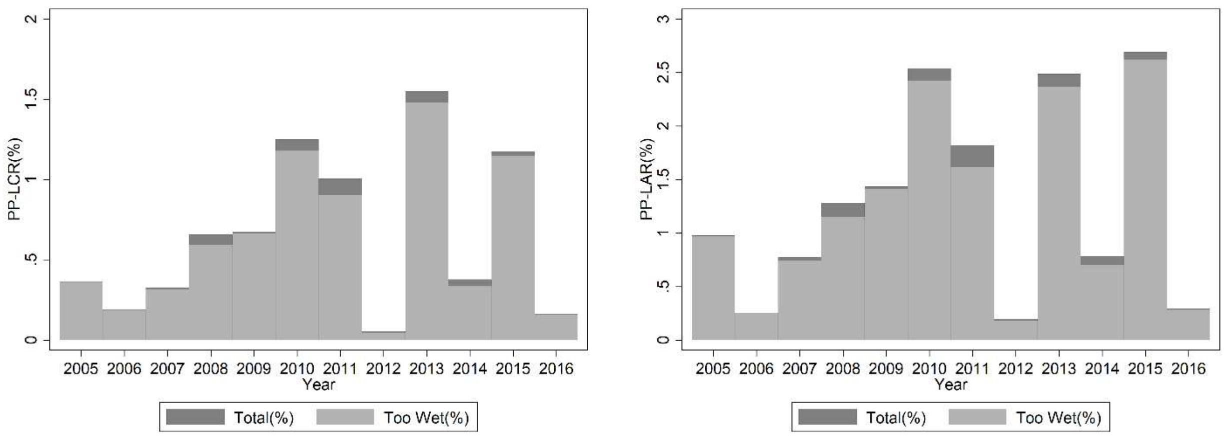
Independent Variables	Dependent Variable: PP-LCR			
	Assumed correlation between CC and county-time varying unobservables			
	0.1	0.2	0.3	0.4
CC	-0.1335***	-0.2399***	-0.3633***	-0.5228**
	(-3.89)	(-3.21)	(-2.68)	(-2.10)
April-May DD10	0.0088***	0.0081***	0.0074***	0.0064***
	(6.79)	(5.80)	(4.49)	(2.90)
April-May DD1029	-0.0002	-0.0005	-0.0009	-0.0014
	(-0.15)	(-0.43)	(-0.68)	(-0.88)
April-May DD30+	-3.9848	-4.6949	-5.5190	-6.5838
	(-0.30)	(-0.34)	(-0.37)	(-0.39)
April-May PDSI_W	0.3819***	0.3872***	0.3933***	0.4012***
	(14.65)	(14.24)	(13.32)	(11.70)
April-May PDSI_D	-0.1552***	-0.1611***	-0.1679***	-0.1768***
	(-3.07)	(-3.07)	(-2.98)	(-2.78)
No. of Obs.	7752	7752	7752	7752

Notes: All columns include county and year fixed effects; Errors are clustered by county; t statistics are shown in parenthesis; * p<0.1, ** p<0.05, *** p<0.01.

Table A.10. Robustness check: Regression results using kinky least squares regression with PP-LAR as a dependent variable

Independent Variables	Dependent Variable: PP-LAR			
	Assumed correlation between CC and county-time varying unobservables			
	0.1	0.2	0.3	0.4
CC	-0.1952***	-0.3440***	-0.5166***	-0.7397**
	(-4.07)	(-3.30)	(-2.73)	(-2.13)
April-May DD10	0.0177***	0.0167***	0.0157***	0.0143***
	(9.77)	(8.56)	(6.84)	(4.66)
April-May DD1029	0.0002	-0.0002	-0.0008	-0.0015
	(0.15)	(-0.14)	(-0.42)	(-0.66)
April-May DD30+	-0.6666	-1.6598	-2.8124	-4.3017
	(-0.04)	(-0.08)	(-0.13)	(-0.18)
April-May PDSI_W	0.5438***	0.5512***	0.5597***	0.5708***
	(14.92)	(14.49)	(13.56)	(11.90)
April-May PDSI_D	-0.2600***	-0.2683***	-0.2779***	-0.2903***
	(-3.68)	(-3.65)	(-3.53)	(-3.26)
No. of Obs.	7752	7752	7752	7752

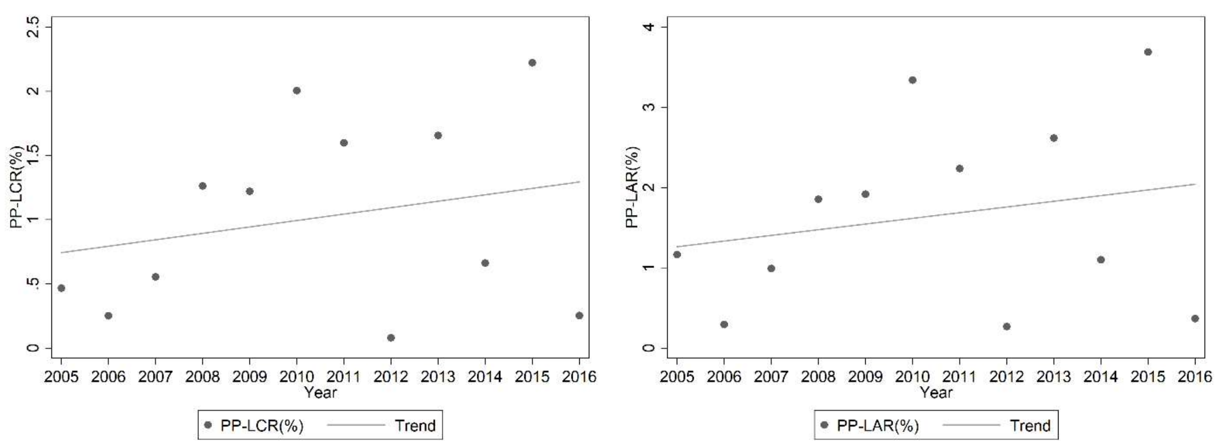
Notes: All columns include county and year fixed effects; Errors are clustered by county; t statistics are shown in parenthesis; * p<0.1, ** p<0.05, *** p<0.01.



(a) PP-related Loss Cost Ratio (PP-LCR, %) (b) PP-related Loss Acres Ratio (PP-LAR, %)

Figure A.1. Yearly prevented planting losses caused by "too wet" conditions (all crops), 2005-2016.

Note: "Too Wet" PP-LCR="Too Wet" PP-Indemnities/Total liabilities; "Too Wet" PP-LAR="Too Wet" Indemnified PP-Acres/Net insured acres



(a) PP-related Loss Cost Ratio (PP-LCR, %) (b) PP-related Loss Acres Ratio (PP-LAR, %)

Figure A.2. Average yearly prevented planting losses and their linear trend (all crops), 2005-2016

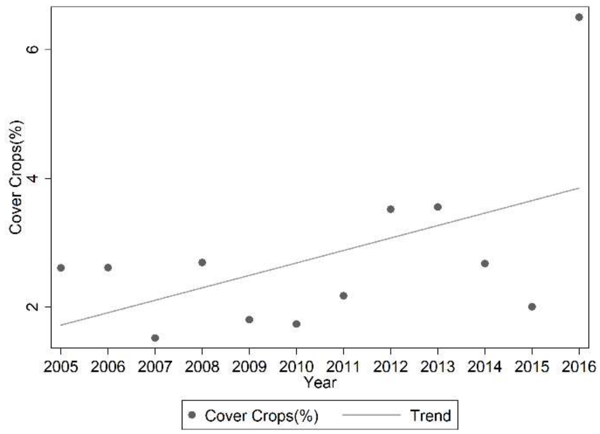


Figure A.3. Average yearly cover crop adoption (all crops, %), 2005-2016

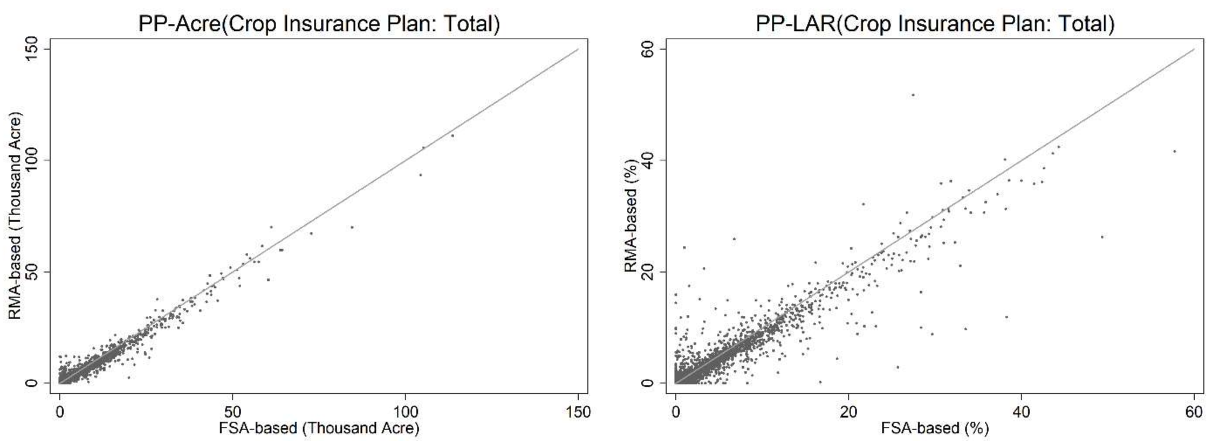


Figure A.4. Comparison of prevented planting measures from RMA vs. from FSA

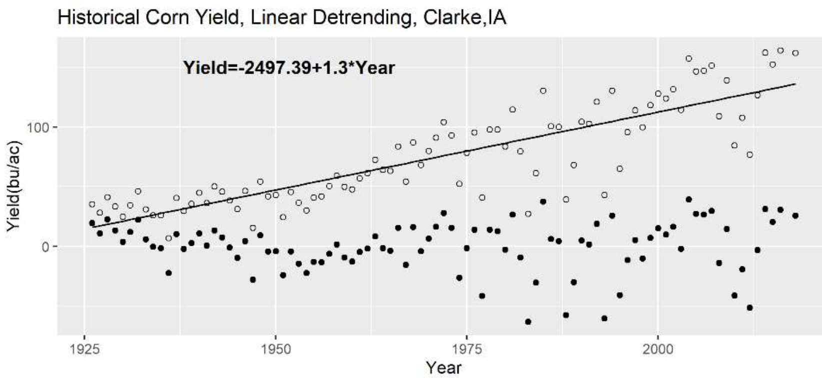
Appendix B

Appendix for Chapter 2

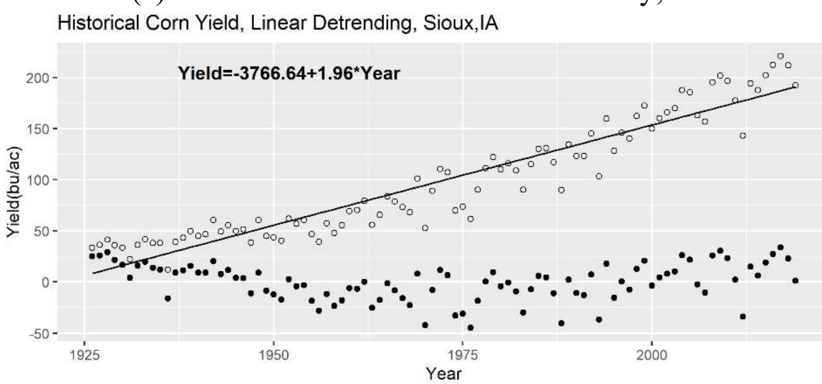
Detrended Corn Yield

In our analysis, linear detrending is selected among detrending methods (e.g., linear, quadratic, and cubic). Each county has a different slope and intercepts: the average slope is 1.46. For example, Figure B.1 shows detrend results of two different Iowa counties. Circles (hollow dots) represent actual corn yield and line represents linear trend based on ordinary least squares (OLS) regression. Solid dots show detrended yields (Residuals) which is considered as “observed” yield in our analysis of yield distribution.

Figure B.2 represents how residuals change over time. The variation of detrended yield has been increased over time. As shown in Figure B.3, the detrended yield has a strong temporal correlation and spatial correlation. In panel (a) of Figure B.3, the blue dotted line gives the values beyond which the temporal autocorrelations are statistically significantly different from zero. We also conduct a global spatial autocorrelation test of detrended yield data using Moran’s I (Moran, 1948). A positive value of Moran’s I indicates that detrended yields tend to cluster spatially (i.e., high detrended yields cluster near other high detrended yields; low detrended yields cluster near other low detrended yields). As illustrated in panel (a) of Figure B.3, Moran’s I is positive and statistically significant at less than the 0.01% level in each year of the data.



(a) Historical Corn Yield: Clarke County, IA



(b) Historical Corn Yield: Sioux County, IA

Figure B.1. Linear Detrending

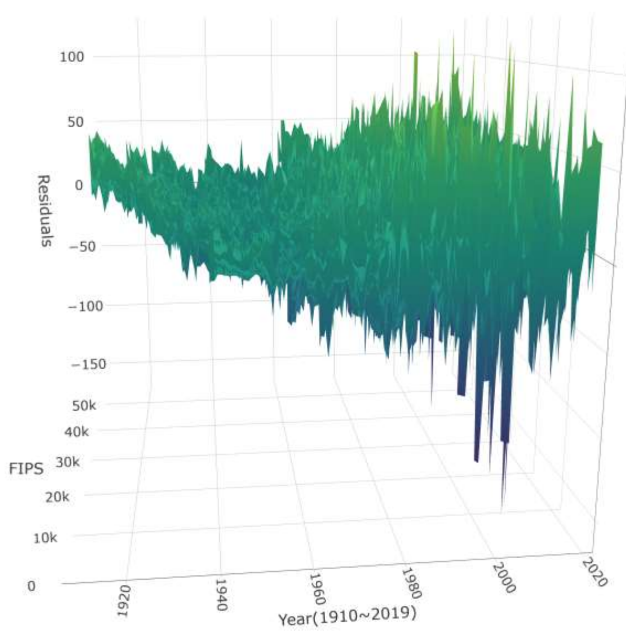
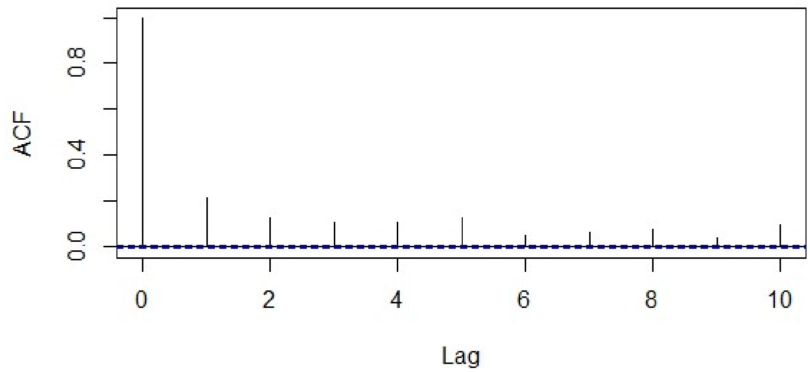
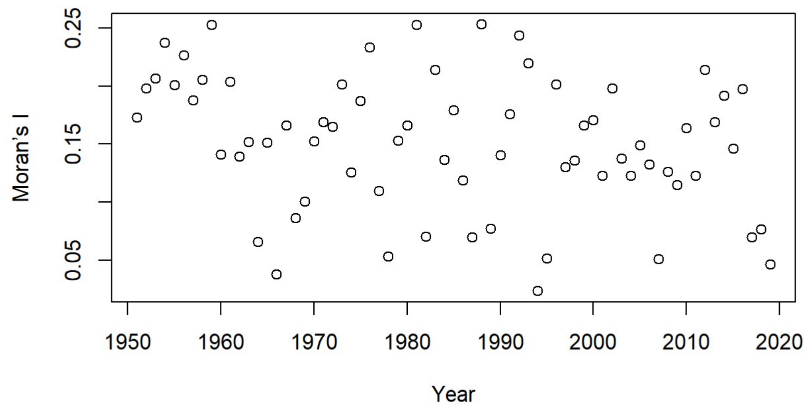


Figure B.2. Detrended Corn Yield, 1910~2019



(a) Temporal correlation



(b) Spatial correlation (Moran's I)

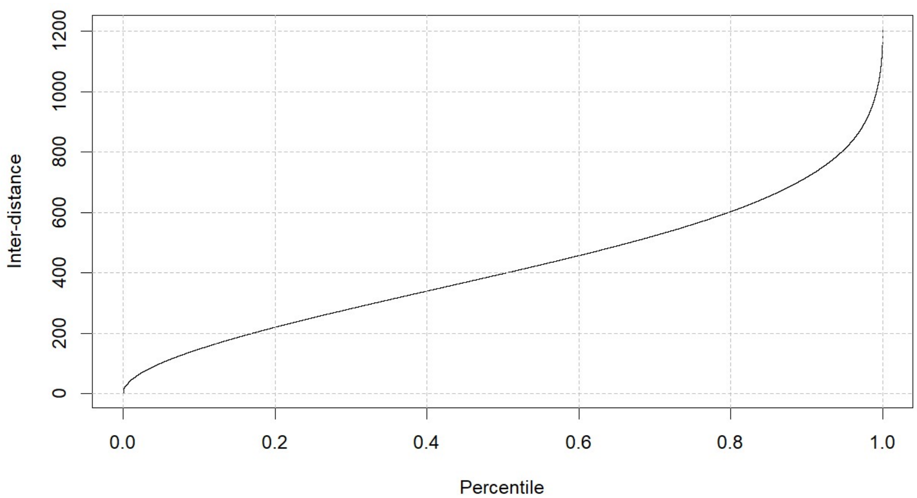
Figure B.3. Correlation, Detrended corn yields in the Corn Belt (919 counties)

Appendix C

Appendix for Chapter 2

Visualizing percentile of inter-distance among counties in the Corn Belt

Figure C.1 shows percentiles of inter-distances among centroid of counties in the Corn Belt.



Percentile	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
Distance	100.85	147.39	185.48	219.69	251.17	281.57	310.76	339.55	368.38	397.13

Figure C.1. Percentiles of inter-distances among counties in the Corn Belt

Appendix D

Appendix for Chapter 2

Spatial Blocking

In our strategy for blocking, we divided our study region (corn production in the Corn Belt, Figure A2.5) into a checkerboard pattern with a pre-specified block size (300.8 miles block size). The plotted block size is based on the median of the spatial autocorrelation ranges using a 69-layered corn yield map. By assessing variability between all possible pairs of counties, the effective spatial autocorrelation range is measured which means two yield observations over 300.8 miles away are independent. Based on randomly assigned 10 blocks, train and test groups of observation are clustered to conduct out-of-sample performance (Figure D.1).

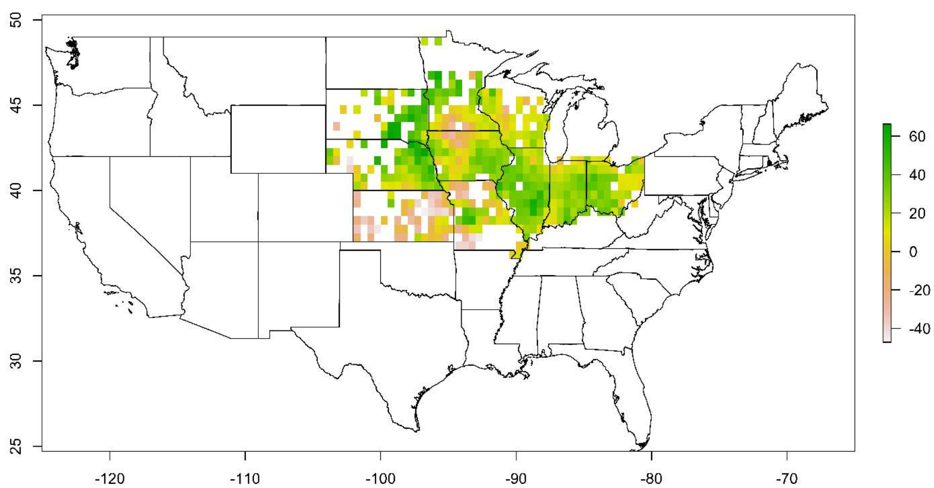


Figure D.1. Detrended Corn Yield in the Corn Belt, 2018(bu/ac)

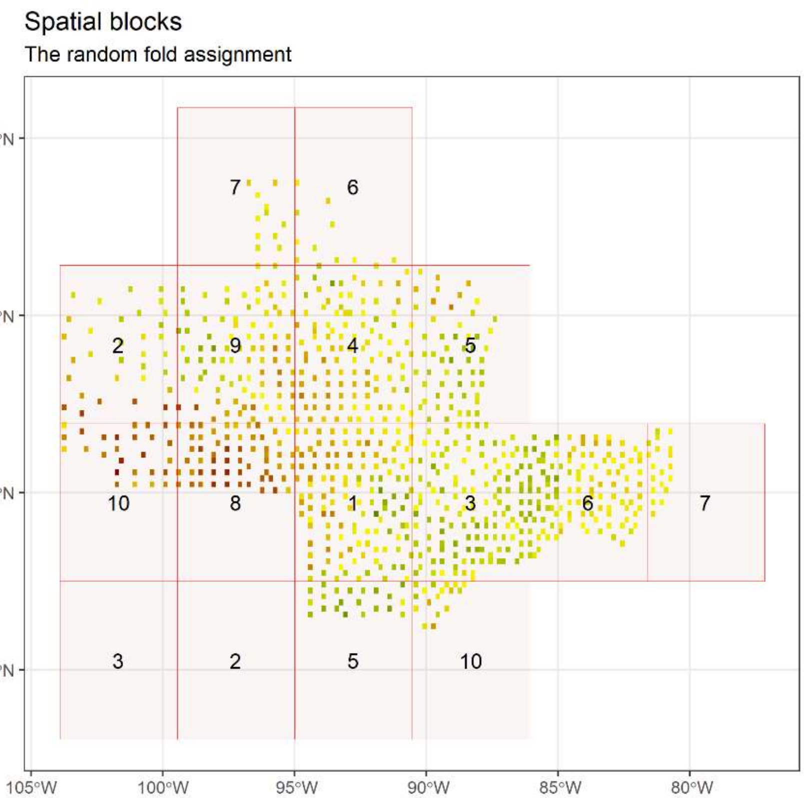


Figure D.2. Spatial blocks using (Block size: 496126 meters, 308 miles)