THREE ESSAYS ON RISK MANAGEMENT AND
IRRIGATION WATER DEMAND IN AGRICULTURE

By

Pin Lu

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Agricultural, Food, and Resource Economics—Doctor of Philosophy

2022

ABSTRACT

Both extensive (share of insured acres in total insurable acres) and intensive (coverage level choice) margin participation rates in the U.S. crop insurance program have increased due to generous subsidies. On a national scale, this program has been well rated to satisfy the actuarial fairness requirement by USDA Risk Management Agency. However, sizeable spatial heterogeneity remains across the Great Plains and Corn Belt regions. If subsidies were to be reduced in the future because of financial constraints, such heterogeneity might be detrimental to the sustainability of the crop insurance program. A central theme of this dissertation is to investigate how farmers make participation decisions when risk factors exist. In a separate but related line of work, this dissertation also explores the irrigation water usage in the Great Lakes region because farmers' irrigation behavior reflects their risk preferences and impacts their incentives for enrolling in the program. The dissertation consists of three essays on farmers' decisions regarding premium mispricing, basis risk, and irrigation water usage.

The first essay proposes a novel resampling procedure to estimate farm-level actuarially fair premiums. The resampling procedure mainly contains two parts: (i) semi-parametric quantile regression; and (ii) rejection method. Many previous studies explore whether county-level mispricing exists based on the historical loss ratio records. However, we can identify farm-level mispricing by imputing actuarially fair premiums based on historical yield records, consistent with theory. We find that farmers with lower land quality cropland paid fewer premiums than they should, but a contrary case happens for farmers with higher land quality cropland. Empirical evidence shows farmers may be more concerned about mispricing than subsidy transfer. Regression results support a conclusion that such farm-level mispricing deters farmers' crop insurance demand. Our analysis sheds light on the policy-making that: (i) mispricing may be a

substitution of subsidy so mitigating mispricing can maintain high participation while saving subsidies; and (ii) imputation of premiums based on historical yield records can apply.

The second essay focuses on the impact of basis risk on participation rates in the U.S. crop insurance program. In recent years, basis risk has been increasingly recognized as an essential driver for deterring insurance uptake. Most research concentrates on index insurance contracts; however, few investigate the effect of mismatch between cash and futures markets on farmers' insurance decisions. We first build a conceptual model to show farmers' acreage response to basis risk within the expected utility framework. Next, we apply the Fractional Probit with Control Function for the empirical analysis and find that the effects of basis risk on participation rates are significantly negative for nearly all insurance contracts. Our analysis implies that: (i) to remove basis risk, revision for revenue contract may be considered; (ii) subsidy structure may be adjusted to be consistent with the underlying basis risk.

The third essay investigates irrigation water usage in the Great Lakes region. Although the water conservation policy was implemented, there has been an upward trend in irrigation water demand from 2003 to 2018, including irrigated acres and total water usage. We employ firm-level irrigation data to examine what factors impact farmers' response to irrigation water usage. We find that: (i) price elasticities vary significantly according to model specifications and water costs; (ii) demand at both extensive (irrigated acres) and intensive (water application per acre) margins is input price inelastic; and (iii) price elasticities are homogeneous across crops but heterogeneous across states. For the policy-making, if there is a 10% tax on irrigation water cost, total water usage decreases by about 4% for corn and soybean, respectively.

I dedicate this dissertation to my parents, Yiming Lu and Xiaozhi Zhang.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

INTRODUCTION

Understanding farmers' decision-making in the crop insurance program is crucial because the federal government has provided generous subsidies for the development of the program. The dual goals of the federal government are higher crop insurance participation and coverage levels (USGAO 2015). However, although participation rates have been boosted over twenty years, significant regional disparities still exist in the Great Plains and Corn Belt regions (Babcock et al. 2004; Glauber 2004; Woodard et al. 2012; Chen et al. 2020). Suppose subsidies were to be reduced in the future due to financial constraints. In that case, such heterogeneity might be detrimental to the sustainability of the crop insurance program because farmers with lower risks might opt out of this program. One potential consequence is that the insurance pool will become very risky, and indemnity will be far beyond premiums. This dissertation explores how farmers make insurance purchasing decisions when yield or price risks exist. Our analysis might shed light on policymaking for premium rate-setting, contract design, and irrigation water restriction.

The first essay explores whether systematic mispricing exists in the yield contract and how farmers respond to such a bias when choosing among multiple coverage level choices. The motivation is that many studies investigate farmers' coverage choices, assuming farmers' premiums (including subsidies) are actuarially fair (Du et al. 2017; Jensen et al. 2018). Although some papers contribute to a deep understanding of rating methods of USDA Risk Management Agency (Goodwin 1993; Babcock et al. 2004; Woodard et al. 2011), to our best knowledge, there is no research investigating whether farmer's premiums are actuarially fair based on yield records. Therefore, one contribution of this study is that we propose a novel resampling procedure to estimate farm-level actuarially fair premiums at all coverage choices (see, for example, Chen and Yu 2016; Price et al. 2019). With the estimated premiums, we can identify

the mispricing for each coverage level choice for each farm. Unlike many previous studies which examine the mispricing based on loss ratio, our analysis is based on actual yields, consistent with theory. After comparing premiums charged by USDA Risk Management Agency and the estimated actuarially fair premiums, we find that a farm with lower (higher) quality cropland in a county has a better (worse)-than-actuarially-fair premium, which implies that a farm with lower (higher) quality cropland paid fewer (more) than they should. Regression results show that such mispricing deters farmers' crop insurance demand (i.e., coverage level choices). Our analysis shows that the adjustment of mispricing might be a supplement of federal subsidies.

The second essay investigates how farmers respond to the basis risk when participating in the crop insurance program. Many previous studies have explored whether basis risk impedes insurance uptake for index contracts (Doherty and Richter 2002; Deng et al. 2007; Cole et al. 2014; Cai et al.2020; Ohashi 2022). The basis risk in index contracts is the imperfect association between experienced losses and indemnification based on index values (Jensen et al. 2018). This essay investigates whether the mismatch between cash and futures markets affects farmers' insurance decisions. The reason is that the crop revenue contract has been the most popular product offered by the U.S. Federal Crop Insurance Corporation. The mismatch between cash and futures prices is inherent in revenue contracts because of contract designs. As in the literature, we define basis as the difference between cash and futures prices and basis risk as basis variation. We develop a conceptual model within the expected utility framework to identify farmers' acreage response to basis risk. With elevator-level basis data, we apply the Fractional Probit with Control Function and find that basis risk deters farmers' insurance demand for nearly all contracts. Our analysis implies that mitigating basis risk by revising revenue contract designs might be an option to maintain high participation rates while reducing the subsidies.

In the third essay, we turn to investigate irrigation water usage in the Great Lakes region because there has been an upward trend from 2003 to 2018. Such increased water withdrawal induces a concern regarding the future's water scarcity, especially in the context that current water restriction policies are inefficient. With the farm-level irrigation data, this essay constructs two water costs (average energy cost and marginal irrigation cost) to estimate various price elasticities of irrigation water usage. The marginal irrigation cost is widespread in studies for U.S. western states (Casewell and Zilberman 1986; Gonzalez-Alvarez et al. 2006; Mieno and Brozović; 2017), but some current research, such as Ito (2014), finds that consumers are more concerned about average cost than marginal cost. Therefore, we extend Kornelis and Norris (2020) with marginal irrigation cost, error tolerances, model specifications, and state-specific price elasticities. Our findings are: (i) AEC and MIC have similar performance for extensive (irrigated acres) and intensive (water application per acre) margins, although price elasticities based on MIC are slightly smaller; (ii) for crop-specific values, price elasticities based on MIC are sensitive to error tolerances and might be underestimated; (iii) demand at both extensive and intensive margins are input price inelastic; (iv) price elasticities are homogeneous across crops but heterogeneous across states; and (v) If there is an arbitrary 10% tax on irrigation water cost, total water usage (acre-inch) will decrease by 4% for corn and soybean, respectively.

CHAPTER 1 Crop Insurance Rate Making, Land Quality, and Adverse Selection

Abstract

Over the past two decades, the U.S. federal government has devoted considerable efforts to encouraging participation rates in the crop insurance program. However, many farmers within a county still feel that their premium rates are too high even after large subsidies have been provided. This paper proposes a novel resampling procedure based on large volumes of unit-level yield data to estimate actuarially fair premiums at all coverage level choices, consistent with theory. By comparing RMA's premiums with our estimated actuarially fair premiums, we find a systematic mispricing situation in the U.S. crop insurance program. That is, farmers with higher quality cropland have worse-than-actuarially-fair premiums, implying that farmers owning good land quality pay more premiums than they should. We find that farmers might be more concerned about mispricing than subsidy transfer. The regression results show that such mispricing deters farmers' crop insurance demand. Therefore, the adjustment of mispricing might be a supplement to federal subsidies.

Introduction

The U.S. federal crop insurance was authorized in the 1930s, and the Federal Crop Insurance Corporation (FCIC) was created in 1938 to carry out the whole program. The crop insurance program was experimental for a long time, and participation rates were low. After the 2000 Agricultural Risk Protection Act (ARPA), the number of insurance products and participation rates have been boosted due to increased premium subsidies. The increased individual-level loss experience provides crucial information for future loss expectations. USDA Risk Management Agency (RMA) relies on historical loss experience data to evaluate underlying yield risks, aiming to obtain actuarially sound premium rates equal to expected indemnities. Actuarial fairness is required for this program, and the loss ratio equal to 1.0 is the target in the 2008 Farm Bill. Over the past two decades, the crop insurance program has been well-rated nationally, indicating that the total premiums (including subsidies) equal the total indemnities. However, significant regional disparities of participation rates exist while the aggregate actuarial fairness has improved (Babcock et al. 2004; Glauber 2004; Sherrick et al. 2004; Woodard et al. 2012; Chen et al. 2020).

Corn Extensive | Corn Intensive

(0.93,1.25] (0.88,0.93] (0.83,0.88]
(0.74,0.83] [0.17,0.74] No data

(0.80,0.86] (0.77,0.80] (0.74,0.77]
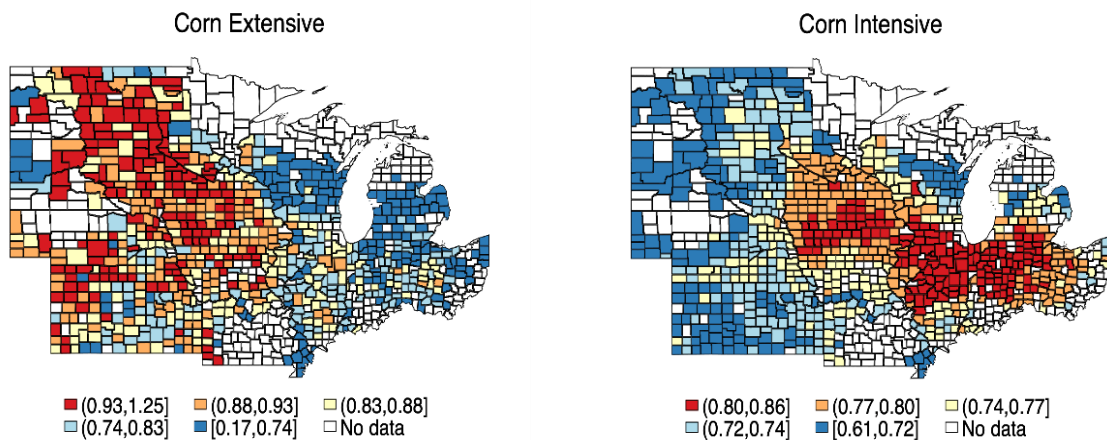(0.72,0.74] [0.61,0.72] No data

Figure 1.1 The Extensive and Intensive Margin Participation Rates for Corn Over 1989-2020. Note: The extensive margin is defined as the ratio of insured acres to total insurable acres; the intensive margin is defined as the acreage-weighted coverage level choices.

Actuarial fairness is quite crucial because it matters to the program's sustainability. Without the generous subsidies, producers whose expected indemnities exceed the premiums costs are more likely to purchase insurance; those whose costs exceed their expected indemnities are less likely to purchase (Glauber 2004). After generous subsidies were provided, farmers enrolled in the program if the expected indemnities exceeded their self-paid premiums. In other words, generous subsidies can cover the potential mispricing and support high participation rates. However, an adverse selection problem might exist if subsidies were reduced and farm-level premiums were not actuarially sound. Farmers with higher expected yields or lower risks might opt out of the crop insurance program. Then the entire insurance pool becomes risky, and indemnities might exceed premiums, which will be detrimental to the development of the program (Skees and Reed 1986; U.S. General Accounting Office 1993; Coble and Knight 2002; Babcock et al. 2004; Glauber 2004). In addition, due to the increased indemnities, RMA might adjust the premium rating upward and crowd more farmers with good land quality out of the insurance pool, further exacerbating the program (Coble et al. 2010; Price et al. 2019).

Many previous studies have explored multiple topics regarding USDA RMA's ratemaking approach. For example, there are the Actual Production History (APH) rating method (Goodwin 1993), assumption of constant rate relativities across coverage levels (Babcock et al. 2004), actuarial implications of the loss cost ratio (LCR) rate-making methodology (Woodard et al. 2011), and improvement of rating efficiency with integrating widely available high-resolution soil data (Woodard and Verteramo‐Chiu 2017). However, most research directly focuses on the historical premiums and indemnities because the targeted loss ratio (expected indemnity/ premium) is 1.0. One disadvantage of the USDA/RMA loss ratio approach is that expected indemnity is estimated based on the limited length of time for which historical experience is

available. However, in theory, expected indemnity should be estimated based on yield distribution. Therefore, we employ large volumes of farm-level actual yield records to estimate the actuarially fair premiums, with which systematic mispricing can be identified.

Therefore, this study aims to: (i) estimate farm-level actuarially fair premiums at all coverage choices; (ii) investigate whether actuarially unfairness exists; and (iii) examine the impact of the unfairness on farmers' coverage choices. The motivation to investigate mispricing is that many previous studies analyse farmers' choices assuming premiums are actuarially fair (LaFrance et al. 2002; Du et al. 2017; Jensen et al. 2018). However, the assumption of actuarial fairness might be too restrictive to obtain a proper conclusion.

Our analysis builds on and contributes to three primary works of literature. First, we propose a novel resampling procedure to have a representative subsample, with which the underlying risk might be estimated reasonably. Many methods regarding yield distributions have been applied in the literature, including parametric distribution (Babcock and Blackmer 1992; Borges and Thurman 1994; Babcock and Hennessy 1996; Coble et al. 1996; Sherric et al. 2004; Claassen and Just 2011), and nonparametric distribution (Goodwin and Ker 1998; Ker and Goodwin 2000; Ker and Coble 2003; Woodard et al. 2011; Zhu et al. 2011). Due to limited subsidies, participation rates in the crop insurance program were meager before 1999. Therefore, if we estimated yield distribution directly based on yield records, actuarially fair premiums might be underestimated because there was no disaster between 1999 and 2008, the time interval most observations lie in. The advantages of our proposed resampling procedure are that it: (i) effectively imputes missing observations before 1999; and (ii) reweights farms based on long-term county-level historical records. Our procedure mainly includes two parts: (i) semi-parametric quantile regression associated with penalized B-splines (see, for example, Chen and

Yu 2016); and (ii) the rejection-acceptance method.

Second, this study highlights whether systematic rate bias exists and how such bias impacts farmers' insurance purchasing decisions. After obtaining a representative subsample, we estimate county-level actuarially fair premiums and then adjust county-level to individual-level premiums at all coverage choices per RMA's rules (Coble et al. 2010; Price et al. 2019). Finally, there are RMA premiums and estimated actuarially fair premiums at all coverage level choices (see, for example, Du et al. 2017). We then construct a wedge variable, the ratio of RMA's premium to the estimated actuarially fair premium. Since the premiums are always positive, the wedge range is from zero to positive infinity, and the wedge equal to 1 indicates that RMA's premium is actuarially sound. Conversely, a wedge higher (lower) than 1 indicates farmers paid more (less) than they should. With farm-level RMA's and estimated actuarially fair premiums, we can identify that: (i) whether farms with better (worse) land quality have higher (lower) wedges; and (ii) whether counties with better (worse) land capability have higher (lower) wedges. More information about wedges (e.g., distribution) can help policymakers understand the potential actuarial unfairness, which might help improve the rate-setting procedure in the crop insurance program.

Third, our analysis provides insights into saving subsidies while maintaining high participation rates. Large subsidies have been provided to encourage a high participation rate and to ensure the sustainability of the crop insurance program. From 2007 to 2027, the Federal Crop Insurance Program has cost and will cost the federal government about $7.5 billion yearly (Rosa 2018a; Rosa 2018b). In this study, farmers may not be concerned about the subsidy transfer as in some previous studies (see, for example, Du et al. 2017). However, the empirical evidence shows that farmers might be concerned about actuarial fairness because the coverage level with the

lowest wedge is the most popular. Regression results robustly support the conclusion that mispricing adversely affects farmers' coverage level choice. Therefore, the mitigation of mispricing can supplement subsidy, which implies that the subsidy structure can be adjusted to increase farmers' crop insurance demand while providing lower subsidies (Babcock et al. 2004).

The rest of this paper is organized as follows. In section 2, we apply the standard expected utility framework to derive how actuarial fairness affects farmers' coverage level choices. Section 3 reports all data employed in this study. Section 4 shows empirical evidence for our hypotheses and regression results. Section 5 is the final part concluding with some discussions.

Basic Model

This study focuses on yield contracts because only yield risk is contained. On the other hand, revenue contracts include the price risk, making the analysis hard, although our main conclusion will not be changed. Let z denote the approved yield history (i.e., the institutional estimate of farm-level mean yield), which represents land quality on a given insured unit and is used to estimate insurance premium cost by RMA. Let φ denote the coverage level with $0 < \varphi < 1$, for a stochastic yield y on a given insured unit, the actuarially fair premium (expected payout) can be calculated as:

(1.1) $\quad \Gamma(\varphi, z) = E[\max(\varphi z - y, 0)] = \int_0^{\varphi z}(\varphi z - y)dF(y),$

where $E[\cdot]$ is the expectation operator; φz is guaranteed yield below which farmers will receive indemnities from insurers; $F(y)$ is a cumulative distribution function with the density function; $f(y)$ on $0 < y < +\infty$. Without loss of generality, we assume the crop price is equal to 1 and total liability is φz.[1] The response of actuarially fair premium to coverage level is

---

[1] In the empirical part, we assume the corn price is \$4/bu and the soybean price is \$10/bu.

$d\Gamma(\varphi, z)/d\varphi \equiv \Gamma_\varphi(\varphi, z) = zF(\varphi z)$.

We define a wedge factor as $w(\varphi) = APR(\varphi)/OPR(\varphi)$, where $APR(\varphi)$ is actual premium rate made by RMA at coverage level φ and $OPR(\varphi)$ is the estimated actuarially fair premium rate. [2] The wedge range is $(0, +\infty)$ because both $APR(\varphi)$ and $OPR(\varphi)$ are positive. Then, a farmer's actual premium payment (including subsidy) is $w(\varphi)\Gamma(\varphi, z)$, and the response to coverage level is $w_\varphi(\varphi)\Gamma(\varphi, z) + w(\varphi)\Gamma_\varphi(\varphi, z)$, where $w_\varphi(\varphi)$ and $\Gamma_\varphi(\varphi, z)$ are the appropriate function derivatives. Denote $s(\varphi)$ as the subsidy rate for coverage level φ with $0 < s(\varphi) < 1$. The subsidy's dollar value is $S(\varphi, z) = s(\varphi)w(\varphi)\Gamma(\varphi, z)$, which represents the portion undertaken by federal government. Finally, a farmer's self-paid premium is $\chi(\varphi, z) = [1 - s(\varphi)]w(\varphi)\Gamma(\varphi, z)$.

Denote $s_\varphi(\varphi)$ and $w_\varphi(\varphi)$ as the derivative of subsidy rate function and wedge function, respectively, then we have the responses of subsidy's dollar value $S(\varphi, z)$ and farmer's self-paid premium $\chi(\varphi, z)$ to coverage level φ as

(1.2a) $\quad S_\varphi(\varphi, z) = \left[\dfrac{s_\varphi(\varphi)}{s(\varphi)} + \dfrac{w_\varphi(\varphi)}{w(\varphi)} + \dfrac{zF(\varphi z)}{\Gamma(\varphi, z)} \right] \times S(\varphi, z)$;

(1.2b) $\quad \chi_\varphi(\varphi, z) = \overbrace{w_\varphi(\varphi)\Gamma(\varphi, z) + w(\varphi)zF(\varphi z)}^{\text{Wedge Effect}} - \overbrace{S_\varphi(\varphi, z)}^{\text{Subsidy Transfer Effect}}$.

When RMA premium rates are actuarially fair, then $w(\varphi) \equiv 1$ and $w_\varphi(\varphi) \equiv 0$, which indicates that the wedge effect vanishes. Then Eqn.(1.2a) becomes

(1.2a') $\quad \tilde{S}_\varphi(\varphi, z) = \left[\dfrac{s_\varphi(\varphi)}{s(\varphi)} + \dfrac{zF(\varphi z)}{\Gamma(\varphi, z)} \right] \times S(\varphi, z)$.

---

[2] In the literature, wedge is sometimes defined as the difference between actual premium payout and actuarially fair premium (i.e., expected indemnity) for an insurance product (see more in Deng et al. 2007). As in Du et al. (2017), the actual premium payout equals actuarially fair premium multiplied by (1 + loading factor). The wedge factor in this study is the same as (1 + loading factor).

Denote $U(\cdot)$ as the Bernoulli utility function for farmer's income π; denote $U_\pi(\cdot)$ and $U_{\pi\pi}(\cdot)$ as appropriate first and second derivative functions respectively with $U_\pi(\cdot) > 0 > U_{\pi\pi}(\cdot)$. Let ε represent other sources of individual farm's income with a cumulative distribution function $G(\varepsilon)$ on a compact set $[\underline{\varepsilon}, \overline{\varepsilon}]$. Then, if purchasing the yield insurance, the farmer's income is $\pi(\varphi, z, y, \varepsilon) \equiv \max(\varphi z, y) - C(\varphi, z, \overline{c}, \varepsilon)$, where $C(\varphi, z, \overline{c}, \varepsilon) \equiv \chi(\varphi, z) + \overline{c} - \varepsilon$, and $\overline{c}$ is fixed cost. Therefore, the farmer's expected utility is:

$$(1.3) \quad E[U(\pi)] = F(\varphi z) \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} U\big(\varphi z - C(\varphi, z, \overline{c}, \varepsilon)\big) dG(\varepsilon)$$

$$+ \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} \int_{\varphi z}^{+\infty} U\big(\pi(\varphi, z, \overline{c}, \varepsilon)\big) dF(y) dG(\varepsilon).$$

Suppose a farmer is a rational and risk-averse EU maximiser, the first derivative of $E[U(\cdot)]$, with regard to coverage level φ will be

$$(1.4) \quad \frac{\partial E[U(\cdot)]}{\partial \varphi} = \underbrace{zF(\varphi z) \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} U_\pi\big(\varphi z - C(\varphi, z, \overline{c}, \varepsilon)\big) dG(\varepsilon)}_{\geq 0}^{\text{Insurance Effect}} - \underbrace{\chi_\varphi(\varphi, z) E\big[U_\pi\big(\pi(\varphi, z, y, \varepsilon)\big)\big]}_{?}^{\text{Wedge Effect−Subsidy Transfer Effect}}$$

where $E[U_\pi(\varphi, z, y, \varepsilon)] = \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} \int_{\varphi z}^{+\infty} U_\pi(\pi(\varphi, z, y, \varepsilon)) dF(y) dG(\varepsilon) \geq 0$.

We define the first term of Eqn.(1.4) as Insurance Effect because both subsidy rate and wedge factor are not included. The second term of Eqn.(1.4) shows the difference between Wedge Effect and Subsidy Transfer Effect. If premium rates are actuarially fair, the second term of Eqn.(1.4) degenerates to the condition that only Subsidy Transfer Effect is included (see, for example, Du et al. 2017). If premium rates are not actuarially fair and $\chi_\varphi(\varphi, z) \leq 0$ (i.e., Case 2 of Figure 1.2), then $\partial E[U(\cdot)]/\partial \varphi > 0$ and a rational farmer will choose the highest coverage level because marginal utility increases as the coverage level increases. However, we know this is not true in the real world because it indicates that farmers can pay less premiums when they choose higher coverage levels. Therefore, we will assume $\chi_\varphi(\varphi, z) > 0$ (i.e., Case 1 of Figure

1.2), which implies that farmers' self-paid premium increases as the coverage level increases.
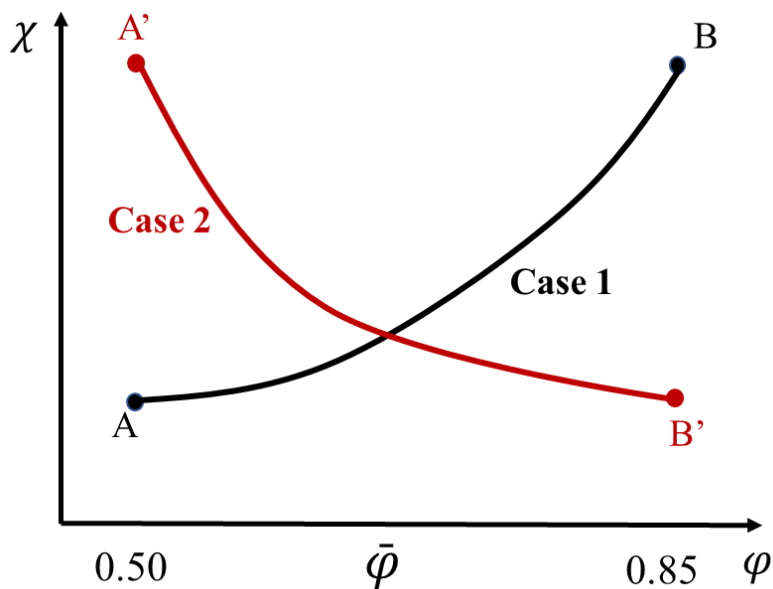
Figure 1.2 Possible Relationships Between Farmer's Self-Paid Function and Coverage Level.
Note: In case 1, $\chi_\varphi > 0$; in case 2, $\chi_\varphi < 0$. The farmer's self-paid premium is the premium rates after subsidy.

Remark 1. Given $\chi_\varphi(\varphi, z) > 0$, suppose $\chi_{\varphi\varphi}(\varphi, z) > 0$, if $\exists \bar{\varphi}$ with $0 < \bar{\varphi} < 1$, s.t $\chi_{\bar{\varphi}} =$

$zF(\varphi z) \int_{\underline{\varepsilon}}^{\bar{\varepsilon}} U_\pi(\varphi z - C(\cdot))dG(\varepsilon)/E[U_\pi(\pi(\cdot))]$, then

i) an interior solution exists whenever $\chi_{\varphi=0.50} < \chi_{\bar{\varphi}} < \chi_{\varphi=0.85}$;

ii) coverage level 0.50 will be chosen whenever $\chi_{\bar{\varphi}} < \chi_{\varphi=0.50}$;

iii) coverage level 0.85 will be chosen whenever $\chi_{\bar{\varphi}} > \chi_{\varphi=0.85}$.

Hypothesis 1 (H1). *Farmers' coverage choices were either clustered at one single coverage level or two adjacent levels.*

We now turn to investigate the effect of the wedge factor $w(\varphi)$ on the subsidy's dollar value $S(\varphi, z)$. The motivations are: (i) the subsidy transfer is considered as an incentive for

farmers to choose higher coverage level (see, for example, Du et al. 2017; Cai et al. 2020); (ii) as

in Eqn.(1.2a), the subsidy transfer is impacted by subsidy rate $s(\varphi)$ and wedge factor $w(\varphi)$.

Back to Eqn. (1.1), we obtain $zF(\varphi z)/\Gamma(\varphi, z) \geq 1/\varphi$. Then, Eqn.(1.2a) can be re-written as

$S_\varphi(\varphi, z) \overset{sign}{=} s_\varphi(\varphi)/s(\varphi) + w_\varphi(\varphi)/w(\varphi) + 1/\varphi$, which implies that the sign of $S_\varphi(\varphi, z)$ only

depends on subsidy rate, wedge factor, and their first derivative functions. Assume that

$S_\varphi(\varphi, z) > 0$ since the higher subsidies are used to offset the excessive premium charge and

make higher coverage levels more attractive (Babcock et al. 2004). Then, we have

      Remark 2. A sufficient condition for $S_\varphi(\varphi, z) > 0$ is

(1.5) $\frac{\varphi s_\varphi(\varphi)}{s(\varphi)} + \frac{\varphi w_\varphi(\varphi)}{w(\varphi)} \equiv \Delta^s(\varphi) + \Lambda^w(\varphi) > -1$

where $\Delta^s(\varphi)$ and $\Lambda^w(\varphi)$ are the elasticities of subsidy rate $s(\varphi)$ and the wedge factor $w(\varphi)$ with

respect to φ respectively.

Table 1.1 Premium Subsidies on Yield Contracts for BaU and OpU.

Coverage Level φ	CAT	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85
$s(\varphi)$	1	0.67	0.64	0.64	0.59	0.59	0.55	0.48	0.38

Source: The 2014 Farm Bill. BaU represents Basic Unit; OpU represent Optional Unit.

      As in Table 1.1, RMA subsidy rate $s(\varphi)$ decreases when coverage level φ increases,

which implies that $s_\varphi(\varphi) \leq 0$. Since $0 < \varphi < 1$ and $s(\varphi) > 0$, then $\Delta^s(\varphi) \leq 0$. Suppose

Eqn.(1.5) holds, the lower bound of $\Lambda^w(\varphi)$ will be $-1 - \Delta^s(\varphi)$. Since $w(\varphi) > 0$, we have: (i)

when $\Delta^s(\varphi) < -1$, $\Lambda^w(\varphi) > 0$ and $w_\varphi(\varphi) > 0$; and (ii) when $\Delta^s(\varphi) > -1$, the sign of $\Lambda^w(\varphi)$

is undecided, which implies that $w_\varphi(\varphi)$ may also be negative and positive. We may ask why

discussion for the above elasticities is important. This is because if premium is assumed to be

actuarily fair, then $w_\varphi(\varphi) \equiv 0$, which implies that the sufficient condition for $S_\varphi(\varphi, z) > 0$ is only $\Delta^s(\varphi) > -1$. Therefore, a confusing situation will exist if farmers do not choose highest coverage levels when $\Delta^s(\varphi) > -1$. Our analysis helps explain the confusing case by exploring the role of actuarial unfairness - $w(\varphi)$ and $w_\varphi(\varphi)$.

Hypothesis 2 (H2). *Farmers choose the highest coverage level, at which the sum of the elasticities of subsidy rate and wedge factor is larger than 1, i.e., $\Delta^s(\varphi) + \Lambda^w(\varphi) > -1$.*

The H2 is proposed based on a predetermined belief that farmers prefer higher subsidy transfer, i.e., $S_\varphi(\varphi, z) > 0$. If $\Delta^s(\varphi) + \Lambda^w(\varphi) \leq -1$, then $S_\varphi(\varphi, z) \leq 0$ implying that higher coverage level will not be attractive to farmers, which violates RMA's target. It is noteworthy that the subsidy rate is readily obtained from USDA/RMA. However, the wedge factor is only maintained by farmers if assuming farmers know their underlying production risks. As such, we may ask whether farmers are more concerned about the wedge information.

Hypothesis 3 (H3). *Farmers choose the coverage level at which the wedge (including subsidies) arrives at the minimum value.*

If H3 holds, there will be a crucial policy implication that the federal government may investigate the wedge conditions experimentally and adjust wedges at all coverage levels. For example, suppose the federal government's goal is still encouraging farmers to purchase the higher coverage levels. Then, wedges at higher coverage levels can be adjusted downwards to make higher coverage levels more attractive. If farmers can be aware of the wedge, as discussed above, we can further investigate the effect of the wedge on farmers' coverage level choices in different quality cropland.

Hypothesis 4 (H4). *Farmers with higher (lower) land quality pay more (fewer) premiums than they should. In other words, a county's higher (lower) quality cropland has a*

worse (better)-than-actuarially-fair premium rating.

In what follows, we will test our hypotheses based on the estimation of the wedge factor, which is the ratio of RMA premium to the estimated actuarially fair premium. The wedge values can be observed at all coverage levels. We will provide the empirical evidence to test H1-H4. Furthermore, to examine the effect of wedge factor on coverage level choice, we employ the Ordered Logit Model (OLM) for farm-level coverage choice; Ordinary Least Squares (OLS) and Weighted Least Square (WLS) for county-level coverage choices.

Data Description

We focus on APH yield contracts in this study because yield contracts only contain yield risk and crop prices are constant. [3] If revenue contracts are included, unnecessary modeling complications will also be introduced. When price risks are in a limited range, the analysis in this study can be extended to revenue contracts, although all conclusions are based on yield contracts.

Table 1.2 Definition of Variables.

	Variable	Description
Premium	APR	Premium rates (including subsidies) from RMA
	OPR	Estimated actuarially fair premium rates based on RMA actual yield data
Wedge	w_i	Farm-level wedge at each coverage level
	Δw_i	Farm-level wedge differential between higher and lower coverage levels
	w_c	County-level average wedge (acreage-weighted)
	Δw_c	County-level wedge differential between higher and lower coverage levels (acreage-weighted)
Land Capability	LCC	Percentage of Class I-II in Class I-VIII
Weather Determinants	$\bar{G}_c$	Average GDD over 1989-2007
	$\bar{S}_c$	Average SDD over 1989-2007

The targeted area is the twelve states in Midwest and Great Plain region (IL, IN, IA, KS,

---

[3] Insurance plan code is 90; abbreviation is APH; the name is Actual Production History.

MI, MN, MO, NE, ND, OH, SD and WI). Corn and soybean are selected for crop-specific

analysis. The Table 1.2 reports the definition of main variables, including premiums, wedge

factors, land capability and weather determinants. The Table 1.3 reports the statistics summary of

main variables.

Table 1.3 Summary of Main Variables.

	Variable	Obs.	Mean	Std.Dev.	Min	Max
Corn	$APR(\varphi = 0.65)$	94,401	22.27	13.85	6.83	174.66
	$OPR(\varphi = 0.65)$	94,401	15.42	14.75	2.31	282.94
	w_i	94,401	1.73	0.65	0.17	20.76
	Δw_i	94,401	0.06	0.17	-9.24	1.02
	w_c	818	1.72	0.51	0.43	4.35
	Δw_c	818	0.04	0.15	-0.91	0.47
Soybean	$APR(\varphi = 0.65)$	108,792	6.23	3.27	2.22	66.9
	$OPR(\varphi = 0.65)$	108,791	3.84	3.63	0.85	196.1
	w_i	108,791	2.07	0.99	0.14	19.18
	Δw_i	108,791	0.11	0.19	-7.12	1.95
	w_c	791	1.88	0.87	0.15	9.26
	Δw_c	791	0.095	0.16	-0.43	0.67
Land Capability	LCC	2,902	0.25	0.21	0.00006	0.93
Weather Determinants	$\bar{G}_c$	1,006	1248.7	191.1	186.6	1692.2
	$\bar{S}_c$	1,006	20.48	22.9	0	126.02

Estimated Actuarially Fair Premium Rates by RMA Rules

Denote $OPR_{i,c}(\varphi)$ as the estimated premium rate for coverage level φ for farm i in

county c; APH_i as Actual Production History reflecting the institutional estimate of mean yield

by USDA/RMA; ref yield$_c$ as county average yield; $-E_c$ as the exponent reflecting the

correlation between county and individual yields; $ULR_c(\varphi)$ as county unloaded rate estimated

from our proposed procedure, which is also the main independent variable in this study;

$fixed\ load_c$ as the load reflecting potential disaster and production failure; $cov\ diff(\varphi)$ as

coverage level differential used to adjust rates from the base 65% coverage to other coverage levels. Then our estimated farm-level premium rate at each coverage level φ is as:

$$(1.6) \quad OPR_{i,c}(\varphi) = \left(\frac{APH_i}{\text{ref yield}_c}\right)^{-E_c} \times ULR_c(\varphi) + fixed\ load_c \times cov\ diff(\varphi)$$

In Eqn.(1.6), our proposed procedure is used to estimate the county-level unloaded premium rate $ULR_c(\varphi)$ at the coverage level φ with $\varphi \in \{0.50, \dots, 0.85\}$. All other parameters are from USDA/RMA. It is noteworthy that in RMA's methodology, county-level unloaded premium rate ULR_c is only for a 65% coverage level (see Coble et al. 2010). Premium rates at other coverage levels are obtained by applying coverage level differentials $cov\ diff(\varphi)$. For example, in RMA's methodology, the premium at $\varphi = 0.75$ equals to $ULR(\varphi = 0.65) \times cov\ diff(\varphi = 0.75)$. Since our procedure can estimate the county-level premium rates at all coverage levels, coverage level differential will impact the farm-level premium rates only through the county-level fixed load $fixed\ load_c$.

Now we turn to introduce how $ULR_c(\varphi)$ is estimated from the 2008 USDA/RMA farm-level yield data. In this yield dataset, each farm has a specific APH and four-to-ten-year yield historical records. There are 18 yield types including actual yield, transitional yield, the exceptional transitional yield for a new producer, and simple average transitional yield for added land. In this study, we only employ the actual yield because other types are RMA's imputation, accounting for a small proportion in the whole sample (less than 20% in western Great Plains and less than 15% in Corn Belt). Finally, there are about 9 million yield records in 882 counties for corn and about 8.6 million records in 828 counties for soybean.

Our proposed procedure contains four parts: (i) *yield detrend*: all historical farm- and county-level actual yields are adjusted to the 2009 technological level; (ii) *semiparametric*

quantile regression imputation (SQRI): impute missing observations over 1970-1998 based on penalized B-splines and quantile regression approaches, because the most observations in the yield dataset lie between 1999 and 2008 and no extreme climate change happens in this 10-year interval; (iii) *rejection sampling*: a more representative sample with farm-level records can be obtained from long-term county-level yields; and (iv) *actuarially fair premium estimation*: impute $ULR_c(\varphi)$ at all coverage levels based on the univariate penalized B-spline method.

After obtaining $ULR_c(\varphi)$, we employ the 2009 USDA/RMA farm-level contract choice data to obtain farm-level premiums as in Eqn.(1.6). The primary Information in this dataset contains the farm's location information (state and county), coverage level chosen, the premium rate paid (including subsidies), insured acres, elected crop price, approved production history (APH), actual yield, and various parameters such as fixed load or coverage differential.

Figure 1.3 reports the coefficient of variation (i.e., standard deviation/mean value) of actual yields for corn in each Crop Reporting Region (CRD). The upper panel of Figure 1.3 shows the coefficient of variation for the original data; the lower panel shows that for the resampling data. We find that the pattern of coefficient of variation does not change after our resampling procedure, implying no systematic bias. Figure 1.4 reports county-level yield distribution after the resampling procedure. We find that the yield distributions in the main production area (e.g., IA, IL, IN) have the similar patterns. However, the distribution becomes dispersed as one move towards the western Great Plains. [4]

---

[4] After resampling, we obtain 1,000 yields records in a county. Denote y^{det} as the detrended farm-level yields, then the nonparametric kernel estimates at a given point y can be defined as $\hat{g}(y) = (1/Nh)\sum_{i=1}^{N} K[(y_i^{\text{det}} - y)/h]$, where $N = 1,000$; $K(\cdot)$ is the Gaussian kernel function which is nonnegative and satisfies $\int K(z)dz = 1$, $\int zK(z)dz = 0$, $\int z^2 K(z)dz < \infty$. Parameter h is the smoothing parameter which can be defined as $\hat{h} = 0.9 \times \min[\sigma_y, IQR_y/1.34] \times N^{-0.2}$, where σ_y is the standard deviation; IQR_y is the interquartile range (i.e., the difference between 25th and 75th quantiles) of y^{det}.
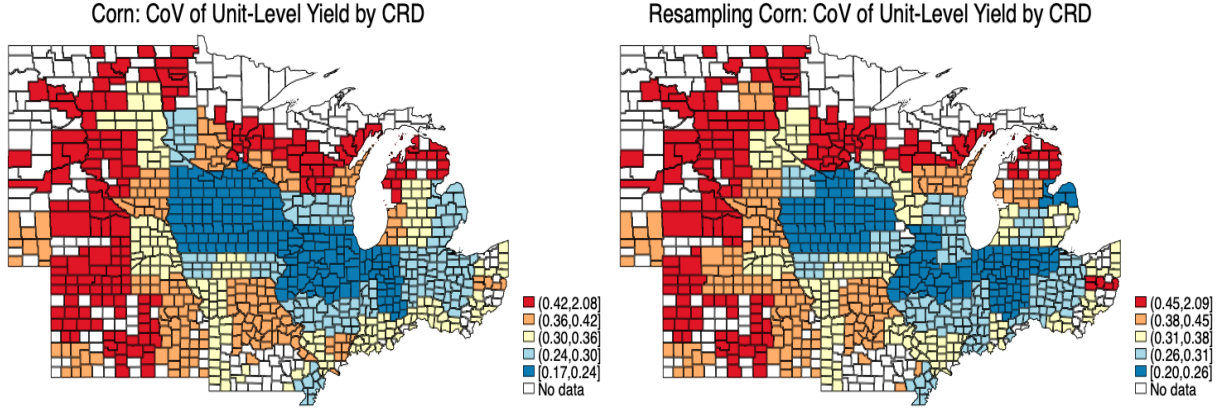
Figure 1.3 Coefficient of Variation from Original and Resampling Data for Corn by Crop Reporting District (CRD).

Wedge Factor

RMA's premium rates at all coverage levels can be imputed from the 2009 USDA/RMA farm-level contract choice data (see, for example, Du et al.2017). After obtaining the estimated actuarially fair premium rates, we can measure farm- and county-level actuarial fairness. Denote $APR_i(\varphi)$ as RMA's actual premium rates (including subsidies) at the coverage level φ for farm i,; $OPR_i(\varphi)$ as the actuarially fair premium rates; $a_{i,c}$ represents insured acreages of farm i; $s_{i,c} = a_{i,c}/\sum_i a_{i,c}$ represents a weight. m corresponds to the coverage level with $\varphi_1 = 0.5$ until $\varphi_8 = 0.85$. Then four wedge factors can constructed as:

(1.7a)   farm average:
$$w_i = \frac{\sum_{m=1}^{8} APR_i(\varphi_m)}{\sum_{m=1}^{8} OPR_i(\varphi_m)};$$

(1.7b)   farm differential:
$$\Delta w_i = \frac{\sum_{m=5}^{8} APR_i(\varphi_m)}{\sum_{m=5}^{8} OPR_i(\varphi_m)} - \frac{\sum_{m=1}^{4} APR_i(\varphi_m)}{\sum_{m=1}^{4} OPR_i(\varphi_m)};$$

(1.7c)   county average:
$$w_c = \frac{\sum_i \sum_{m=1}^{8} APR_i(\varphi_m) \times s_{i,c}}{\sum_i \sum_{m=1}^{8} OPR_i(\varphi_m) \times s_{i,c}};$$

(1.7d)   county differential:
$$\Delta w_c = \frac{\sum_i \sum_{m=5}^{8} APR_i(\varphi_m) \times s_{i,c}}{\sum_i \sum_{m=5}^{8} OPR_i(\varphi_m) \times s_{i,c}} - \frac{\sum_i \sum_{m=1}^{4} APR_i(\varphi_m) \times s_{i,c}}{\sum_i \sum_{m=1}^{4} OPR_i(\varphi_m) \times s_{i,c}}.$$
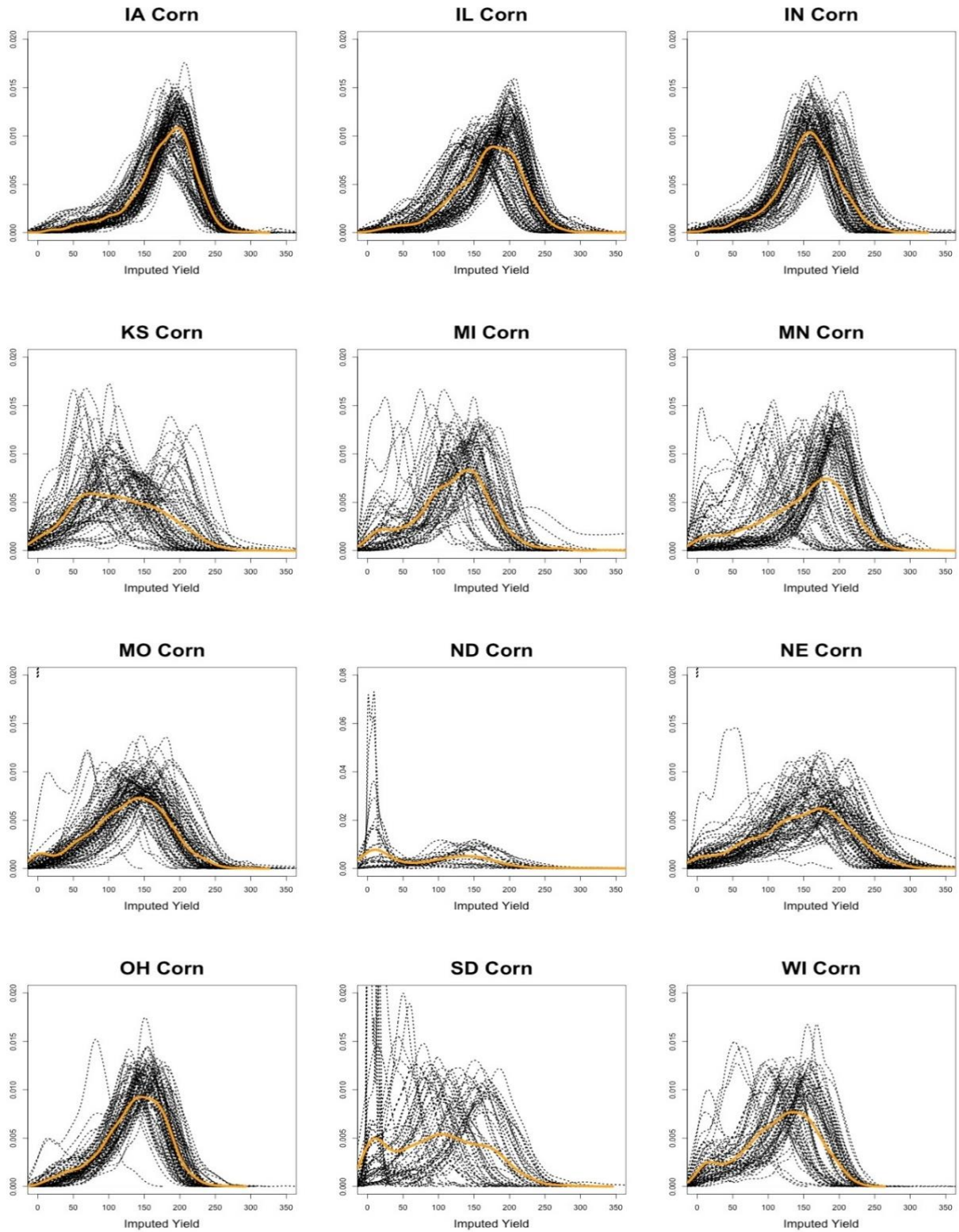
Figure 1.4 Detrended Unit-Level Imputed Yield Densities for Corn by County and States.
Note: "dashed curve" shows density for each county; "yellow curve" shows density for the state (randomly draw from all observations in this state).
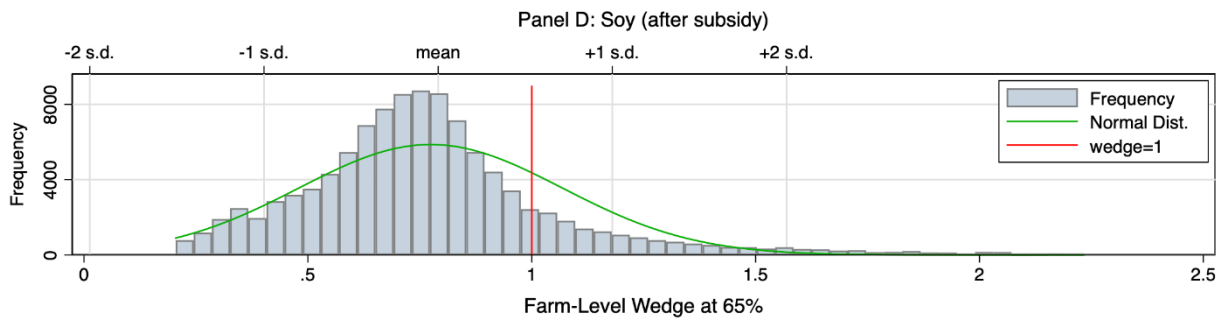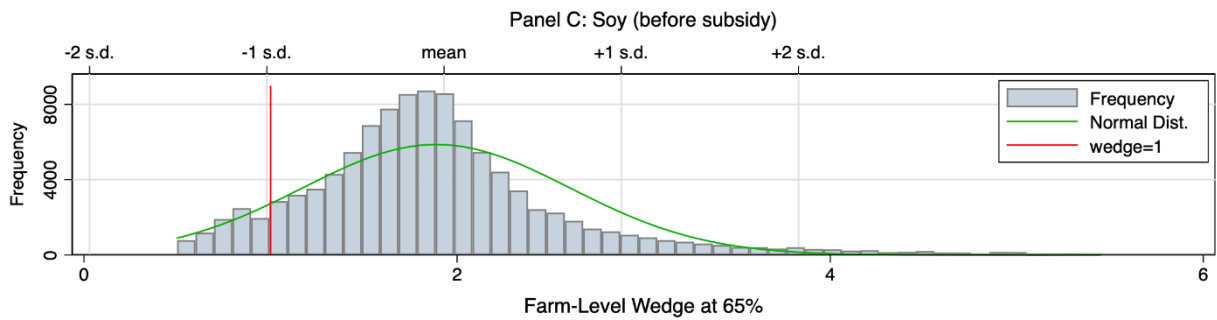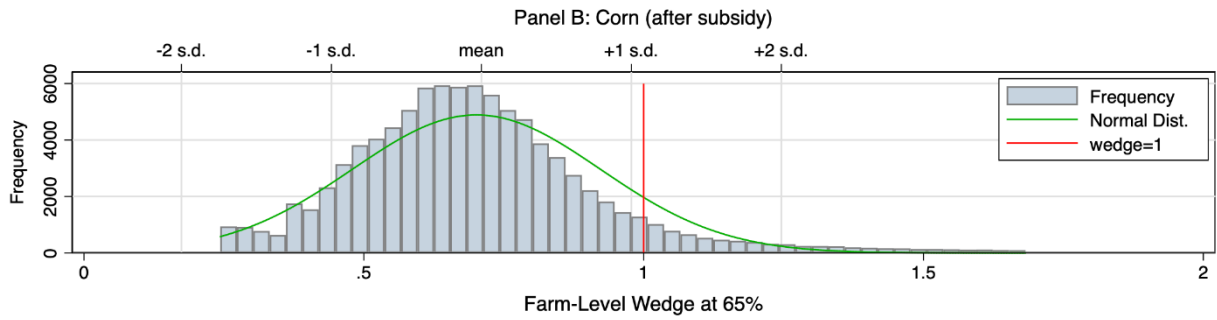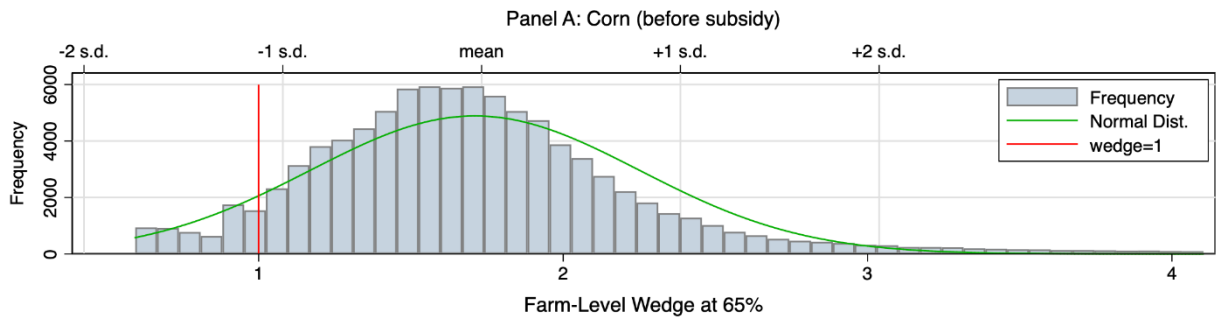
Figure 1.5 Distributions of Farm-level Wedges (before and after subsidies) for Corn and Soybean at the 65% Coverage Level.
Note: The green curve shows the normal distribution. The red line shows the actuarial fairness condition (loss ratio=1).

The average wedges (Eqn.(1.7a) and (1.7c)) capture the farmer's response to the systematic pricing bias; the differential wedges (Eqn.(1.7b) and (1.7d)) can capture farmers' response to benefits from higher coverage levels. The motivation for constructing both average and differential wedges is that USDA RMA encourages farmers to choose higher coverage levels. Therefore, wedges at higher coverage levels may differ from those at lower coverage levels.

Figure 1.5 shows the distributions of the farm-level wedges (before and after subsidies) for corn and soybean at the 65% coverage level. By comparing Panel A with B, we know the effect of subsidies on the farm-level wedges for corn. Before subsidy, most farm-level wedges are larger than the actuarial target (i.e., 1), which indicates that RMA's premiums in most farms are higher than the actuarial fair premiums. After subsidy, most farm-level wedges are less than 1, which indicates that subsidy policy may cover actuarial unfairness.[5] A similar pattern is also applied to soybeans (Panels C and D).

An inquiry may be generated because it seems that RMA's premiums (including subsidy) are systematically higher than our estimated premium. To investigate whether and why the systematic bias happens, we compare the inverse of wedge and loss ratio as in Figure 1.6. In our construction, the inverse of the wedge is the ratio of expected indemnity (the estimated actuarially fair premium) to RMA's premium, which is the loss ratio used by RMA. Therefore, the inverse of the wedge and loss ratio should be consistent. Panel A and B of Figure 1.6 report the inverse of wedge and loss ratio for corn, respectively. We find that: first, the actuarially fair premiums (i.e., expected payout) are higher than actual indemnities in the Corn Belt region,

---

[5] For example, the mean wedge before subsidy for corn is 1.732, which is changed to 0.71 after subsidy, consistent with the subsidy rate at a 65% coverage level ($0.71=1.73*(1-0.59)$).

implying RMA's premiums might be underestimated in the main production area; second, the inverses of the wedge are lower than the loss ratios when moving outside the main production area, implying that RMA's premiums might be overestimated. Patterns for soybean are similar (Panel C and D of Figure 1.6). The findings above indicate that actuarial unfairness might be correlated with land quality.

To examine whether wedge is related to land quality, we set six quantiles of land quality (10th,25th,50th,75th, 90th, and 95th) based on the farm's APH in each Cropping Record District (CRD). It is noteworthy that we are interested in whether the farmer with higher quality cropland receives a worse-than-actuarially-fair rating. Therefore, the 5th quantile of land quality is not included. Finally, we may ask why the quantiles are divided in each CRD. The reasons are that: first, yields corresponding quantiles are not too concentrated so that we can avoid many same observations being divided into one pool; second, yield risk in one CRD is similar, so yields corresponding quantiles are not too dispersed.

Figure 1.7 reports the wedge means for each quantile land quality. We find: (i) wedge (before subsidy) decreases with coverage level in lower land quality, but a contrary case exists in higher land quality; (ii) wedge (after subsidy) increases with coverage level in all land qualities; (iii) wedges (after subsidy) in lower land quality (10th and 25th quantiles) are less than 1 (actuarial fair target), but some wedges in higher land quality are higher than 1; and (iv) for higher land quality (50th -95th quantiles), wedges (after subsidy) increase with coverage level and intersect with 1 at 80% coverage level.
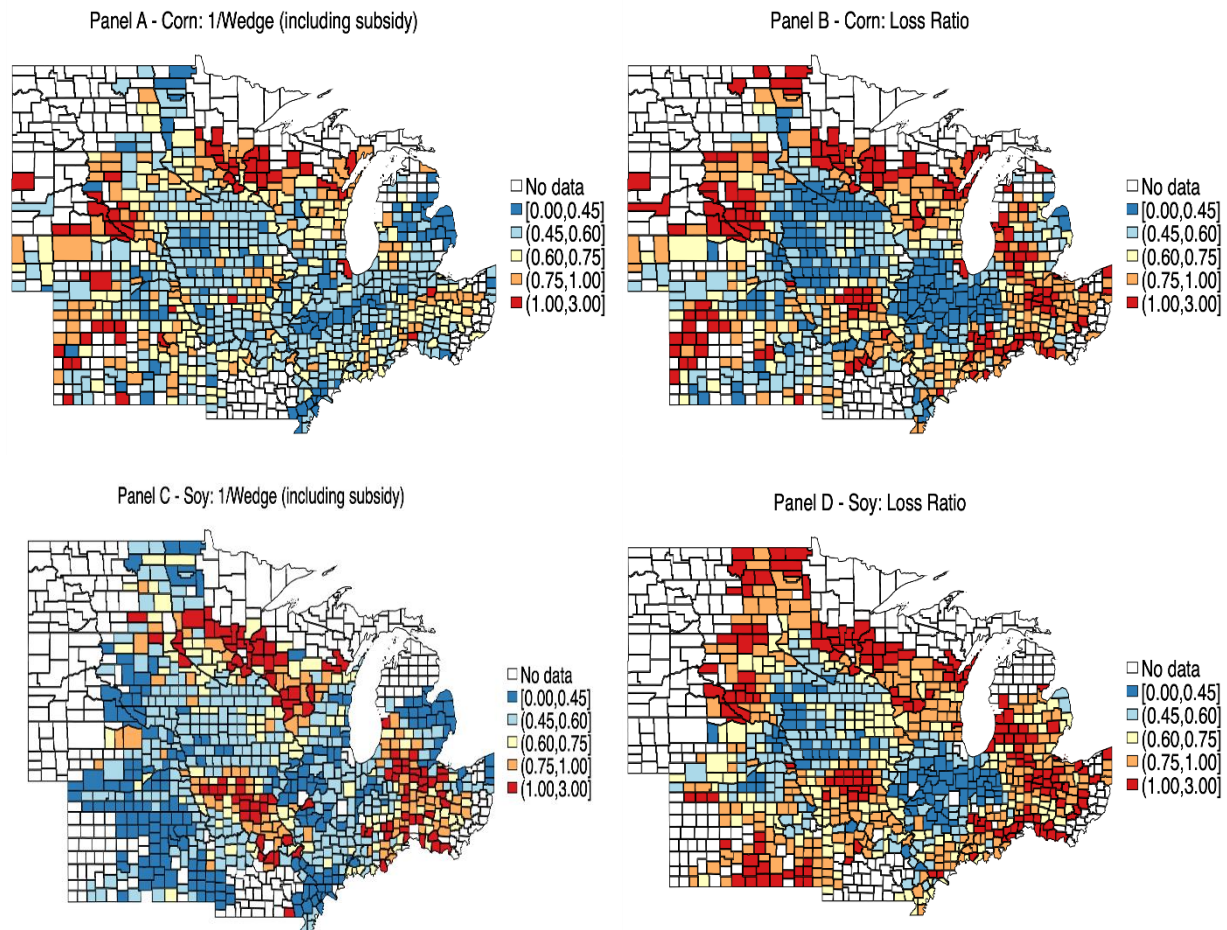
Figure 1.6 Comparison between Expected Loss Ratio (1/Wedge) and Average Loss Ratio.
Note: As the construction in this study, 1/Wedge represents the ratio of expected indemnity to actual premium, which means that 1/Wedge is expected loss ratio.

Land Capability

Soil quality data are from National Resource Inventory (NRI). As in USDA (2015), there are eight Land Capability Classes (LCC). To be specific, Class I soils have limitations that restrict their use. However, Class VIII soils and miscellaneous areas have limitation that precludes their use for commercial plant production and limits their use for recreation, wildlife, water supply, or aesthetic purposes. As in the literature, we measure the land capability by calculating the percentage of good land in total land. We employ Class I-II as in Goodwin et

24

al. (2004), although Du et al. (2017) choose Class I-IV.

We now turn to show whether actuarial unfairness exists in a macro-level region since land capability has inter-county heterogeneities. As in Figure 1.8, the x-axis shows the ranking of land capability (01-lowest; 12-highest) and state name. The distribution in each state shows county-level wedge heterogeneities at the 65% coverage level. For corn, we find that wedges (before subsidies) are lower in states with lower land capability (e.g., WI, SD, MN); however, wedges (before subsidies) are higher in the main production area (e.g., IA, IL, IN). This pattern implies that the wedge (before subsidies) might be inversely correlated with land quality. It is noteworthy that: (i) for corn, some states with lower land capability have high wedges, e.g., MO, MI, and NE; (ii) for soybean, the relationship between wedges and land capabilities is not entirely clear. Therefore, we will apply empirical methods for further analysis.

Weather Determinants

As in the literature (e.g., Schlenker and Roberts 2009; Xu et al. 2013), we construct two variables-Growing Degree Days (GDD) and Stress Degree Days (SDD)-to represent beneficial heat and heat stress during the growing season (April to September) as below:

(1.8a) $GDD_{c,t} = \sum_{d \in \Omega_t} \left[0.5 \left(min\left(max\left(T_{c,d,t}^{max}, T^l\right), T^h\right) + min\left(max\left(T_{c,d,t}^{min}, T^l\right), T^h\right) \right) - T^l \right]$

(1.8b) $SDD_{c,t} = \sum_{d \in \Omega_t} \left[0.5 \left(max\left(T_{c,d,t}^{max}, T^k\right) + max\left(T_{c,d,t}^{min}, T^k\right) \right) - T^k \right]$

where c is county; d is day; Ω_t is the set of growing season days in year t. The thresholds are $T^l = 10°C$, $T^h = 30°C$ and $T^k = 32.2°C$. We use climatological normal around 20 years $\{1989, ..., 2007\}$ to control for weather effect on coverage level choice as:

(1.9a) $\bar{G}_c = (1/19) \sum_{j=1989}^{2007} GDD_{c,j}$;

(1.9b) $\bar{S}_c = (1/19) \sum_{j=1989}^{2007} SDD_{c,j}$.
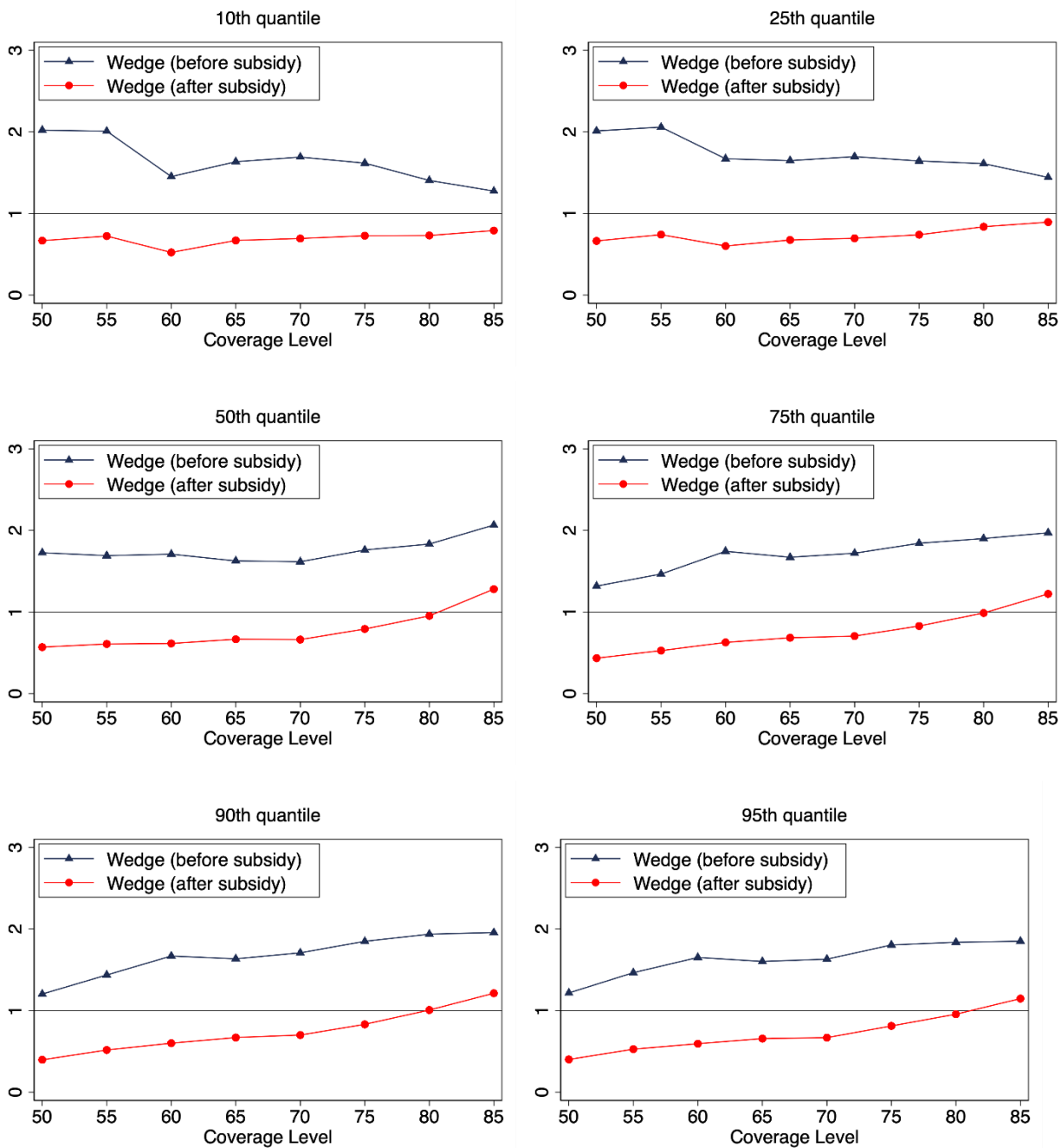
Figure 1.7 Comparisons of Wedges (before and after subsidies) among Different Land Qualities for Corn.
Note: Land quality is measured by farm-level APH within a Cropping Record District. The tolerance is 20 bushels for each quantile. For instance, suppose the 25th quantile of APH is 112, then observations who own APH within the interval (102, 122) will be used for the 25th quantile part.
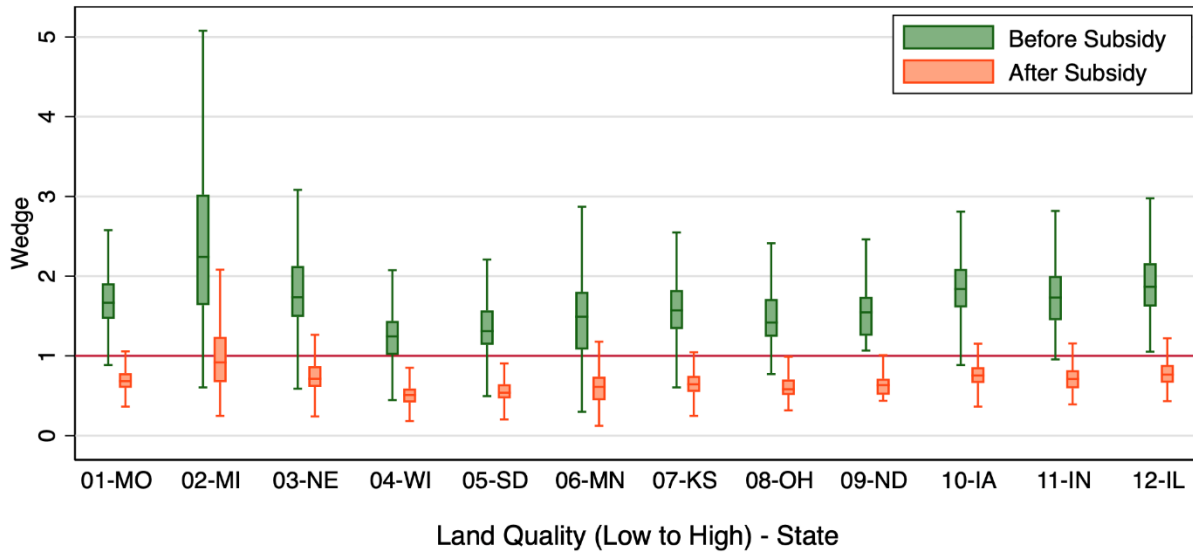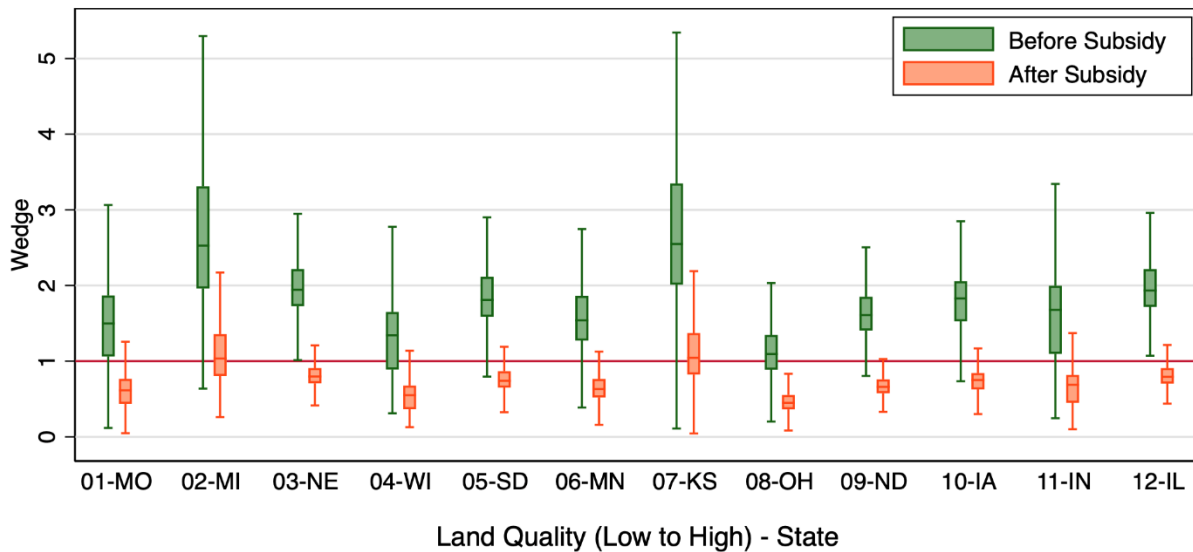
Figure 1.8 Wedge Comparison at 65% Coverage Level.
Note: The distribution on each box is based on farm-level observations in each state. Central mark is median; edges are 25th (Q1) and 75th (Q3) quantiles respectively. Upper and lower whiskers include data points within (+)- 1.5×IQR where IQR=Q3-Q1. The distribution is based on farms in each state.

Empirical Results

Evidence on Hypotheses

We now turn to examine the hypotheses H1-H4. Recall that H1 states that farmers'

coverage choices are clustered at one single coverage level or two adjacent levels. From Table

1.4, we find that farmers choosing 65% and 70% coverage levels account for over 55% of the

whole sample, which supports the H1. This conclusion holds for different crops and unit types.

Furthermore, the inner solution indicates that the gradient of the self-paid premium function (χ_φ)

is between the two end coverage levels (0.5 and 0.85).

Table 1.4 Evidence for H1.

Crop	Plan 90	% obs. at 50%	% obs. at 55%	% obs. at 60%	% obs. at 65%	% obs. at 70%	% obs. at 75%	% obs. at 80%	% obs. at 85%
Corn	All	0.12	0.02	0.05	0.30	0.25	0.18	0.06	0.03
	OpU	0.10	0.01	0.05	0.29	0.27	0.18	0.06	0.03
	BaU	0.14	0.02	0.05	0.32	0.21	0.18	0.06	0.03
Soybean	All	0.12	0.01	0.04	0.30	0.25	0.18	0.06	0.03
	OpU	0.10	0.01	0.05	0.28	0.28	0.19	0.06	0.03
	BaU	0.14	0.02	0.05	0.33	0.21	0.17	0.06	0.02

Note: the sample only includes OpU and BaU.

Recall that H2 states farmers choose the highest coverage level, at which sum of

elasticities of subsidy rate and wedge factor is larger than -1. To test this hypothesis, we first

define two common elasticities as:

(1.10a)   Conservative: $\Delta_1^s(\varphi) = \frac{s(\varphi_+)-s(\varphi_-)}{\varphi_+-\varphi_-} \times \frac{\varphi_+}{s(\varphi_+)}$,

$\Lambda_1^w(\varphi) = \frac{w(\varphi_+)-w(\varphi_-)}{\varphi_+-\varphi_-} \times \frac{\varphi_+}{w(\varphi_+)}$;

(1.10b)   Arc Elasticity: $\Delta_2^s(\varphi) = \frac{s(\varphi_+)-s(\varphi_-)}{\varphi_+-\varphi_-} \times \frac{\varphi_++\varphi_-}{s(\varphi_+)+s(\varphi_-)}$;

$$\Lambda_2^w(\varphi) = \frac{w(\varphi_+) - w(\varphi_-)}{\varphi_+ - \varphi_-} \times \frac{\varphi_+ + \varphi_-}{w(\varphi_+) + w(\varphi_-)},$$

where φ_+ and φ_- are adjoining coverage levels with $\varphi_+ > \varphi_-$ and $\varphi_+, \varphi_- \in \{0.5, \ldots, 0.85\}$. Both elasticities are employed to eliminate the potential mismeasurement due to the different start and end points. From Table 1.5, 75% should be farmers' favorite coverage level because the 75% coverage level is the highest level satisfying the sum of elasticities larger than -1. Since the percentage of observations at the 65% coverage level is the highest in Table 1.4, H2 is not supported, which implies that we cannot conclude that farmers are concerned about the subsidy transfer.

Table 1.5 reports evidence for H3, which states that farmers choose the coverage level at which the wedge (including subsidies) arrives at the minimum value. Wedges for corn and soybeans arrive at the minimum levels at the 65% coverage choices, implying that farmers are concerned about wedges (including subsidies). Therefore, policymakers might consider adjusting actuarial unfairness rather than maintaining large subsidies. For example, if the federal government's target encourages farmers to choose higher coverage levels, wedges at higher levels can be adjusted downward. It is noteworthy that the adjustment of wedges cannot be uniform nationwide. If the actuarially fair premiums are constant, the downward adjustment of wedges indicates RMA's premiums should be reduced, which might increase the loss ratio.

Recall that H4 states farmers with higher (lower) land quality cropland pay more (fewer) premiums than they should. However, as in Figure 1.9, the mean values of RMA's premiums (before subsidy) are higher than that of the actuarially fair premiums for all land qualities. Moreover, the difference between RMA's and actuarially fair premiums becomes more significant as the coverage level increases. Therefore, H4 is not robustly supported if we compare the mean values of RMA's premiums and the actuarially fair premiums.

Table 1.5 Evidence for H2 and H3.

Coverage Level φ	CAT	50%	55%	60%	65%	70%	75%	80%	85%
Subsidy Rate $s(\varphi)$	1	0.67	0.64	0.64	0.59	0.59	0.55	0.48	0.38
$\Delta_1^s(\varphi)$			-0.52	0	-1.32	0	-1.09	-2.33	-4.47
$\Delta_2^s(\varphi)$			-0.48	0	-1.02	0	-1.02	-2.11	-3.84
Corn Wedge $w(\varphi)$	NA	1.73	1.72	1.69	1.66	1.71	1.75	1.78	1.80
$\Lambda_1^w(\varphi)$			-0.12	-0.18	-0.24	0.43	0.37	0.26	0.18
$\Lambda_2^w(\varphi)$			-0.12	-0.18	-0.23	0.42	0.36	0.25	0.17
Soybeans Wedge $w(\varphi)$	NA	1.93	1.89	1.85	1.82	1.87	1.93	1.98	2.02
$\Lambda_1^w(\varphi)$			-0.27	-0.24	-0.26	0.43	0.44	0.40	0.35
$\Lambda_2^w(\varphi)$			-0.25	-0.23	-0.25	0.42	0.43	0.39	0.34
Corn									
$\Delta_1^s(\varphi)+\Lambda_1^w(\varphi)$			-0.64	-0.18	-1.56	0.43	-0.72	-2.07	-4.29
$\Delta_2^s(\varphi)+\Lambda_2^w(\varphi)$			-0.60	-0.18	-1.25	0.42	-0.66	-1.86	-3.67
Soybeans									
$\Delta_1^s(\varphi)+\Lambda_1^w(\varphi)$			-0.79	-0.24	-1.58	0.43	-0.65	-1.93	-4.12
$\Delta_2^s(\varphi)+\Lambda_2^w(\varphi)$			-0.73	-0.23	-1.27	0.42	-0.59	-1.72	-3.50

Table 1.6 Coverage Choices by Land Qualities.

| Land Quality (quantile)	Coverage Level							
	% obs. at 50%	% obs. at 55%	% obs. at 60%	% obs. at 65%	% obs. at 70%	% obs. at 75%	% obs. at 80%	% obs. at 85%
Corn								
10th	12.8	1.9	6.5	36.9	28.5	11.3	1.8	0.2
25th	11.7	1.8	6.2	32.7	28.0	16.3	2.8	0.6
50th	11.7	1.6	4.8	27.9	25.6	20.5	5.5	2.5
75th	10.4	1.4	3.6	26.0	21.1	22.8	9.4	5.2
90th	10.6	1.2	4.0	26.2	20.9	21.6	9.5	6.1
95th	11.4	1.4	3.7	25.0	20.8	20.9	10.5	6.3
Soybeans								
10th	7.8	1.0	3.9	46.8	31.9	7.8	0.6	0.1
25th	11.3	1.3	6.1	39.7	31.3	8.3	1.8	0.2
50th	13.1	1.6	5.4	33.5	28.5	14.8	2.5	0.5
75th	14.7	1.5	4.4	28.2	24.7	19.6	5.2	1.8
90th	13.3	1.5	4.3	25.9	20.6	22.3	8.3	3.8
95th	10.2	1.6	3.6	24.3	19.9	23.6	10.6	6.2
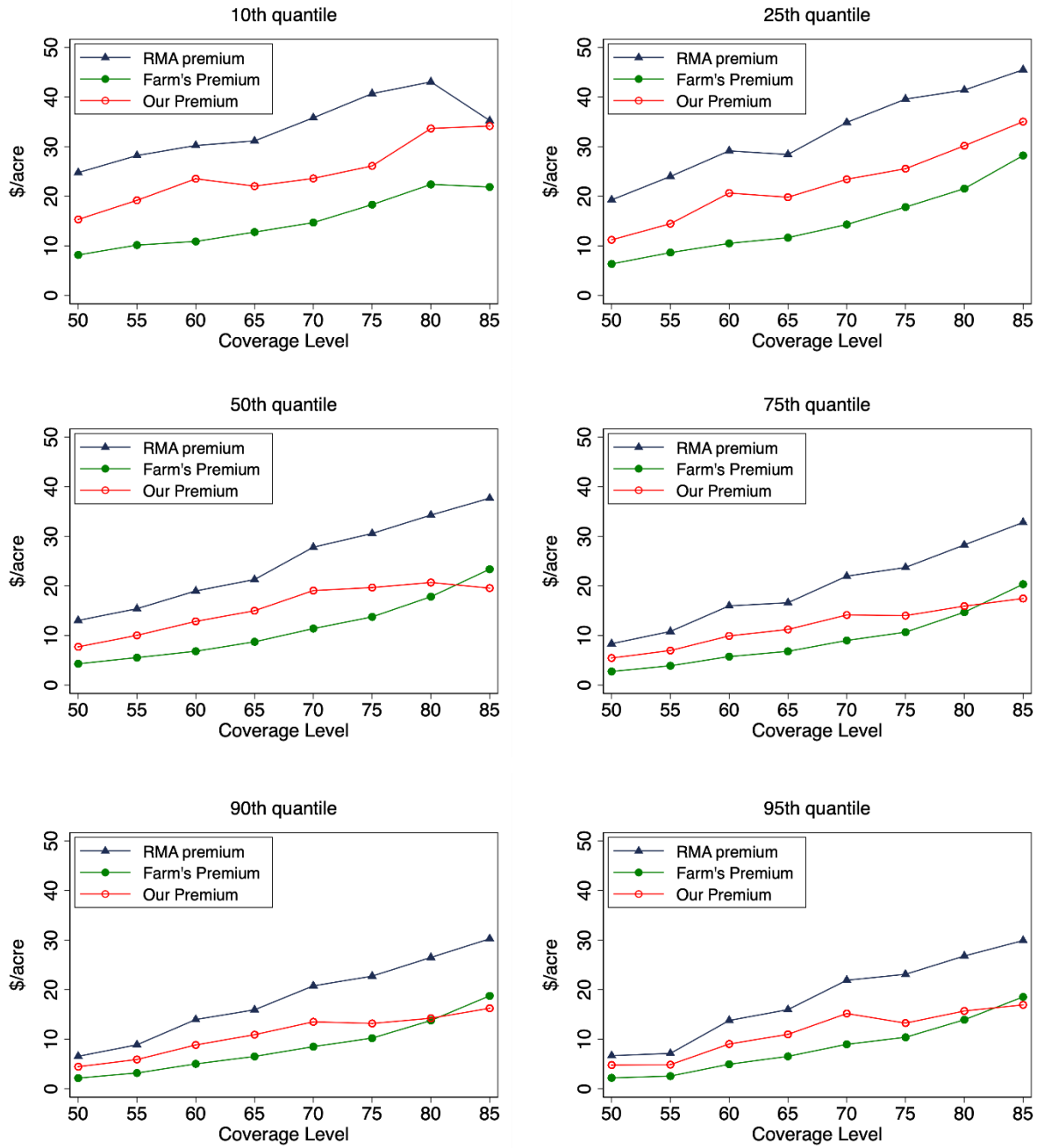
Figure 1.9 RMA Premiums (including subsidies), Farm's Self-Paid Premium and Our Estimated Actuarially Fair Premiums for Corn.
Note: Only mean values are reported here.

Table 1.6 reports the farmer's choices in terms of land qualities. We find that: (i) the 65%

coverage level is the most popular level for all land qualities, and (ii) the percentage for higher

coverage level choices increases when land quality increases. For example, the percentage of

75% coverage level for the 10th land quality is 11.3, but that for the 95th land quality increases to

20.9. Results in Table 1.6 imply that farmers with better land quality are inclined to choose

higher coverage levels.

Regression Results

We first employ the ordered logit model (OLM) to analyse the farm-level coverage

choice. The dependent variable is each farm's actual coverage choice; the independent variables

of most interest are average wedge across all coverage levels (i.e., w_i) and wedge differential

(i.e., Δw_i). The model specification we employed is as follows:

(1.10a) $\qquad Prob(\varphi_i = \varphi_1 | \mathbf{X_i}, \mathbf{b}, c) = 1 - \frac{exp(\mathbf{X_i b} - c_1)}{1 + exp(\mathbf{X_i b} - c_1)};$

(1.10b) $\qquad Prob(\varphi_i = \varphi_j | \mathbf{X_i}, \mathbf{b}, c) = \frac{exp(\mathbf{X_i b} - c_{j-1})}{1 + exp(\mathbf{X_i b} - c_{j-1})} - \frac{exp(\mathbf{X_i b} - c_j)}{1 + exp(\mathbf{X_i b} - c_j)}, j = 2,3,...,7;$

(1.10c) $\qquad Prob(\varphi_i = \varphi_8 | \mathbf{X_i}, \mathbf{b}, c) = \frac{exp(\mathbf{X_i b} - c_7)}{1 + exp(\mathbf{X_i b} - c_7)},$

where $i \in \{1,2,...,N\}$ is the index of each farm in the sample; $\varphi \in \{0.5,...,0.85\}$ with $\varphi_1 = 0.5$,

$\varphi_2 = 0.55$ until $\varphi_8 = 0.85$; $\mathbf{X_m}$ is the vector of the exogenous variables $\{w_i, \Delta w_i, LCC, \bar{G}_c, \bar{S}_c\}$; $\mathbf{b}$

is the coefficient vector; $c_j, j \in \{1,...,7\}$ are cut points of the distribution. To investigate the

effect of land quality on farmer's coverage level choice, we add an indicator variable

$index\_CRD$ showing quantiles of land quality in a Cropping Report District. Specifically,

$index\_CRD = 1$ represents 10th quantile and $index\_CRD = 6$ represents 95th quantile. For the

robustness checks, we also employ a county's longitude and lattitude in order to control for some

non-climate features such as technological adoption and infrastructure availability.

Next, by calculating weights based on farm-level insured acres, we investigate that how

wedges (i.e., w_c and Δw_c) affect coverage choices from a county-level perspective. We employ

Ordinary Least Squares (OLS) for the primary estimation, and use Weighted Least Squares

(WLS) for robustness checks. In WLS estimation, suppose the model specification be $y =$

$\mathbf{X}^{\mathbf{T}}\boldsymbol{\beta} + \varepsilon$ and $\varepsilon^2 = \mathbf{X}^{\mathbf{T}}\boldsymbol{\gamma} + u$, we employ a procedure as: i) regress y on $\mathbf{X}$ by OLS and have

residuals $\hat{\varepsilon}$; ii) run $\hat{\varepsilon}^2$ on $\mathbf{X}$ and have $\hat{\boldsymbol{\gamma}}$; and iii) use $1/\sqrt{\mathbf{X}^{\mathbf{T}}\hat{\boldsymbol{\gamma}}}$ as the weight to correct

heterogeneity in ε.

Table 1.7 reports the regression results for the farm-level coverage choices. The farmers'

coverage choice responses to wedge average and wedge differential are significantly negative for

corn, although they are unclear to soybean. However, land quality variables – land quantile and

land capability- have significantly positive effects on farmers' choices, implying that farmers

with better land quality are inclined to choose higher coverage levels.

Table 1.8 reports OLS estimation results, and Table 1.9 reports the WLS results. Again,

like the farm-level analysis, the wedge effect (average) on county-level coverage choice is

significantly negative for corn but unclear for soybean. As in Table 1.8, for corn, the coefficient

of wedge average is around -1.3, indicating the marginal effect of wedge average is also around -

1.3 since we use OLS estimation. Then the semi-elasticity of coverage choice is -2.2, implying

that the mean coverage level will increase to 0.69 from 0.67 if the wedge average decreases by

1%. Furthermore, the wedge effect (differential) is unclear for corn and soybean.

Conclusion

Over the past twenty years, participation rates in the U.S. federal crop insurance program

have improved nationwide; however, significant regional disparities still exist. This study

contributes to the literature by proposing a novel resampling procedure, estimating the

actuarially fair premiums at all coverage levels, and then examining the effect of actuarial unfairness on farmers' insurance demand.

We find that: *first*, farmers are more concerned about the actuarial unfairness than the subsidy transfer, implying that mitigation of mispricing supplements the subsidy; *second*, actuarial unfairness is positively correlated with land quality, indicating that farmers with higher land quality cropland pay more premiums than they should. Finally, regression results show that such mispricing deters farmers' insurance uptake, which is robust from the farm- or county-level perspectives.

Our analysis might help the sustainability of the crop insurance program because an adverse selection problem might exist if federal subsidies were to be reduced in the future. When farm-level mispricing exists and subsidy reduction starts, farmers with higher land quality cropland will opt out of the program. If this is the case, the entire insurance pool becomes risky, and indemnities might exceed premiums, which will be detrimental to the development of the program. In addition, due to the increased indemnities, RMA might adjust the premium rating upward and crowd more farmers with good land quality out of the insurance pool, further exacerbating the program.

The effect of mispricing on coverage level choice is significant for corn but not for soybean. According to our estimation, if the wedge (=RMA's premium/the actuarially fair premium) decreases by 10%, county-level coverage choices of corn will increase by 0.2 percentage points. Therefore, the mitigation of mispricing provides an opportunity to adjust the current subsidy structure.

Table 1.7 Estimation Results for Ordered Logit Model.

Dependent Variable	Farm-Level Coverage Choice							
Estimation Method	Ordered Logit Model							
Crop	Corn				Soybeans			
Wedge Average	-0.082***	-0.090***	-0.074***	-0.081***	0.020***	0.020***	0.008	0.009
	(-7.42)	(-8.01)	(-6.66)	(-7.15)	(3.36)	(3.38)	(1.26)	(1.49)
Wedge Differential		-0.158***		-0.133***		-0.007		-0.125***
		(-3.49)		(-2.93)		(-0.22)		(-3.76)
Land Quantile	0.007	0.010**	0.009**	0.012***	0.039***	0.040***	0.036***	0.039***
	(1.63)	(2.29)	(2.17)	(2.69)	(9.22)	(9.13)	(8.47)	(9.07)
Land Capability	1.349***	1.367***	1.353***	1.368***	1.887***	1.888***	1.879***	1.883***
	(36.93)	(36.98)	(36.95)	(36.94)	(54.50)	(54.42)	(53.75)	(53.81)
Control Variable								
Growing Degree Days	-0.000***	-0.000***	0.000	0.000	-0.001***	-0.001***	-0.000***	-0.000**
	(-3.02)	(-2.76)	(0.30)	(0.37)	(-16.74)	(-16.68)	(-2.62)	(-2.52)
Stress Degree Days	-0.014***	-0.014***	-0.011***	-0.011***	-0.008***	-0.008***	-0.005***	-0.005***
	(-24.52)	(-24.75)	(-15.47)	(-15.69)	(-14.46)	(-14.13)	(-8.38)	(-7.85)
Longitude			0.025***	0.025***			0.014**	0.011**
			(4.71)	(4.72)			(2.48)	(1.97)
Latitude			0.081***	0.079***			0.176***	0.178***
			(8.61)	(8.38)			(23.81)	(23.97)
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	81,088	81,088	81,088	81,088	104,209	104,209	104,209	104,209
Pseudo R^2	0.042	0.042	0.042	0.042	0.046	0.046	0.048	0.048

Note: The t-statistics are in the parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.8 The OLS Estimation Results for County-Level Choice.

Dependent Variable	Acreage-Weighted Coverage Choice							
Estimation Method	OLS							
Crop	Corn				Soybeans			
Wedge Average	-1.288***	-1.314***	-1.222***	-1.244***	-0.406*	-0.362	-0.409*	-0.357
	(-3.79)	(-3.83)	(-3.66)	(-3.69)	(-1.81)	(-1.58)	(-1.84)	(-1.58)
Wedge Differential		-0.740		-0.614		-0.851		-1.027
		(-0.63)		(-0.52)		(-0.72)		(-0.86)
Land Capability	5.053***	5.111***	4.918***	4.967***	4.022***	4.030***	3.949***	3.955***
	(4.77)	(4.80)	(4.68)	(4.69)	(4.58)	(4.58)	(4.51)	(4.51)
Growing Degree Days	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.000	-0.000
	(-0.72)	(-0.67)	(-0.46)	(-0.43)	(-0.58)	(-0.56)	(-0.09)	(-0.05)
Stress Degree Days	-0.031*	-0.032*	-0.041**	-0.041**	0.003	0.002	-0.001	-0.002
	(-1.71)	(-1.74)	(-2.09)	(-2.10)	(0.16)	(0.14)	(-0.08)	(-0.11)
Longitude			-0.503*	-0.493*			-0.377	-0.392
			(-1.94)	(-1.92)			(-1.52)	(-1.57)
Latitude			0.040	0.035			0.456	0.474
			(0.09)	(0.08)			(1.04)	(1.09)
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CRD FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	776	776	776	776	751	751	751	751
Pseudo R^2	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58

Note: The t-statistics are in the parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.9 The WLS Estimation Results for County-Level Choice.

Dependent Variable	Acreage-Weighted Coverage Choice							
Estimation Method	WLS							
Crop	Corn				Soybeans			
Wedge Average	-1.305***	-1.305***	-1.253***	-1.25***	-0.257	-0.218	-0.241	-0.193
	(-3.95)	(-3.94)	(-3.76)	(-3.69)	(-1.09)	(-0.92)	(-1.02)	(-0.81)
Wedge Differential		0.001		0.119		-1.167		-1.349
		(-0.00)		(0.11)		(-1.06)		(-1.22)
Land Capability	4.453***	4.452***	4.375***	4.367***	3.636***	3.614***	3.554***	3.526***
	(5.14)	(5.12)	(4.99)	(4.96)	(4.13)	(4.11)	(3.96)	(3.93)
Growing Degree Days	-0.001	-0.001	-0.001	-0.001	0.00001	0.0001	0.003	0.001
	(-0.61)	(-0.60)	(-0.37)	(-0.39)	(0.01)	(-0.08)	(0.23)	(0.35)
Stress Degree Days	-0.025	-0.025	-0.031	-0.03**	0.011	0.01	0.005	0.004
	(-1.28)	(-1.28)	(-1.49)	(-1.48)	(0.51)	(0.48)	(0.23)	(0.17)
Longitude			-0.317	-0.32			-0.31	-0.34
			(-1.29)	(-1.3)			(-1.2)	(-1.30)
Latitude			0.011	0.01			0.12	0.15
			(0.03)	(0.03)			(0.32)	(0.39)
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CRD FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	770	770	770	770	733	733	733	733
Pseudo R^2	0.52	0.52	0.52	0.52	0.50	0.50	0.50	0.50

Note: The t-statistics are in the parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

REFERENCES

Babcock, B.A. and A.M., Blackmer. 1992. The value of reducing temporal input nonuniformities. *Journal of Agricultural and Resource Economics*: 335-347.

Babcock, B.A., E.K. Choi and E. Feinerman. 1993. Risk and probability premiums for CARA utility functions. *Journal of Agricultural and Resource Economics*: 17-24.

Babcock, B.A. and D.A., Hennessy.1996. Input demand under yield and revenue insurance. *Amer. J. of Agr. Econ*, 78(2): 416-427.

Babcock, B. A., Hart, C. E., & Hayes, D. J. 2004. Actuarial fairness of crop insurance rates with constant rate relativities. *Amer. J. of Agr. Econ*. 86(3): 563-575.

Borges, R.B. and Thurman, W.N., 1994. Marketing quotas and random yields: marginal effects of inframarginal subsidies on peanut supply. *Amer. J. of Agr. Econ.* (4): 809-817.

Calvin, L. 1992. Participation in the US federal crop insurance program (No. 1800). US Department of Agriculture, Economic Research Service.

Chen, S. and L.Y. Cindy. 2016. Parameter estimation through semiparametric quantile regression imputation. *Electronic Journal of Statistics*, 10(2): 3621-3647.

Chen, Z., Dall'Erba, S. and Sherrick, B.J., 2020. Premium misrating in federal crop insurance programs: scale, geography, and fiscal impacts. *Agricultural Finance Review*, 80(5): 693-713.

Chite, R. 1988. Federal crop insurance: background and current issues. In CRS report for Congress (USA). Congressional Research Service.

Claassen, R. and R.E Just, 2011. Heterogeneity and distributional form of farm‐level yields. *Amer. J. of Agr. Econ.* 93(1): 144-160.

Coble, K.H., T.O. Knight, R.D. Pope and J.R. Williams, 1996. Modeling farm‐level crop insurance demand with panel data. *Amer. J. of Agr. Econ*. 78(2): 439-447.

Coble, K.H., T.O. Knight, R.D. Pope and J.R. Williams. 1997. An expected-indemnity approach to the measurement of moral hazard in crop insurance. *Amer. J. of Agr. Econ.* 79(1): 216-226.

Coble, K.H., T.O. Knight, B.K. Goodwin, M.F. Miller, R.M. Rejesus and G. Duffield. 2010. A comprehensive review of the rma aph and combo rating methodology: Final report. prepared by sumaria systems for the risk Management agency.

Deng, X., B.J. Barnett and D.V. Vedenov. 2007. Is there a viable market for area-based crop insurance?. *Amer. J. of Agr. Econ*. 89(2): 508-519.

Du, X., D.A. Hennessy and H. Feng. 2014. A natural resource theory of US crop insurance

contract choice. *Amer. J. of Agr. Econ.* 96(1): 232-252.

Du, X., H. Feng and D.A. Hennessy. 2017. Rationality of choices in subsidized crop insurance markets. *Amer. J. Agr. Econ.* 99(3): 732-756.

Feng, H., X. Du and D.A. Hennessy. 2020. Depressed demand for crop insurance contracts, and a rationale based on third generation Prospect Theory. *Agricultural Economics*, 51(1): 59-73.

Firpo, S., A.F. Galvao, M. Kobus, T. Parker and P. Rosa-Dias. 2020. Loss aversion and the welfare ranking of policy interventions. arXiv preprint arXiv:2004.08468.

Gardner, B.L. and R.A. Kramer. 1986. Experience with crop insurance programs in the United States.

Glickman, D., 2000. Testimony of Dan Glickman before the House Committee on Agriculture.

Gollier, C. and J.W. Pratt. 1996. Risk vulnerability and the tempering effect of background risk. Econometrica: *Journal of the Econometric Society*: 1109-1123.

Goodwin, B.K., 1993. An empirical analysis of the demand for multiple peril crop insurance. *Amer. J. Agr. Econ.* 75(2): 425-434.

Goodwin, B.K. 1994. Premium rate determination in the federal crop insurance program: what do averages have to say about risk?. *Journal of Agricultural and Resource Economics*: 382-395.

Goodwin, B.K. and A.P Ker, 1998. Nonparametric estimation of crop yield distributions: implications for rating group‐risk crop insurance contracts. *Amer. J. Agr. Econ.* 80(1): 139-153.

Goodwin, B.K. 2001. Problems with market insurance in agriculture. *Amer. J. of Agr. Econ.* (3): 643-649.

Hojjati, B. and N.E. Bockstaal. 1988. Modeling the demand for crop insurance. International Food Policy Research Institute.

Hirshleifer, J., H. Jack and J.G. Riley, 1992. The analytics of uncertainty and information. Cambridge University Press.

Jensen, N.D., A.G. Mude and C.B. Barrett, 2018. How basis risk and spatiotemporal adverse selection influence demand for index insurance: Evidence from northern Kenya. *Food Policy*, 74: 172-198.

Joe, H. and J.J. Xu. 1996. The estimation method of inference functions for margins for multivariate models.

Just, R. E., and Q. Weninger. 1999. "Are Crop Yields Normally Distributed?" *Amer. J. Agr.*

Econ. 81: 287-304.

Ker, A.P. and B.K., Goodwin. 2000. Nonparametric estimation of crop insurance rates revisited. *Amer. J. Agr. Econ.* 82(2): 463-478.

Ker, A.P. and Coble, K., 2003. Modeling conditional yield densities. *Amer. J. Agr. Econ.* 85(2): 291-304.

Knight, T.O. and K.H., Coble. 1999. Actuarial effects of unit structure in the US actual production history crop insurance program. Journal of Agricultural and Applied Economics, 31(3): 519-535.

Koenker, R. and G. Bassett Jr. 1978. Regression quantiles. *Econometrica*: *journal of the Econometric Society*: 33-50.

Kousky, C. 2017. Disasters as learning experiences or disasters as policy opportunities? Examining flood insurance purchases after hurricanes. *Risk analysis*, 37(3): 517-530.

LaFrance, J. T., Shimshack, J. P., and Wu, S. 2002. The Environmental Impacts of Subsidized Crop Insurance: Crop Insurance & the Extensive Margin.

Miranda, M.J.1991. Area‐yield crop insurance reconsidered. *Amer. J. Agr. Econ.*73(2): 233-242.

Miranda, M. J., and J. W. Glauber. 1997. "Systemic Risk, Reinsurance, and the Failure of Crop Insurance Markets." *Amer. J. Agr. Econ.* 79:206-215.

Nelson, C. H., & Loehman, E. T. (1987). Further toward a theory of agricultural insurance. *Amer. J. Agr. Econ.* 69(3): 523-531.

Nelson, C.1990. "The Influence of Distribution Assumptions of the Calculation of Crop Insurance Premia." *N. Cent. J. Agr. Econ.* 12:71-78.

Nelson, C., and P. Preckel.1989. "The Conditional Beta Distribution as a Stochastic Production Function." *Amer. J. Agr. Econ.* 71(2):370-378.

Norwood, B., M.C. Roberts and J.L. Lusk. 2004. Ranking crop yield models using out-of-sample likelihood functions. *Amer. J. Agr. Econ.* 86(4): 1032-1043.

O'Donoghue, E. 2014. The effects of premium subsidies on demand for crop insurance. USDA-ERS economic research report, (169).

Price, M.J., C.L. Yu, D.A. Hennessy and X. Du. 2019. Are actuarial crop insurance rates fair?: an analysis using a penalized bivariate B-spline method. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 68(5): 1207-1232.

Quiggin, J. C., G. Karagiannis and J. Stanton. 1993. Crop insurance and crop production: an empirical study of moral hazard and adverse selection. *Australian Journal of Agricultural*

Economics, 37(429-2016-29192), 95-113.

Ramirez, O. 1997. "Estimation and Use of a Multivariate Parametric Model for Simulating Heteroskedastic, Correlated, Nonnormal Random Variables: The Case of Corn Belt Corn, Soybean, and Wheat Yields." *Amer. J. Agr. Econ*. 79:191-205.

Ramirez, O.A. and J.S. Shonkwiler. 2017. A probabilistic model of the crop insurance purchase decision. *Journal of Agricultural and Resource Economics*: 10-26.

Ramirez, O., S. Misra and J. Field.2003. "Crop-Yield Distributions Revisited." *Amer. J. Agr. Econ.* 85(1): 108-120.

Robinson, P.M., 1987. Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form. *Econometrica*: *Journal of the Econometric Society*: 875-891.

Rosch, Stephanie. 2021. "Federal Crop Insurance: A Primer", Report No. R46686.

Rosa, I. 2018a. Federal crop insurance: Program overview for the 115th congress. Report R45193.

Rosa, I. 2018b. Farm bill primer: Federal crop insurance. Congressional Research Service.

Schmidt, U., C. Starmer and R. Sugden. 2008. Third-generation prospect theory. *Journal of Risk and Uncertainty,* 36(3): 203-223.

Schwarz, G.1978. "Estimating the Dimension of a Model." *Annals of Statis*. 6:461-464.

Sheather, S.J. and M.C. Jones. 1991. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society: Series B (Methodological),* 53(3): 683-690.

Sherrick, B.J., Zanini, F.C., Schnitkey, G.D. and Irwin, S.H., 2004. Crop insurance valuation under alternative yield distributions. *Amer. J. Agr. Econ.* 86(2): 406-419.

Skees, J.R., 1987. Future research needs on federal multiple peril crop insurance (No. 2096-2018-3251).

Skees, J. R., and Reed. M.R. 1986. Rate making for farm-level crop insurance: Implications for adverse selection. *Amer. J. Agr. Econ*. 68(3): 653-659.

Smith, V.H. and B.K. Goodwin. 1996. Crop insurance, moral hazard, and agricultural chemical use. *Amer. J. Agr. Econ.* 78(2): 428-438.

Swinton, S., and R. King.1991. "Evaluating Robust Regression Techniques for Detrending Crop Yield Data with Non-Normal Errors." *Amer. J. Agr. Econ.* 73:446-461.

Taylor, C. 1990. "Two Practical Procedures for Estimating Multivariate Nonnormal Probability Density Functions." *Amer. J. Agr. Econ.* 72: 210-217.

U.S. General Accounting Office. 1989. "A Disaster Assistance: Crop Insurance Can Provide
      Assistance More Effectively than Other Programs: GAO/RCED-89-211, September.

U.S. General Accounting Office. 1993. "A Crop Insurance: Federal Program Has Been Unable to
      Meet Objectives of 1980 Act." GAO/T-RCED-93-12, March 3.

US GAO.2014 Considerations in Reducing Federal Premium Subsidies. Report GAO-14-700.
      US Gov. Account. Off., Washington DC.

US GAO. 2015. In areas with higher crop production risks, costs are greater, and premiums may
      not cover expected losses. Report GAO-15-215. US Gov. Account. Off., Washington DC.

Vuong, Q. H. 1989. "Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses."
      Econometrica 57: 307-333.

Wood, S.N., N. Pya and B. Säfken. 2016. Smoothing parameter and model selection for general
      smooth models. *Journal of the American Statistical Association*, 111(516): 1548-1563.

Woodard, J.D., Sherrick, B.J. and Schnitkey, G.D., 2011. Actuarial impacts of loss cost ratio
      ratemaking in US crop insurance programs. *Journal of Agricultural and Resource
      Economics*: 211-228.

Woodard, J.D., Schnitkey, G.D., Sherrick, B.J., Lozano‑Gracia, N. and Anselin, L., 2012. A
      spatial econometric analysis of loss experience in the US crop insurance
      program. *Journal of Risk and Insurance*, 79(1):261-286.

Woodard, J.D. and Verteramo‑Chiu, L.J., 2017. Efficiency impacts of utilizing soil data in the
      pricing of the federal crop insurance program. *Amer. J. Agr. Econ.* 99(3): 757-772.

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

Yoshida, T., 2013. Asymptotics for penalized spline estimators in quantile
      regression. Communications in Statistics-Theory and Methods, (just-accepted).

Yu, J. and D.A. Sumner. 2018. Effects of subsidized crop insurance on crop
      choices. *Agricultural Economics*, 49(4): 533-545.

Zulauf, C., J. Coppess, G. Schnitkey and N. Paulson. 2018. Premium Subsidy and Insured US
      Acres: Differential Impact by Crop. *farmdoc daily*, 8.

Zhu, Y., B. K Goodwin., S. K. Ghosh. 2011. Modeling yield risk under technological change:
      Dynamic yield distributions and the US crop insurance program. *Journal of Agricultural
      and Resource Economics*: 192-210.

APPENDIX A: Resampling and Premium Estimation

Step 1: Detrended Yields

Let $y_{i,c,t}$ be the unit-level yield for farm i in county c and year t, $y_{c,t}$ be the county-level yield. Two log-linear trend equations will be estimated as (Miranda and Glauber 1997; Deng et al. 2007):

(A1.1) $\log(y_{i,c,t}) = \beta_0 + \beta_1(2009 - t) + \varepsilon;$

(A1.2) $\log(y_{c,t}) = \gamma_0 + \gamma_1(2009 - t) + u,$

where $t \in \{1970, \dots, 2008\}$ for unit-level yields and $t \in \{1951, \dots, 2018\}$ for county-level yields. The reason is that long-term county-level historical yields may represent the long-term variation and be more convincible. Then the detrend unit- and county-level yields will be calculated as:

(A1.3) $y_{i,c,t}^{det} = \frac{y_{i,c,t}}{\hat{y}_{i,c,t}} \times \hat{y}_{i,c,2009}$

(A1.4) $y_{c,t}^{det} = \frac{y_{c,t}}{\hat{y}_{c,t}} \times \hat{y}_{c,2009}$

where $\hat{y}_{i,c,t}$ is the predicted unit-level yield, and $\hat{y}_{c,t}$ is the predicted county-level yield. Both yields are adjusted to 2009 technological level.

Step 2: Semiparametric Quantile Regression Imputation (SQRI)

Observations in our yield sample are from 1970 to 2008 but mostly lie between 1999 and 2008. This may induce temporal bias since no extreme climate change happens in the 10-year time interval. Aiming to making the sample more representative for a long-term period, we adopted semi-parametric quantile regression imputation using penalized B-splines as in Chen and Yu (2016).

Consider a model for $(x_i, y_i)^T$, $i \in \{1, \dots, n\}$ that is a set of *i.i.d* observations of random variable (X, Y), where Y is the response variable that may be subject to missing, and X is an

univariate variable that can be observed. Let $q_\tau(x)$ be the unknown conditional $\tau - th$ quantile

of response Y given $X = x$. For a given $\tau \in (0,1)$, the conditional quantile function $q_\tau(x)$ is

defined as: $P(Y < q_\tau(x)|X = x) = \tau$. Then

(A1.5) $\qquad q_\tau(x) = \arg \min_{h(x)} E\{\rho_\tau(Y - h(x))|X = x\},$

where $\rho_\tau(u) = u \times (\tau - I(u < 0)) = \begin{cases} \tau|u| & u \geq 0 \\ (1-\tau) & u \leq 0 \end{cases}$ is the check function (see more in

Koenker and Bassett 1978).

Here $q_\tau(x)$ will be estimated with penalized B-spline and $h(x)$ will be represented vias

basis expansion. Let $K_n - 1$ be the number of knots within the support of X, and p be the degree

of B-splines. Define equidistantly located knots as $\kappa_k = K_n^{-1}k, k = -p+1,..., K_n + p$. As such, the

p-th B-spline basis is

(A1.6) $\qquad B(x) = \left(B_{-p+1}^{[p]}(x), B_{-p}^{[p]}(x), ..., B_{-p}^{[p]}(x) \right)^T$

where $B_k^{[p]}(x), k = -p + 1, ..., K_n$ are defined recursively as:

- For $s = 0$:

(A1.7) $\qquad B_k^{[0]}(x) = \begin{cases} 1, & \kappa_{k-1} < x < \kappa_k \\ 0, & \text{otherwise} \end{cases}$ where $k \in \{-p + 1, ..., K_n + p\}$;

- For $s \in \{1,2,...,p\}$:

(A1.8) $B_k^{[s]}(x) = \frac{x - \kappa_{k-1}}{\kappa_{k+s-1} - \kappa_{k-1}} B_k^{[s-1]}(x) + \frac{\kappa_{k+s} - x}{\kappa_{k+s} - \kappa_k} B_{k+1}^{[s-1]}(x),$

where $k \in \{-p + 1, ..., K_n + p - s\}$.

The estimated conditional quantile regression function is:

(A1.9) $\qquad \hat{q}_\tau(x) = B^T(x)\hat{b}(\tau)$

Where $\hat{b}(\tau)$ is a $(K_n + p) \times 1$ vector obtained by:

(A1.10) $\quad \hat{b}(\tau) = \arg\min\limits_{b(\tau)} \sum_{i=1}^{n} \delta_\tau \rho_\tau \left[y_i - B^T(x)\hat{b}(\tau) \right] + \frac{\lambda_n}{2} b^T(\tau) D_m^T D_m b(\tau),$

Where $\lambda_n (> 0)$ is the smoothing parameter, δ_τ is the "learning rate", D_m is the $m - th$ difference matrix and is $(K_n + p - m) \times (K_n + p)$ dimensional with its element defined as

(A1.11) $\quad d_{ij} = \begin{cases} (-1)^{|i-j|} \binom{m}{|i-j|}, & 0 \leq j - i \leq m \\ 0, & \text{otherwise} \end{cases},$

where $\binom{m}{k}$ is the choose function given by $(k!\,(m-k)!)^{-1} m!$; m is the order of penalty. The difference penalty $b^T(\tau) D_m^T D_m b(\tau)$ is to solve computational difficulty as in Yoshida (2013).

As shown above, smoothing penalties will vary as λ_n changes. From a Bayesian viewpoint, $\hat{b}(\tau)$ is a posterior mode for true parameter $b(\tau)$. As in Wood et al. (2016), the conditional distribution of $b(\tau)$ based on y is $b(\tau)|y \sim N\left(\hat{b}(\tau), \left(\Xi + (D_m^T D_m)^\lambda \right)^{-1} \right)$ where Ξ is the expected negative Hessian of the log-likelihood at $\hat{b}(\tau)$ and log-likelihood is $l\big(b(\tau), q_\tau(x)\big) = \log f\big(y|b(\tau), q_\tau(x)\big)$ where f_λ is given by $N\big(0, (D_m^T D_m)^{\lambda-}\big)$. Then smoothing parameters λ_n can be estimated by maximizing the log marginal likelihood

(A1.12) $\quad V(\lambda) = \log \int f\big(y|b(\tau)\big) f_\lambda(b(\tau)) db(\tau)$

Ideally, τ should be randomly drawn from [0,1]. To save computational time, the alternative procedure for semi-parametric quantile regression imputation is:

- Divided [0,1] into $J = 100$ equally distance sub-intervals;

- Obtain $\hat{b}(\tau_j)$ $j \in \{1, \dots, J-1\}$ by using best "learning rate" δ_i and the smoothing parameter λ_n as in Wood (2016);

- For a given detrended county-level yield $y_{c,t}$ where c represents county and t represent year, the imputed detrended unit-level yield at $\tau - th$ quantile is $y_{j,c,t}^{\text{det}*} = B^T(y_{c,t}^{\text{det}})\hat{b}(\tau_j)$. Randomly draw 1,000 observations from all years for one county. For instance, if $t \in$

45

$\{1970, \dots, 2008\}$, there are total 39 years and 3,861 (=39*99) imputed unit-level yield for this county, then 1,000 observations will be chosen.

Step 3: Rejection Sampling

Based on SQRI using penalized B-splines, we have 1,000 unit-level yields for each county in 12 states (IA, IL, IN, KS, MI, MN, MO, NE, ND, OH, SD and WI). The imputed yields in one county consist of the target density for resampling. Rejection sampling method will be adopted in this paper to get a more representative sample. The steps are as follows:

- All yields are adjusted to 2009 technological level. Let y_c^* be imputed unit-level yield for county c in the SQRI part and set the distribution $f(y_c^*)$ as the target distribution

- Let y_c be the detrended unit-level yield for county c with an auxiliary distribution $f(y_c)$[1]

- Set the "envelope constant" A, such that $0 < A < +\infty$ and multiply by the auxiliary distribution to create a "blanket function", $Af(y_c)$. The selection of $f(y_c)$ and A must satisfy the condition: $Af(y_c) \geq f(y_c^*)$

- Choose y_c by the criteria: when $Af(y_c) \geq f(y_c^*)$ then y_c will be accepted, when $Af(y_c) \leq f(y_c^*)$ then y_c will be rejected. For the sake of desired sample size, we will still randomly choose 1,000 observations for each county

Step 4: Actuarially Fair Premium

After the Steps 1-3 are finished, there is a more representative sample for each county and then each sample has 1,000 observations. Here we will calculate the actuarially fair premium based on the new sample and the procedure in Price et al. (2019). The main idea is to use a modified stochastic yield function:

(A1.13) $y_{i,c,t}^{\text{det}} = \alpha_0 + m(z_i) + \sigma(z_i)\varepsilon_{i,t}$

---

[1] Both distributions are obtained by "density" and "approxfun" in R with the default setting.

where $y_{i,c,t}^{\text{det}}$ is the detrended farm-level yield for farm i in county c in year t; $m(z_i)$ and $\sigma(z_i)$ are respectively yield mean and yield standard deviation functions which depend on the unit's four-to-ten-year average actual yield (APH) labelled as z_i; residual ε has zero mean and CDF $G(\varepsilon):[\underline{\varepsilon},\overline{\varepsilon}] \rightarrow [0,1]$. Here, the distribution of residuals is not restricted.

Let $E[\cdot]$ be the expectation operator, $\psi = [\varphi z - m(z)]/\sigma(z)$, the actuarially fair premium rate from buyer's perspective (i.e., expected indemnity) is:

(A1.14) $\qquad ULR(\varphi, z) = \frac{E[\max{(\varphi z - y, 0)}]}{\varphi z} = \frac{\sigma(z)E[\max{(\psi - \varepsilon, 0)}]}{\varphi z} = \frac{\sigma(z)}{\varphi z}\int_{\underline{\varepsilon}}^{\psi} G(\varepsilon)d\varepsilon,$

where ULR is County Unloaded Rate; y is actual yield for each unit; z is county yield in 2009; φ is coverage level.

It is clear that $\widehat{m}(z)$, $\widehat{\sigma}(z)$ and $\widehat{G}(\varepsilon)$ need to be estimated from the regression equation. Insured price p is calculated by averaging out February futures prices for upcoming year's harvest contracts (December contracts for corn and November contracts for soybean). The procedure to obtain those estimators is:

- Regress $y_{i,c,t}^{\text{det}}$ on z_i using the univariate penalized B-spline method to estimate coefficients for mean function $\widehat{m}(z)$;
- Extract residuals $\hat{r}_{i,t}$ from step 1 and regress $\hat{r}_{i,t}^2$ on z_i with the penalized B-spline method to estimate $\widehat{\sigma}(z)$;
- Standardize the residuals as $\hat{\varepsilon}_{i,t} = \hat{r}_{i,t}/\sigma(z)$ and obtain the empirical cumulative density function estimate $\widehat{G}(\varepsilon_{i,t})$[1];
- Integrate the cumulative density function estimate and obtain the actuarially fair premium for each (coverage level, APH) pair.

---

[1] See "ecdf" function in R.

Parameters are specified as: (a) degree of spline, p , is 3, i.e., cubic B-spline; (b) degree of penalty, m, is 2 as in Yoshida (2013); (c) number of knots is 5; (d) smoothing parameter λ is determined by GCV (generalized cross-validation).

Let z be median APH in resampling data, then we have reference rate ULR for all coverage levels. By a slight modification of Base Producer Rate (BPR) as in Coble and Goodwin (2010), the estimated actuarially fair BPR will be:

(A1.15) $\qquad \Gamma_i(\varphi, z_i) = \left(\frac{z_i}{\text{ref yield}_{i,c}}\right)^{-E_c} \times ULR_c(\varphi) + \text{fixed load} \times \text{cov diff}(\varphi)$

where z_i is Actual Production History of farm i; ref yield$_{i,c}$ is the reference yield in county c where farm i locates in; $-E_c$ is the negative exponent provided by RMA by which a farm with better land quality will have a low premium rate; $ULR_c(\varphi)$ is the county unloaded rate we have estimated; fixed load includes prevented planting rate load, replant rate load, quality adjustment rate load and state catastrophic rate load; cov diff is equal to 1 when $\varphi = 0.65$.

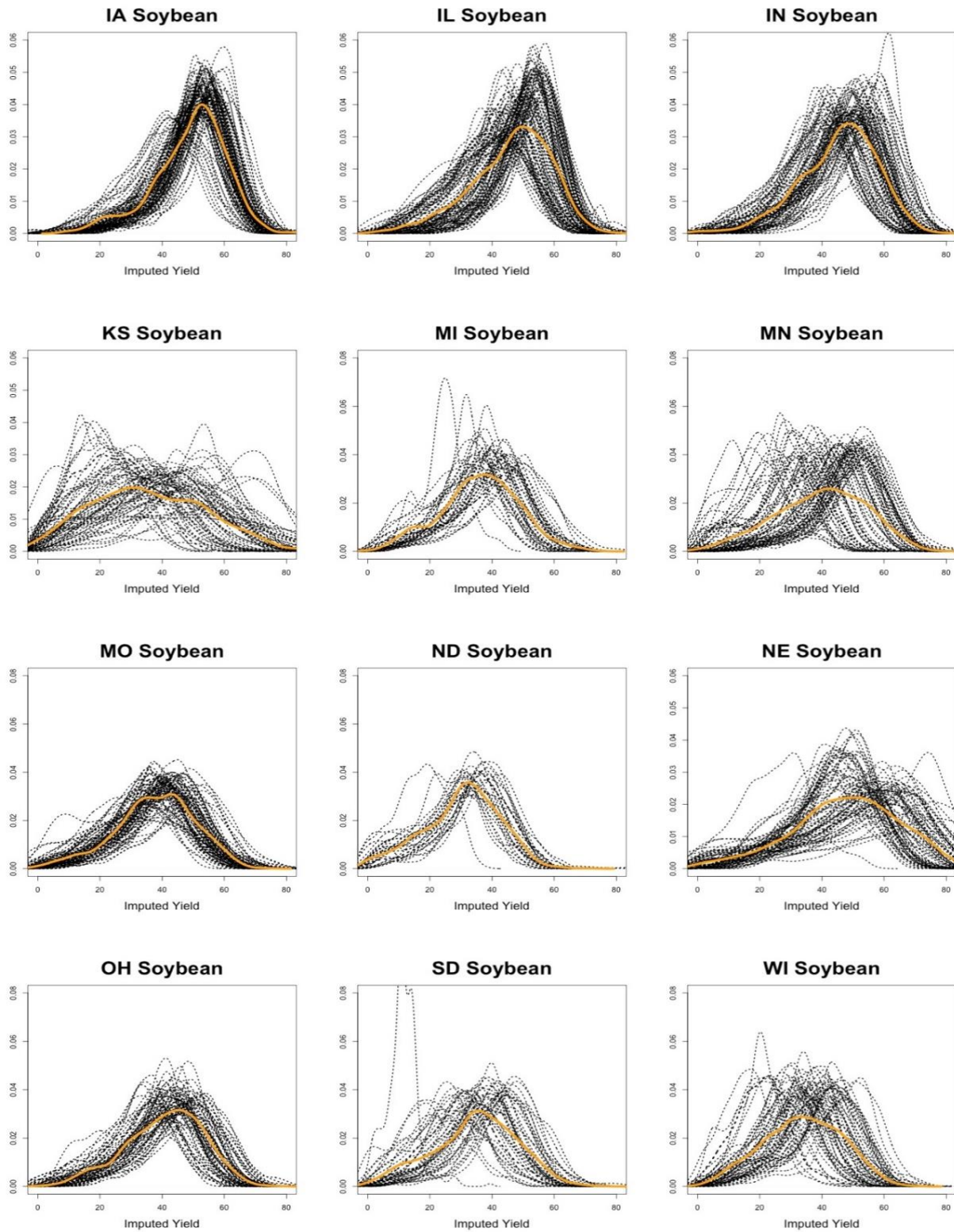**APPENDIX B: Supplemental Figures and Tables**

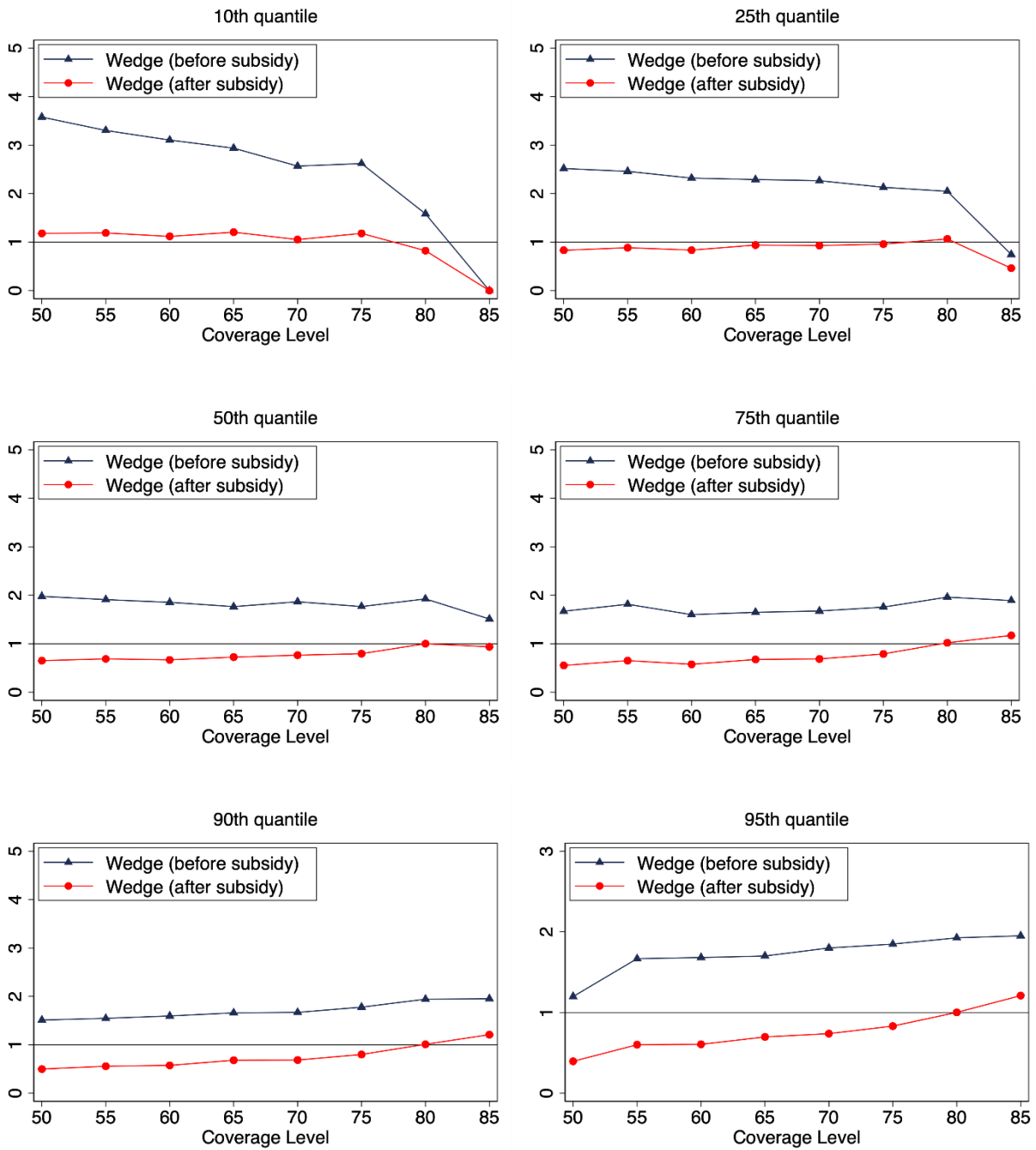Figure 1B.1 Imputed County-Level Densities for Soybeans.

Figure 1B.2 Comparisons of Wedges (before and after subsidies) among Different Land Qualities for Soybean.
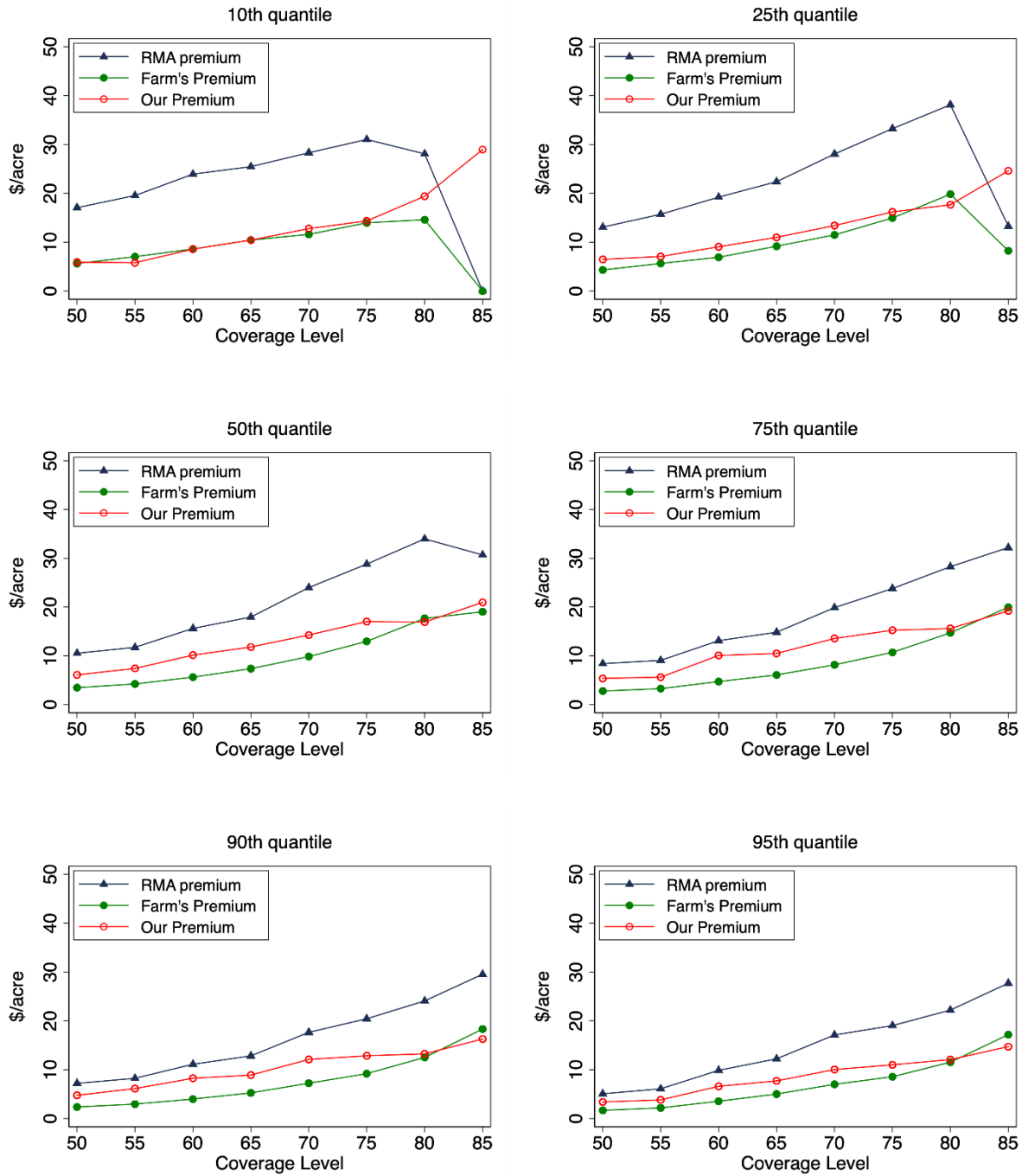
Figure 1B.3 RMA Premiums (including subsidies), Farm's Self-Paid Premium and Estimated Actuarially Fair Premiums for Soybean.
Note: Only mean values are reported here.

CHAPTER 2 Basis Risk and Farmers' Participation in the U.S. Federal Crop Insurance

Program: A Conceptual Framework and its Application

Abstract

Since 2000, both insured acres and coverage level choices in U.S. Federal Crop Insurance Program have been boosted due to generous subsidy policies. However, spatially heterogeneous participation patterns exist across the Great Plains and Corn Belt regions. Recently, the basis risk has been increasingly recognized as an essential driver for deterring the crop insurance uptake of weather index contracts. This study investigates whether the basis risk inversely affects the insurance demand for revenue and yield contracts. We build a conceptual model to explore farmers' acreage response to basis risk within an Expected Utility framework. With elevator-level basis risk, we apply Fractional Probit with Control Function and find that the effects of basis risk on participation rates are significantly negative for nearly all insurance contracts. Our analysis implies that: (i) to remove basis risk, revision for revenue contract may be considered; and (ii) subsidy structure may also be adjusted to be consistent with the underlying basis risk. The policy adjustments for basis risk might help save large subsidies while keeping high participation rates.

Introduction

Famers face multiple risks that can generate large fluctuations in their income. The U.S. government provides large subsidies to boost participation rates in the Federal Crop Insurance Program (FCIP) to shield farmers from risks. Crops and contracts have also been expanded to satisfy farmers' demand for diversification. The 2000 Agricultural Risk Protection Act (ARPA) provides $8.2 billion for expanding participation through increased premium subsidies (Goodwin 2001; Coble and Goodwin 2010). Unsurprisingly then, both extensive (share of insured acres in total insurable acres) and intensive (coverage level choice) margin participation rates have increased over the past twenty years (Che et al. 2020). However, one concern increases because significant regional disparities in insurance uptake exist, especially across the Great Plains and Corn Belt (Innes and Ardila 1994; LaFrance et al. 2002; Goodwin et al. 2004; Clark 2016; Jensen et al. 2018).

In this study, we focus on the effect of basis risk on the participation rate. As in the literature, the basis is defined as the difference between spot and futures prices, and then basis risk is the variation of the basis which can be defined as, for example, the standard deviation of the basis. Our motivation mainly comes from (i) basis risk is inherent in the revenue contracts because indemnity measurement is based on futures price, but farmer's sales revenue is based on cash price; (ii) crop revenue insurance, first introduced in the 1990s, has been the most popular product offered by Federal Crop Insurance Corporation (Du et al. 2014; Goodwin and Hungerford 2014); and (iii) many studies explore the effect of basis risk on insurance uptake focuses on weather index insurance products (Doherty and Richter 2002; Deng et al. 2007; Cole et al. 2014; Jensen et al. 2018; Cai et al. 2020; Gaurav and Chaudhary 2020; Lichtenberg and Martinez 2022; Ohashi 2022), but few studies investigate this topic using information from

the cash and futures markets.

There is considerable interest in whether basis risk impedes crop insurance uptake in the United States. Clarke (2016) provides a formal analysis of how and why downside basis risk can discourage farmers' purchase of index insurance and then argues that basis risk may pose a considerable obstacle to designing index insurance programs. The reason is that index insurance decreases farmers' income in the wrong states of nature (suffer losses but are not indemnified) and increases it in the good states of nature (do not suffer losses but are indemnified). Then risk-averse farmers have little incentive to purchase index insurance (Miranda and Farrin 2012; Jensen and Barrett 2017; Lichtenberg and Martinez 2022). In this study, the best state of nature (no basis risk) is a perfect match between cash and futures prices, and the worst state of nature (considerable basis risk) is that a fundamental mismatch happens. If basis risk exists, weakening (strengthening) basis indicates farmer's income will decrease (increase) compared to the best state of nature.

Spatial disparities of participation rates show that: *first*, extensive margin (share of insured acres in total insurable acres) is in general highest in North Dakota and decreases when moving from west-southwest toward the eastern Corn Belt; *second*, intensive margin (coverage level choice) is highest in the central and eastern Corn Belt, especially in central Illinois and northern Indiana, but decreases when moving outside this region. The linkage between participation rate and basis risk may exist because basis risk is low in the production center of gravity (e.g., Corn Belt region) and increases when moving outside the main production area (Figlewski 1984; Haushalter 2000; Doherty and Richter 2002).

Our analysis builds on and contributes to three main pieces of literature. First, our study provides insights on observed spatial disparities in the participation rate in U.S. Federal Crop

Insurance Program. Existing research has analyzed factors impacting insurance up-take, such as expected return to insurance (Gardner and Kramer 1986; Goodwin 1993), premium expenditures (Binswanger-Mkhize's 2012; Cole et al. 2013; Du et al. 2017), recency effects (Cole et al. 2014; Gallagher 2014; Kousky 2017; Stein 2018; Che et al. 2020), premium payment at harvest rather than planting time (Casaburi and Willis 2018; Liu et al. 2020) and farmer's Willingness-to-Pay (Feng et al. 2020). However, to our knowledge, little research has investigated the effect of basis risk (mismatch between cash and futures markets) on participation rates.

Second, our empirical results contribute to the literature on the effect of basis risk on participation rates. Regression results show that basis risk inversely affects the participation rates, regardless of the extensive or intensive margin of insurance demands. This conclusion is robust even though changing model specifications or controlling for more variables. Results imply that revising insurance contract design to eliminate the basis risk can improve participation rates. Furthermore, if farmers' indemnity is calculated based on cash rather than futures price, basis risk will vanish, and participation rates will increase. Therefore, eliminating basis risk supplements the subsidy structure, which might help save large subsidies while maintaining high participation rates.

Third, this study sheds light on the impact of basis risk on farmers' willingness-to-pay with the Monte Carlo simulation associated with the Multivariate Gaussian Copula (MGC) method. This method allows a dependence structure for multiple risk sources such as futures price, cash price, and actual yield. We employ unpublished farm-level insurance purchasing data maintained by USDA Risk Management Agency (RMA). Then we estimate the farmer's WTP ($/acre) at all individual coverage levels when assuming a constant absolute risk aversion (CARA) utility function (see, for example, Lapan and Moschini 1994; Babcock and Hennessy

1996; Du and Hennessy 2012; Feng *et al.* 2020). As in our conceptual framework, a farmer's

WTP should decrease when basis risk increases. The implication is that a farmer will be inclined

to insure fewer acreages under a high basis risk than a low risk. We assume that a state should be

ranked highly (towards 12) in participation rate when ranked lowly (towards 1) in basis risk.

Simulation results partly support our argument, but further analysis is needed because the risk-

averse coefficient is set as a constant across all states, which may not be the case in the real

world.

This paper is organized as follows. First, a conceptual model will be derived to explain

the effect of basis risk on the participation rate within a standard Expected Utility framework.

Then, some testable hypotheses will be provided for further investigations. A subsequent section

describes the data and approaches we employed to construct variables of interest, such as

participation rates and basis risks. In the empirical section, we report the empirical results. The

final section contains concluding comments.

A Conceptual Model

This section aims to better understand the effects of basis risk on both extensive (share of

insured acres in total insurable acres) and intensive (coverage level) margin participation rates.

As in Feng et al. (2020), within the expected utility setting, a farmer's WTP for insured acres (or

coverage level) can be implicitly estimated by equating the expected value of the objective

function value with and without insurance products. Such WTP is equivalent to a farmer's

desirability to enroll in the crop insurance program. Given a coverage level and a pre-determined

USDA/RMA premium, a high (low) WTP can raise (deter) crop insurance demand, which also

indicates that extensive margin participation rate will go up (down).

Revenue Protection with the Harvest Price Exclusion (RPHPE) is a simple revenue

contract since the revenue guarantee is based on the projected price only, and so we commence with analyzing this contract form. At harvesting time, two-income sources should be realized: (a) income from crop sales (cash price × actual yield); and (b) indemnity payment after any crop loss. If actual income is lower than the revenue guarantee, the difference will be paid from the insurer to the insured, or 0 otherwise. Denote $p(\varphi)$ as insurance premium price and c as production cost, the farmers' net revenue after purchasing a corn crop revenue insurance product is:

$$(2.1) \quad \tilde{R}^{\text{In}} = -c - p(\varphi) + \overbrace{\tilde{P}_{\text{Oc}}^{c}\tilde{y}}^{\substack{\text{Revenue in} \\ \text{local market}}} + \overbrace{max\left[\phi F_{\text{De,Fe}} y^{APH} - \tilde{F}_{\text{De,Oc}}\tilde{y}, 0\right]}^{\substack{\text{Indemnity based on futures} \\ \text{prices and actual yields}}},$$

where tilde indicates variables random at planting time, $\tilde{P}_{\text{Oc}}^{c}$ is the local cash price when the crop is sold, which is presumed to be October, $\tilde{F}_{\text{De,Oc}}$ is the harvest contract futures price,[8] $F_{\text{De,Fe}}$ is the Springtime price (i.e., the projected price) for corn December futures contract,[9] $\tilde{y}$ is actual yield in the harvesting time, ϕ is coverage level with $\phi \in \{0.5, \dots, 0.85\}$, and y^{APH} is the approved production history (APH) for each farm. With $\tilde{B}_{\text{Oc}}^{c} = \tilde{P}_{\text{Oc}}^{c} - \tilde{F}_{\text{De,Oc}}$ as the basis in October in county c, then (2.1) may be rewritten as

$$(2.2) \quad \tilde{R}^{\text{In}} = -c - p(\varphi) + \tilde{B}_{\text{Oc}}^{c}\tilde{y} + \tilde{F}_{\text{De,Oc}}^{c}\tilde{y} + max\left[\phi F_{\text{De,Fe}} y^{APH} - \tilde{F}_{\text{De,Oc}}\tilde{y}, 0\right]$$

$$= -c - p(\varphi) + \tilde{B}_{\text{Oc}}^{c}\tilde{y} + max\left[\phi F_{\text{De,Fe}} y^{APH}, \tilde{F}_{\text{De,Oc}}\tilde{y}\right]$$

$$= -c - p(\varphi) + \begin{cases} \tilde{B}_{\text{Oc}}^{c}\tilde{y} + \tilde{F}_{\text{De,Oc}}\tilde{y}, & \text{whenever no loss;} \\ \tilde{B}_{\text{Oc}}^{c}\tilde{y} + \phi F_{\text{De,Fe}} y^{APH}, & \text{otherwise;} \end{cases}$$

To account for basis risk, we assume that basis follows a location-scale family with $\tilde{B}_{\text{Oc}}^{c}(\sigma) \equiv \bar{B} + \tilde{\varepsilon}\sigma$ where $E[\tilde{\varepsilon}|\tilde{y}] = 0$ and σ scales basis risk such that risk vanishes whenever

---

[8] The Harvest price for a revenue contract is usually defined as the average of futures contract price during October. For more information, please refer to: https://www.extension.iastate.edu/agdm/crops/html/a1-54.html.
[9] For soybean, the futures contract for new crop is November.

$\sigma = 0$ (Sandmo 1971; Meyer 1987; Meyer and Rasche 1992). Requirement $E[\tilde{\varepsilon}|\tilde{y}] = 0$ is the independent background risk assumption. Re-writing Eqn. (2.2) then we have

(2.3) $\quad \tilde{R}^{\text{In}} = -c - p(\varphi) + \tilde{B}_{\text{Oc}}^{c}(\sigma)\tilde{y} + \max[\phi F_{\text{De,Fe}} y^{APH}, \tilde{F}_{\text{De,Oc}}\tilde{y}]$

$\quad\quad\quad = -c - p(\varphi) + (\bar{B} + \tilde{\varepsilon}\sigma)\tilde{y} + \max[K, \tilde{F}_{De,Oc}\tilde{y}]; \quad\quad K \equiv \phi F_{\text{De,Fe}} y^{APH}$

We may ask what is WTP to remove basis risk? Then WTP as a function of basis risk and coverage level, $W(\sigma; \phi)$, may be defined implicitly as

(2.4) $\quad E\left[U\left([\bar{B} + \tilde{\varepsilon}\sigma]\tilde{y} + \max[K, \tilde{F}_{\text{De,Oc}}\tilde{y}] - W(\sigma; \phi) - c - p(\phi)\right)\right]$

$\quad\quad\quad = E\left[U\left(\bar{B}\tilde{y} + \max[K, \tilde{F}_{\text{De,Oc}}\tilde{y}] - c - p(\phi)\right)\right],$

where independent background risk (Gollier and Pratt, 1996) affects both sides through random yield. A crucial point to note is that if the distributions of $\tilde{F}_{\text{De,Oc}}$ and $\tilde{y}$ do not change then the right-hand term is constant at, say, value κ. Eqn. (2.3) provides a key observation for understanding the role of basis risk.

Observation 1: *For RPHPE, the expected value of the outside option of no insurance is independent of basis risk while the expected value of RPHPE depends on basis risk only through its effect on basis times yield.*

Thus, we may write

(2.5) $\quad E\left[U\left([\bar{B} + \tilde{\varepsilon}\sigma]\tilde{y} + \max[K, \tilde{F}_{\text{De,Oc}}\tilde{y}] - W(\sigma; \phi) - c - p(\phi)\right)\right] = \kappa$

Applying the implicit function theorem, we obtain:

(2.6) $\quad E\left[U'\left(\tilde{R}^{\text{In}} - W(\sigma; \phi)\right)\tilde{\varepsilon}\tilde{y}\right] - E\left[U'\left(\tilde{R}^{\text{In}} - W(\sigma; \phi)\right)\right]\frac{dW(\sigma;\phi)}{d\sigma} = 0,$

and so, using the covariance relation $E[x_1 x_2] = E[x_1]E[x_2] + \text{Cov}(x_1, x_2)$, it follows that

(2.7) $\quad \dfrac{dW(\sigma;\phi)}{d\sigma} = \dfrac{E\left[U'\left(\tilde{R}^{\text{In}}-W(\sigma;\phi)\right)\tilde{\varepsilon}\tilde{y}\right]}{E\left[U'\left(\tilde{R}^{\text{In}}-W(\sigma;\phi)\right)\right]} = \dfrac{E\left\{E\left[U'\left(\tilde{R}^{\text{In}}-W(\sigma;\phi)\right)\tilde{y}|\tilde{y}\right]\right\}\overbrace{E\{E[\tilde{\varepsilon}|\tilde{y}]\}}^{=0}}{E\left[U'\left(\tilde{R}^{\text{In}}-W(\sigma;\phi)\right)\right]}$

$$+\frac{E\left\{Cov\left[U'\left(\tilde{R}^{\text{In}}-W(\sigma;\phi)\right)\tilde{y},\tilde{\varepsilon}|\tilde{y}\right]\right\}}{E\left[U'\left(\tilde{R}^{\text{In}}-W(\sigma;\phi)\right)\right]}=\frac{E\left\{Cov\left[U'\left(\tilde{R}^{\text{In}}-W(\sigma;\phi)\right)\tilde{y},\tilde{\varepsilon}|\tilde{y}\right]\right\}}{E\left[U'\left(\tilde{R}^{\text{In}}-W(\sigma;\phi)\right)\right]},$$

where the $|\tilde{z}$ notation indicates conditional on $\tilde{z}$.

Observe that $d\tilde{\varepsilon}/d\tilde{\varepsilon} > 0$ while $d\left[U'\left(\tilde{R}^{\text{In}} - W(\sigma;\phi)\right)\tilde{y}\right]/d\tilde{\varepsilon} = U''\left(\tilde{R}^{\text{In}} - W(\sigma;\phi)\right)\sigma\tilde{y} \le 0$ whenever $U''(\cdot) \le 0$. Therefore, by the covariance rule (Gollier 2001, p. 32), $Cov\left[U'\left(\tilde{R}^{\text{In}} - W(\sigma;\phi)\right)\tilde{y}, \tilde{\varepsilon}|\tilde{y}\right] \le 0 \;\; \forall\tilde{y}$ and so $E\left\{Cov\left[U'\left(\tilde{R}^{\text{In}} - W(\sigma;\phi)\right)\tilde{y}, \tilde{\varepsilon}|\tilde{y}\right]\right\} \le 0.$

Therefore,

(2.8) $\quad \dfrac{dW(\sigma;\phi)}{d\sigma} \le 0,$

a strong inference given the presence of independent background risk. Eqn.(2.8) indicates that farmers' WTP will decrease as basis risk increase if the coverage level is constant.

Hypothesis 1: *Given a coverage level, a farmer's WTP for crop insurance will decrease whenever basis risk increases.*

Standard revenue insurance (without harvest price exclusion, and labelled as RP) may be written as:

(2.9) $\quad \tilde{R}^{\text{In}} = -c - p(\varphi) + \tilde{B}^c_{\text{Oc}}(\sigma)\tilde{y} + \max\left[\phi\max\left[F_{\text{De,Fe}}, \tilde{F}_{\text{De,Oc}}\right]y^{APH}, \tilde{F}_{\text{De,Oc}}\tilde{y}\right]$

$\quad\quad = -c - p(\varphi) + (\bar{B} + \tilde{\varepsilon}\sigma)\tilde{y} + \max\left[\phi\max\left[F_{\text{De,Fe}}, \tilde{F}_{\text{De,Oc}}\right]y^{APH}, \tilde{F}_{\text{De,Oc}}\tilde{y}\right].$

Specification Eqn.(2.9) shows that Observation 1 applies also for RP in that basis risk matters only through its effect on basis times yield. In summary then

Proposition 1: *All else equal, under expected utility preferences whenever a location has larger futures market basis risk as measured by σ then willingness to pay for revenue insurance will be smaller. This is true for both standard revenue insurance and revenue insurance with harvest price exclusion.*

A natural question to ask is whether basis risk has the same role for yield insurance. Yield insurance indemnifies yield shortfalls at a fixed price rather than at some measure of market price. Letting the fixed price be the February price for the December maturity contract (for corn) then total revenues become

$$(2.10)\quad \tilde{R}^{\mathrm{In}} = -c - p(\varphi) + \overbrace{\tilde{P}^c_{\mathrm{Oc}}\tilde{y}}^{\substack{\text{Revenue in}\\\text{local market}}} + \overbrace{\max\left[\phi F_{\mathrm{De,Fe}} y^{APH} - \tilde{F}_{\mathrm{De,Oc}}\tilde{y}, 0\right]}^{\substack{\text{Indemnity based on futures}\\\text{prices and actual yields}}}$$

$$= -c - p(\varphi) + \tilde{B}^c_{\mathrm{Oc}}\tilde{y} + \tilde{F}_{\mathrm{De,Oc}}\tilde{y} + F_{\mathrm{De,Fe}}\max[\phi y^{APH} - \tilde{y}, 0]$$

$$= -c - p(\varphi) + \tilde{B}^c_{\mathrm{Oc}}\tilde{y} + \left(\tilde{F}_{\mathrm{De,Oc}} - F_{\mathrm{De,Fe}}\right)\tilde{y} + F_{\mathrm{De,Fe}}\max[\phi y^{APH}, \tilde{y}]$$

$$= -c - p(\varphi) + \begin{cases} \tilde{B}^c_{\mathrm{Oc}}\tilde{y} + \tilde{F}_{\mathrm{De,Oc}}\tilde{y}, & \text{whenever no loss;} \\ \tilde{B}^c_{\mathrm{Oc}}\tilde{y} + \left(\tilde{F}_{\mathrm{De,Oc}} - F_{\mathrm{De,Fe}}\right)\tilde{y} + \phi F_{\mathrm{De,Fe}} y^{APH}, & \text{otherwise;} \end{cases}$$

Observation 1 applies, but with the addendum that a random number of futures contract payoffs $\left(\tilde{F}_{\mathrm{De,Oc}} - F_{\mathrm{De,Fe}}\right)\tilde{y}$ also enters grower returns in addition to a revenue guarantee.

Does the distaste for basis risk apply when preferences are other than those under the expected utility paradigm? As we shall show below, there is reason to believe so. Consider the version of prospect theory most relevant to insurance choices, third generation prospect theory (Schmidt, Starmer and Sugden 2008) where the point of reference is a contract absent basis risk. Loss relative to the reference point will then be basis risk times yield, $\tilde{B}^c_{\mathrm{Oc}}(\sigma)\tilde{y} = (\bar{B} + \tilde{\varepsilon}\sigma)\tilde{y}$, as given above when $\sigma = 1$. Following a modification of Firpo et al. (2020), we suppose that the value function, $v(x)$, for relative gains/losses is differentiable on its domain of reals, $x \in \mathfrak{R}$, and satisfies:

(2.11) 1. $v(x)x \geq 0$ $\forall x \in \mathfrak{R}$ with $v(0) = 0$ (preferences for gains over losses);

       2. $v'(x) \geq 0$ $\forall x \in \mathfrak{R}$ (monotonicy);

       3. $-v(-x) \geq v(x)$ $\forall x > 0$ (loss aversion).

The first condition separates the loss and gain domains while the second states that more profit is preferred to less. The third condition, which replaces $v'(-x) \geq v'(x)$ in Firpo et al. (2020), codifies loss aversion because, loss of a given magnitude is more detrimental than a gain of the same magnitude is beneficial. With joint distribution function $G(\tilde{\varepsilon}, \tilde{y})$, then expected value is

(2.12) $\quad W[v(\cdot); G(\tilde{\varepsilon}, \tilde{y})]|_{\sigma=1} = \int v(\tilde{\varepsilon}\tilde{y}) d(\tilde{\varepsilon}, \tilde{y}).$

Suppose then that basis is independent of yield and the distribution of basis is symmetric around zero. In light of point 3. in Eqn. (2.11) above it can be readily shown that, for each yield outcome and each basis outcome, loss in the loss domain dominates the equally probable gain in the gain domain so that basis risk depresses expected value.

Returning to Eqn. (2.7) above, and so to the RPHPE contract, we now ask how the WTP response to basis risk is affected by coverage level. In order to do so we must ask how premium is affected by coverage level. Supposing that the premium is actuarially fair, with payout $\max[\phi F_{\text{De,Fe}} y^{APH} - \tilde{F}_{\text{De,Oc}} \tilde{y}, 0]$ and with joint distribution function $G(\tilde{F}_{\text{De,Oc}}, \tilde{y})$, the premium function is

(2.13) $\quad p(\phi) = E\left[\max[\phi F_{\text{De,Fe}} y^{APH} - \tilde{F}_{\text{De,Oc}} \tilde{y}, 0]\right]$

$$= \int_0^{\phi F_{\text{De,Fe}} y^{APH}} (\phi F_{\text{De,Fe}} y^{APH} - \tilde{F}_{\text{De,Oc}} \tilde{y}) \, dG(\tilde{F}_{\text{De,Oc}}, \tilde{y}),$$

where the lower and upper limits of the integral are 0 and $\phi F_{\text{De,Fe}} y^{APH}$ respectively. Eqn.(2.13) indicates that the indemnity $\phi F_{\text{De,Fe}} y^{APH} - \tilde{F}_{\text{De,Oc}} \tilde{y}$ will be paid by insurers to farmers when farmer's actual revenue in the market $\tilde{F}_{\text{De,Oc}} \tilde{y}$ is less than the guaranteed income $\phi F_{\text{De,Fe}} y^{APH}$. In this sense, actuarially fair premium $p(\phi)$ should be equal to the expected payout. So that the derivative with respect to ϕ is:

(2.14) $p'(\phi) = F_{\text{De,Fe}} y^{APH} \text{prob}[\phi F_{\text{De,Fe}} y^{APH} \geq \tilde{F}_{\text{De,Oc}} \tilde{y}]$.

Eqn.(2.14) indicates that premium rate monotonically increases as coverage level increases because the probability of revenue loss is always non-negative. An extreme case is that farmer's revenue is always more than guaranteed income of the revenue contract, then there will be no insurance demand. A differentiation of Eqn. (2.6), and using identity (2.2) with $\tilde{R}^{\text{net}} = \tilde{R}^{\text{In}} - W(\sigma; \phi)$, generates:

$$(2.15) \quad \frac{d^2 W(\sigma;\phi)}{d\sigma d\phi} = \frac{1}{E[U'(\tilde{R}^{\text{net}})]} \frac{dE[U'(\tilde{R}^{\text{net}})\tilde{\varepsilon}\tilde{y}]}{d\phi} - \frac{E[U'(\tilde{R}^{\text{net}})\tilde{\varepsilon}\tilde{y}]}{\{E[U'(\tilde{R}^{\text{net}})]\}^2} \frac{dE[U'(\tilde{R}^{\text{net}})]}{d\phi}$$

$$= \frac{1}{E[U'(\tilde{R}^{\text{net}})]} \frac{dE[U'(\tilde{R}^{\text{net}})\tilde{\varepsilon}\tilde{y}]}{d\phi} - \frac{1}{E[U'(\tilde{R}^{\text{net}})]} \frac{dW(\sigma;\phi)}{d\sigma} \frac{dE[U'(\tilde{R}^{\text{net}})]}{d\phi}$$

$$\overset{\text{sign}}{\triangleq} \frac{dE[U'(\tilde{R}^{\text{net}})\tilde{\varepsilon}\tilde{y}]}{d\phi} - \frac{dW(\sigma;\phi)}{d\sigma} \frac{dE[U'(\tilde{R}^{\text{net}})]}{d\phi}$$

Now

$$(2.16) \quad \frac{d[\tilde{R}^{\text{net}}]}{d\phi} = \begin{cases} F_{\text{De,Fe}} y^{APH} - \frac{dW(\sigma;\phi)}{d\sigma} - p'(\phi), & \text{whenever } F_{\text{De,Fe}} y^{APH} > \tilde{F}_{\text{De,Oc}} \tilde{y}; \\ \text{undefined}, & \text{whenever } F_{\text{De,Fe}} y^{APH} = \tilde{F}_{\text{De,Oc}} \tilde{y}; \\ -\frac{dW(\sigma;\phi)}{d\sigma} - p'(\phi), & \text{otherwise.} \end{cases}$$

With $I[A]$ as an indicator function conditional on A and with $\tilde{R}^{\text{nor}} \equiv \tilde{F}_{\text{De,Oc}} \tilde{y} / F_{\text{De,Fe}} y^{APH}$, Eqn.(2.15) may be written as

$$(2.17) \quad \frac{d^2 W(\sigma;\phi)}{d\sigma d\phi} \overset{\text{sign}}{\triangleq} E\left[U''(\tilde{R}^{\text{net}})\tilde{\varepsilon}\tilde{y}\left(F_{\text{De,Fe}} y^{APH} I[\phi \geq \tilde{R}^{\text{nor}}] - \frac{dW(\sigma;\phi)}{d\sigma} - p'(\phi)\right)\right]$$

$$- \frac{dW(\sigma;\phi)}{d\sigma} E\left[U''(\tilde{R}^{\text{net}})\left(F_{\text{De,Fe}} y^{APH} I[\phi \geq \tilde{R}^{\text{nor}}] - \frac{dW(\sigma;\phi)}{d\sigma} - p'(\phi)\right)\right]$$

$$= F_{\text{De,Fe}} y^{APH} E\left[U''(\tilde{R}^{\text{net}})\tilde{\varepsilon}\tilde{y}\left(I[\phi \geq \tilde{R}^{\text{nor}}] - \frac{dW(\sigma;\phi)}{d\sigma} \frac{1}{F_{\text{De,Fe}} y^{APH}} - \text{prob}[\phi \geq \tilde{R}^{\text{nor}}]\right)\right]$$

$$- F_{\text{De,Fe}} y^{APH} \frac{dW(\sigma;\phi)}{d\sigma} E\left[U''(\tilde{R}^{\text{net}})\left(I[\phi \geq \tilde{R}^{\text{nor}}] - \frac{dW(\sigma;\phi)}{d\sigma} \frac{1}{F_{\text{De,Fe}} y^{APH}} - \text{prob}[\phi \geq \tilde{R}^{\text{nor}}]\right)\right]$$

$$\stackrel{\text{sign}}{\cong} E\left[U''(\tilde{R}^{\text{net}})\tilde{\varepsilon}\tilde{y}\left(I[\phi \geq \tilde{R}^{\text{nor}}] - \frac{dW(\sigma;\phi)}{d\sigma}\frac{1}{F_{\text{De,Fe}}y^{APH}} - \text{prob}[\phi \geq \tilde{R}^{\text{nor}}]\right)\right]$$

$$-\frac{dW(\sigma;\phi)}{d\sigma}E\left[U''(\tilde{R}^{\text{net}})\left(I[\phi \geq \tilde{R}^{\text{nor}}] - \frac{dW(\sigma;\phi)}{d\sigma}\frac{1}{F_{\text{De,Fe}}y^{APH}} - \text{prob}[\phi \geq \tilde{R}^{\text{nor}}]\right)\right]$$

If farmers choose coverage level to maximize expected utility, where any coverage level were possible, and if Constant Absolute Risk Aversion applied,

$$-U''\left(\tilde{R}^{\text{In}} - W(\sigma;\phi)\right)\Big/U'\left(\tilde{R}^{\text{In}} - W(\sigma;\phi)\right) = \text{positive constant, then the last expectation would}$$

equal zero in light of the first-order condition.[10] To see this, return to Eqn.(2.4) and solve

(2.18) $\displaystyle\max_{\phi} E\left[U\left(\bar{B}\tilde{y} + \max[\phi F_{\text{De,Fe}}y^{APH}, \tilde{F}_{\text{De,Oc}}\tilde{y}] - c - p(\phi)\right)\right]$

with first-order condition that

(2.19) $E\left[U'(\tilde{R}^{\text{In}}) \times \left(I[\phi \geq \tilde{R}^{\text{nor}}] - \text{prob}[\phi \geq \tilde{R}^{\text{nor}}]\right)\right] = 0$

So under optimality for coverage level choice and constant absolute risk aversion then Eqn. (2.14) resolves to

(2.20) $\dfrac{d^2 W(\sigma;\phi)}{d\sigma d\phi} \stackrel{\text{sign}}{\cong} E\left[U''\left(\tilde{R}^{\text{In}}\right)\tilde{\varepsilon}\tilde{y}\left(I[\phi \geq \tilde{R}^{\text{nor}}] - \text{prob}[\phi \geq \tilde{R}^{\text{nor}}]\right)\right]$

As this expression involves multiple random variables and as two terms, $\tilde{\varepsilon}$ and $I[\phi \geq \tilde{R}^{\text{nor}}] - \text{prob}[\phi \geq \tilde{R}^{\text{nor}}]$, are not uniform in sign, establishing a sign on the expectation is challenging.

Remark 1: *Farmer's coverage level choice is undetermined when basis risk increases.*

The relationship between basis risk and coverage level choice may not be robust.

---

[10] There is strong evidence that they do not. See Du et al. (2017), <u>Rationality of choices in subsidized crop insurance markets</u>, X Du et al., American Journal of Agricultural Economics 99 (3), 732-756.
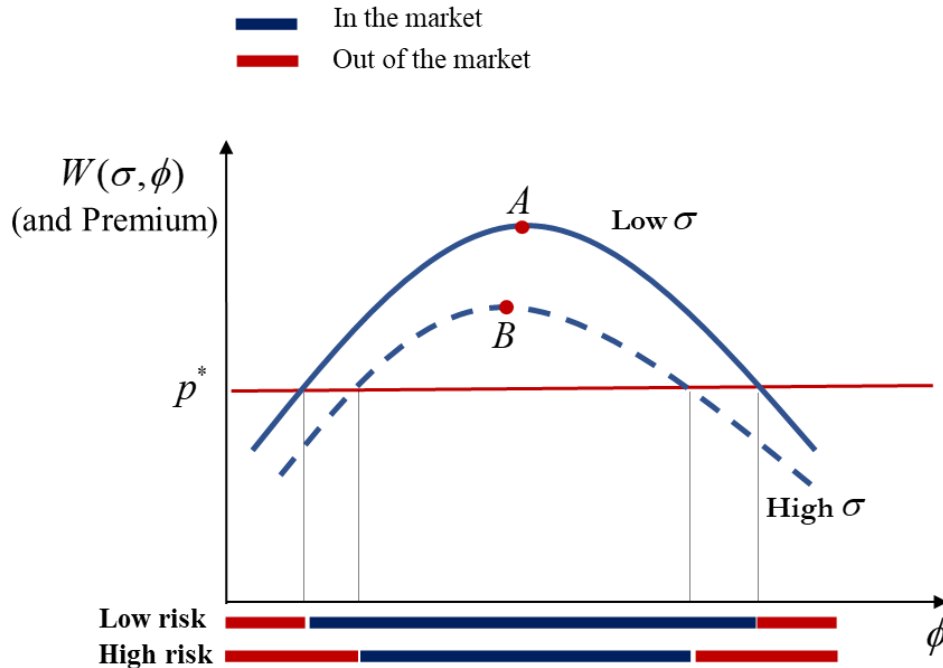
Figure 2.1 Assumptions for the Effect of Basis Risk on Participation Rates.
Note: the y-axis represents the farmer's willingness to pay and RMA's premiums, both of which are measured as \$/acre. The x-axis represents the farmer's coverage level choice. p^* represents a predetermined premium. The upper (lower) curve represents farmers' willingness to pay under a low (high) basis risk condition. "In the market" and "Out of the market" indicate whether farmers will participate in the federal crop insurance program. "A" and "B" show the optimized coverage choices and corresponding WTPs under different conditions, but the two coverage choices are not necessarily the same.

Figure 2.1 explains farmers' acreage response when they face the basis risk. First, we set the farmer's WTP as a quadratic coverage level function, implying that farmers prefer medium coverage levels over end levels. The set-up is consistent with participation rates in the real world. Second, farmers' WTP decreases when basis risk increases (low $\sigma \rightarrow$ high σ), given a coverage level. It is noteworthy that the optimal coverage choices under different basis risks might not be identical. For example, the optimal choice A may be on the left or right side of B. Third, if the premium rate is set as p^*, the blue area's ratio to total area (sum of blue and red) decreases when basis risk increases (see bars under the x-axis). The downside ratio of the blue area indicates that

a farmer will choose to insure fewer acres in the market. An extreme case is that if a farmer's WTP at the optimal coverage level is lower than the premium rate p^*, this farmer will eventually move out of the market and insure zero acres.

Data Description

This study focuses on participation rates for Buy-Up contracts in U.S Federal Crop Insurance Program. Unlike the CAT contract, the Buy-Up contracts allow farmers to choose one among multiple coverage levels. The targeted area contains 12 Midwest and Great Plain states (IL, IN, IA, KS, MI, MN, MO, NE, ND, OH, SD, and WI). Two dependent variables of most interests in the empirical analysis are: (i) the share of insured acres to total insurable acres (e.g., planted acres), which represents the extensive margin participation rate; and (ii) acreage-weighted coverage level choices, which represents the intensive margin participation rate. An extensive margin participation rate for yield and revenue contracts cannot be obtained directly because total insurable acres were not reported.

Our primary research question is this: does basis risk affect participation rates? One of our main works is constructing county-level basis risk based on elevator-level daily basis data. The reason why elevator-level data are employed is that various county-level basis risks can be constructed and compared. We follow the literature and include several variables intended to control land capability (the share of good land in total land), Growing Degree Days (GDD), Stress Degree days (SDD), yield risk, and geographic location. A general form of the model specification for both participation rates (extensive and intensive margins) is:

(2.21) $\text{Participation Rate}_{lc} = f \begin{pmatrix} \text{Basis Risk}_{lc}, \text{Land Capability}_{lc}, \text{Growing Degree Days}_{lc} \\ \text{Stress Degree Days}_{lc}, \text{Yield CV}_{lc}, \text{Geographic Location}_c \end{pmatrix}$,

where $f(\cdot)$ is a function mapping the effect of drivers on participation rates.

Table 2.1 Definition and Summary for Main Variables.

Description	Data Source	Crop	Variable	Obs.	Mean	Std.Dev
Participation Rate						
Share of insured acres in total insurable acres	NASS, FSA, SOB	Corn	e^1	775	0.85	0.12
			e^2	775	0.82	0.13
			e^3	775	0.83	0.13
		Soybeans	e^1	735	0.82	0.13
			e^2	735	0.81	0.12
Acreage-weighted average coverage level	SOB	Corn	ϕ_{yi}	775	0.70	0.06
			ϕ_{re}	775	0.76	0.04
		Soybeans	ϕ_{yi}	735	0.70	0.07
			ϕ_{re}	735	0.76	0.04
Basis						
Normalized Basis risk	Bids Data	Corn	σ	747	0.006	0.005
		Soybeans	σ	728	0.004	0.004
Elevator amount per county	Bids Data	Corn	$EleAmt$	775	5.17	4.27
		Soybeans	$EleAmt$	735	4.87	4.00
Years of elevator records per county	Bids Data	Corn	$AveYear$	775	3.90	1.50
		Soybeans	$AveYear$	735	3.85	1.50
Yield Risk and Stock						
Coefficient of Variation	NASS	Corn	YR	775	0.18	0.07
		Soybeans	YR	735	0.18	0.06
Ratio of Ending Stock to Production	NASS	Corn	$RaES$	775	0.56	0.06
		Soybeans	$RaES$	735	0.46	0.01
Land Capability						
Percentage of acres for Class I-II in total acres for Class I-VIII	NRI		LCC	2,902	0.25	0.21
Weather Determinant						
Growing Degree Days	NOAA		$\bar{G}$	1,016	1248.63	194.56
Stress Degree Days	NOAA		$\bar{S}$	1,016	20.06	22.95

Note: (1) NASS: National Agricultural Statistics Service; (2) FSA: Farm Service Agency; (3) SOB: USDA Summary of Business; (4) NRI: National Resource Inventory; (5) NOAA: National Oceanic and Atmospheric Administration. Bids Data represent elevator-level daily basis data which are purchased from a company.

As discussed above, the effect of basis risk on extensive margin participation rate should

be negative, but that on intensive margin is unclear. The effects of production risks (i.e., weather

determinants and yield risk) are expected to be contrary to extensive and intensive margins. The contrary cases happen because exacerbation of production conditions impacts farmers' trade-off between extensive and intensive margins. For example, bad weather conditions encourage farmers to enroll in the crop insurance program, but the coverage level might be low because farmers' subjective probability of catastrophe may be high. Therefore, for an extensive (intensive) margin participation rate, the effect of GDD is expected to be negative (positive); however, those of SDD and yield risk are expected to be positive (negative).

Furthermore, we will not consider the recency effect of loss events because all variables in this study are aggregate measures from 2009 to 2020. The main reasons for cross-sectional rather than panel analysis are: *first*, basis variation in a year might not be appropriate to measure basis risk; *second*, participation rates in a region (e.g., county) are maintained at a relatively constant level, so yearly variation maybe not enough. A summary of the main variables is provided in Table 2.1.

Measuring Participation Rates

Denote n_{rltc} as the total insured acres for contract $r \in R = $ Yi $\cup$ Re where Yi and Re represent sets including yield and revenue contracts respectively, and crop $l \in L = \{\text{Corn}, \text{Soy}\}$ in year t in county c. Table 2.2 reports the product categories for yield and revenue contracts. For each insurance contract, e.g., revenue contract, we can construct the yearly coverage level as $\phi_{\text{Re}}^{ltc} = \sum_{Re} \phi_r^{ltc} \times w_r^{ltc}$ where ϕ_r^{ltc} is coverage level with $r \in$ Re and $w_r^{ltc} = n_{rltc}/\sum_{\text{Re}} n_{rltc}$ representing weights. In this study, all intensive margin participation rates are acreage-weighted coverage levels. Let $T = \{2009, \dots, 2020\}$, the intensive margin participation rate for revenue contracts will be $\phi_{\text{Re}}^{lc} = \sum_T \sum_{\text{Re}} \phi_r^{ltc} \times (n_{rltc}/n_{lc})$ where $n_{lc} = \sum_T n_{ltc}$. It is noteworthy that both ϕ_{Re}^{lc} and ϕ_{Yi}^{lc} represent the spatial-only variations. All insurance data are from USDA Risk

Management Agency (RMA) Summary of Business (SOB). [11]

Table 2.2 Information for Crop Insurance Plan.

Plan Code	Abbreviation	Name	Contract
Period: 2008-2010			
12	GRP	Group Risk Plan	Yield
25	RA	Revenue Assurance	Revenue
42	IP	Income Protection	Revenue
44	CRC	Crop Revenue Coverage	Revenue
45	IIP	Indexed Income Protection	Revenue
73	GRIP	Group Risk Income Protection	Revenue
90	APH	Actual Production History	Yield
Period: 2011-2020			
01	YP	Yield Protection	Yield
02	RP	Revenue Protection	Revenue
03	RPHPE	Revenue Protection with Harvest Price Exclusion	Revenue
04	GRP	Group Risk Plan	Yield
	AYP	Area Yield Protection	Yield
05	ARP	Area Revenue Protection	Revenue
06	ARPHP	Area Revenue Protection with Harvest Price Exclusion	Revenue

Note: see Du et al. (2017) for period 2008-2010; see USDA RMA (2022) for period 2022-2020.

Measuring acreage response is challenging because there is no direct measure of a county's eligible acres (Goodwin et al. 2004). In our case, SOB cannot be merged with any other data directly to match between insured acres and total eligible (insurable) acres. Therefore, we employ data from USDA National Agricultural Statistics Service (NASS) [12] and Farm Service Agency (FSA) [13] for more information, such as planted, prevented, or failed acres. We may ask why both NASS and FSA are needed in this study. *First*, the FSA dataset only includes farm operators who participate in the government programs, but the NASS survey tries to cover each

---

[11] Data can be downloaded from https://www.rma.usda.gov/SummaryOfBusiness.

[12] Data can be downloaded from https://quickstats.nass.usda.gov.

[13] Data can be downloaded from https://www.fsa.usda.gov/news-room/efoia/electronic-reading-room/frequently-requested-information/crop-acreage-data/index.

farm; as such, NASS is more representative than FSA. *Second*, the FSA dataset has prevented and failed acres; however, NASS does not provide such information.

Denote a^{N1}, a^{N2}, a^{N3} as planted, harvested, and silage acres from USDA/NASS; a^{F1}, a^{F2}, a^{F3} as FSA planted, prevented, and failed acres. Extensive margin participation rates can be constructed as:

(2.22a) $e^1 = \dfrac{n}{\max(a^{N1}, a^{F1}) + a^{F2} + a^{F3}};$

(2.22b) $e^2 = \dfrac{n}{\max(a^{N1}, a^{F1} + a^{F3}) + a^{F2}};$

(2.22c) $e^3 = \dfrac{n}{\max\left(\underbrace{a^{N1} - a^{N3}}_{\text{remove corn silage}}, a^{F1} + a^{F3}\right) + a^{F2}}.$

All elements in (2.22a)-(2.22c) have subscript $rltc$ to signify contract, crop, year and county information. It is noteworthy that NASS planted acres are not necessarily close to FSA planted acres although both of them are labeled as "planted". In some counties, NASS harvested acres are larger than FSA planted acres. Therefore, we construct e^1 and e^2 for both corn and soybeans. Another concern is that NASS records contain acres for corn silage.[14] We construct e^3 for corn grain by removing corn silage from NASS planted acres. Since acres for corn silage are non-negative, i.e., $a^{N3} \geq 0$, then e^3 will exceed e^2.

As in Figure 2.2, many measures are more significant than the theoretical maximum value (i.e., 1.0). It is noteworthy that states outside the main production area have more outliers (extreme values), e.g., North Dakota and Nebraska. When moving from west to east, extensive margin has a decreasing trend, but on the contrary, intensive margin seems to have an increasing trend (see Figures 2.3 and 2.4).

---

[14] In general, silage is usually made from maize, sorghum or other cereals. As one of them, corn silage is a high-quality, popular forage crop for dairy farms and beef cattle farms.
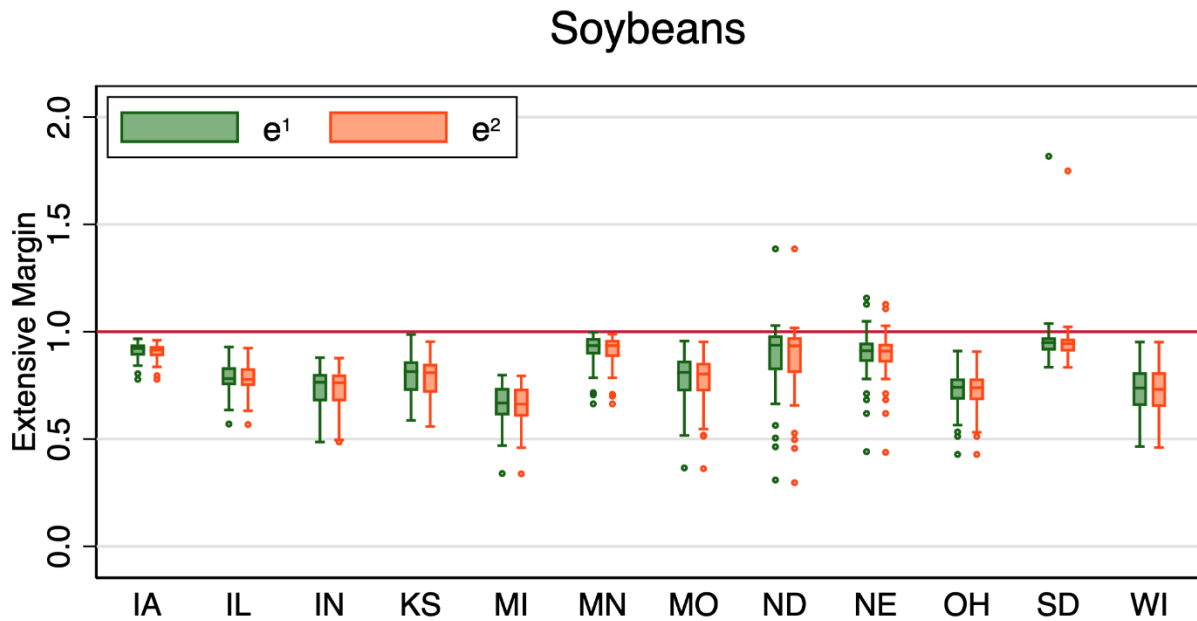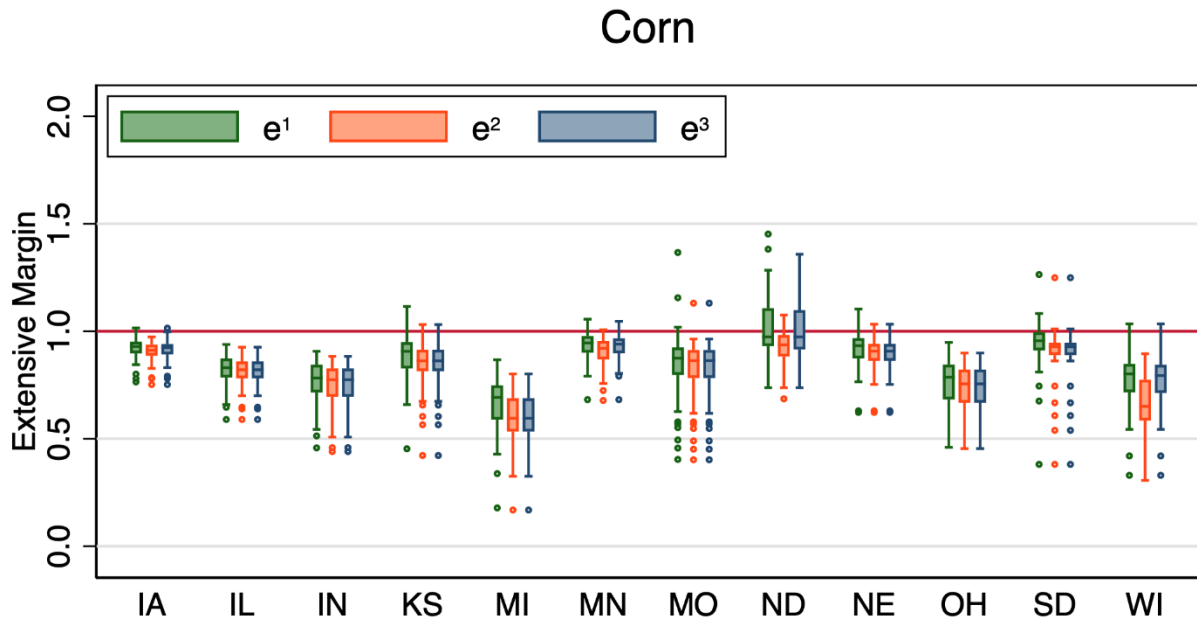
Figure 2.2 Extensive Margin Participation Rates over 2009-2020.
Notes: On each box, the central mark is median. Denote $x_{[25]}$ and $x_{[75]}$ as the 25th and 75th percentiles, then upper and lower adjacent lines represent $x_{[75]} + 1.5 \times IQR$ and $x_{[25]} + 1.5 \times IQR$, respectively, where $IQR = x_{[75]} - x_{[25]}$. Dots represent other observations outside the range above.

Corn

Soybeans

Figure 2.3 Extensive Margin Participation Rates from West to East.
Note: Each point represents one county. For convenience, we just choose e^1.

71

Figure 2.4 Intensive Margin Participation Rates from West to East.
Note: Each point represents one county. Only revenue contracts are included here. Yield contracts will be adopted in the empirical analysis.

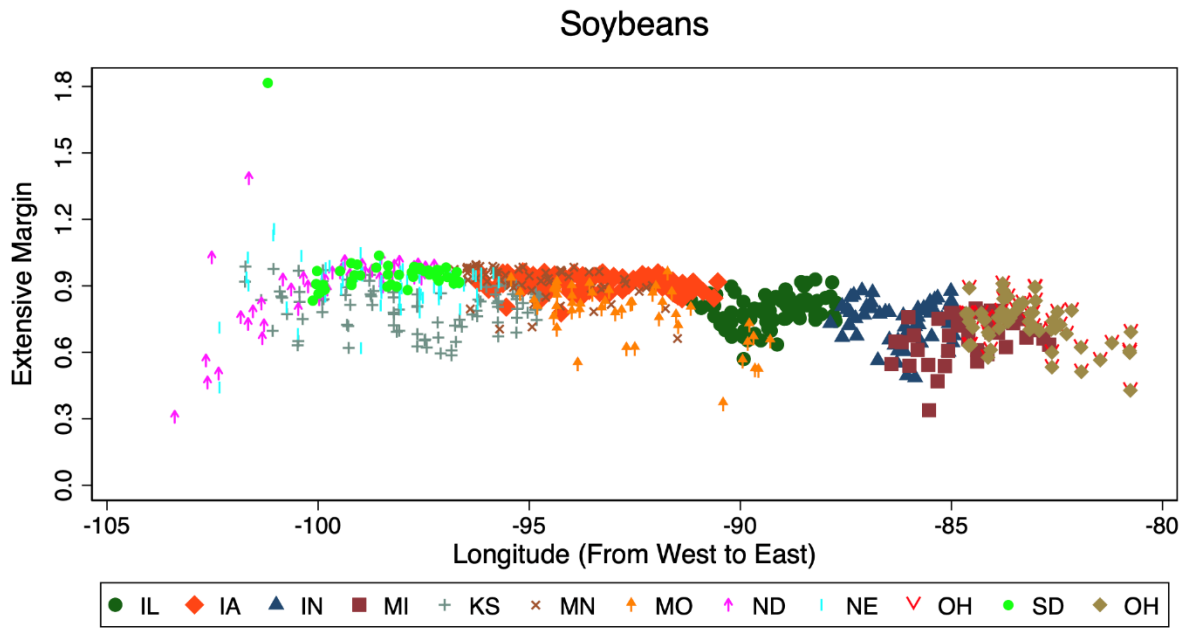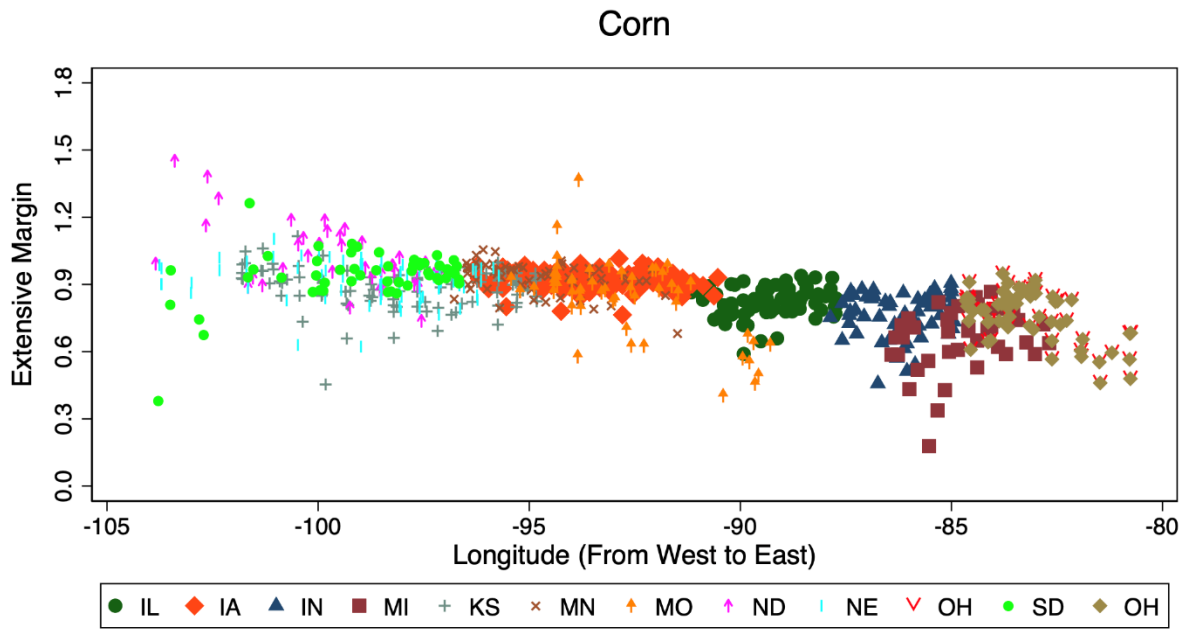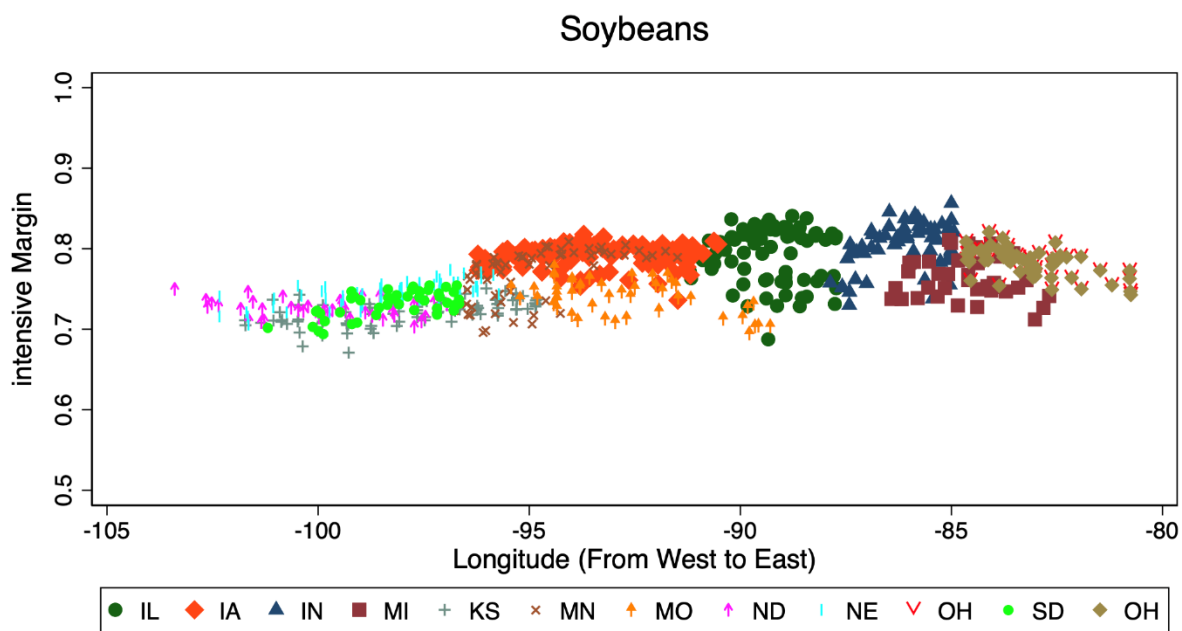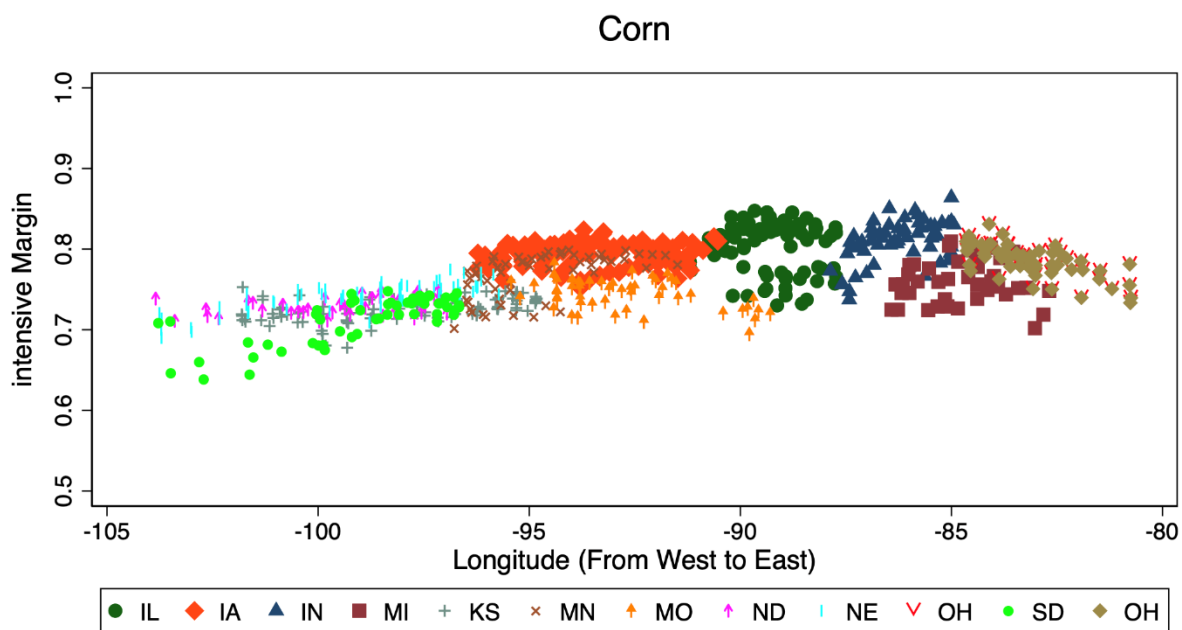*Measuring Basis Risk*

Recall that in the conceptual model, we define basis as $B_{0c}^c(\sigma) \equiv \bar{B} + \varepsilon\sigma$ where 0c indicates the harvesting time (October); $\bar{B}$ indicates the geographical heterogeneity which are county-level constants due to locations; ε is a random variable with $E[\varepsilon] = 0$; σ measures basis risk. Therefore, a simple transformation shows that $\sigma_B \equiv \sigma \times \sigma_\varepsilon$ here $\sigma_B = \sqrt{\text{var}\left(B_{0c}^c(\sigma)\right)}$, and $\sigma_\varepsilon = \sqrt{\text{var}(\varepsilon)}$. Without loss of generality, we normalize $\sigma_\varepsilon = 1$ so that $\sigma_B = \sigma$. Therefore, basis risk can be measured based on the observed cash and futures prices.

Corn December and Soybeans November futures contracts are employed for our analysis. Denote $P_{0c}^{lc,ntd}$ as cash price at the harvesting time (October) for elevator n on trading day d for crop l in county c and year t; $F^{l,td}$ as futures price. Then the expression of normalized basis risk is $B_{0c}^{lc,ntd} = \left(P_{0c}^{lc,ntd} - F^{l,td}\right)/F^{l,td}$. The motivation for normalization is that we are inclined to have scale-free numbers as best possible because participation rates are measured on the interval $[0,1]$. Other normalization techniques can be found in the literature (see more in Nayak et al. 2014); however, these techniques may not be appropriate because they only focus on one input variable.

With normalized basis $B_{0c}^{lc,ntd}$, we turn to constructing the county-level basis risk. Denote D_t as a set including total trading days in year t; N_c as a set including total elevators and T_c as a set including total years in county c. Then the basis risk can be constructed as:

(2.23a) elevator-level yearly basis $\qquad \bar{B}_{lc,nt} = (1/D_t)\sum_{d=1}^{D_t} B_{0c}^{lc,ntd}$;

(2.23b) elevator-level yearly variation $\qquad \sigma_{lc,nt} = \left[(D_t - 1)^{-1}\sum_{d=1}^{D_t}\left(B_{lc,ntd} - \bar{B}_{lc,nt}\right)^2\right]^{1/2}$;

(2.23c) county-level basis risk $\qquad \sigma_{lc} = \frac{1}{N_c T_c}\sum_{t=1}^{T_c}\sum_{n=1}^{N_c} \sigma_{lc,nt}$.

We may ask whether there are other alternatives to measure basis risk because elevator-level basis can be readily aggregated to county-level basis. we first define three bases as:

(2.24a) county-level daily basis $\qquad B_{lc,td} = (N_c)^{-1} \sum_{n=1}^{N_c} B_{lc,ntd};$

(2.24b) county-level yearly basis $\qquad B_{lc,t} = (D_t)^{-1} \sum_{d=1}^{D_t} B_{lc,td};$

(2.24c) county-level long-term average $\qquad B_{lc} = (T_c)^{-1} \sum_{n=1}^{T_c} B_{lc,t}.$

Then three alternatives for basis risk can be constructed as:

(2.25a) $\qquad \sigma_{lc}^{A1} = (1/T_c) \sum_{t=1}^{T_c} \sigma_{lc,t}^{\oplus}$ with $\sigma_{lc,t}^{\oplus} = \left[\frac{1}{D_t-1} \sum_{d=1}^{D_t} (B_{lc,td} - B_{lc,t})^2 \right]^{1/2};$

(2.25b) $\qquad \sigma_{lc}^{A2} = \left[\frac{1}{T_c D_t - 1} \sum_{t=1}^{T_c} \sum_{d=1}^{D_t} (B_{lc,td} - B_{lc})^2 \right]^{1/2};$

(2.25c) $\qquad \sigma_{lc}^{A3} = \left[\frac{1}{T_c-1} \sum_{t=1}^{T_c} (B_{lc,t} - B_{lc})^2 \right]^{1/2}.$

Figure 2.5 shows the kernel density estimations (KDE) for all four basis risks. We find that the first basis risk σ has the minimum means and standard deviations among all measurements. In light of large heterogeneities of scale, it is natural to ask which basis risk is the most appropriate one for our analysis. We assume that elevators in a county are homogenous, and farmers are sensitive to the cash prices of elevators nearby. If this is the case, the variable averaged out from elevator-level yearly variation might capture the true basis risk. As such, county-level basis risk σ in Eqn.(2.23c) will be used (red curve in Figure 2.5). The first three basis risks are highly correlated because $\text{corr}(\sigma, \sigma^{A1}) = 0.8$, $\text{corr}(\sigma, \sigma^{A2}) = 0.7$ and $\text{corr}(\sigma^{A1}, \sigma^{A2}) = 0.7$. However, the last one σ^{A3} is nearly uncorrelated with the others.

Figure 2.6 shows spatial disparities of basis risk. We find that the basis risk for corn is more dispersed than soybeans, which is reasonable because the interstate basis for corn has a more extensive range than soybeans (see Figure 2.7). The scale of basis (i.e., the absolute value of basis) increases in general as one moves away from the main production areas (e.g., IA, IL,

IN), which can be explained due to various costs (e.g., storage or transportation). However, we cannot conclude the basis risk pattern because interstate and intrastate basis risks are dispersed. As in Figure 2.8, a general pattern is that participation rates seem high in areas where basis risk is low, although some states in the western Great Plains (e.g., North Dakota and South Dakota) need to be discussed further.

Table 2.3 provides more information on the elevator-level basis data, including number of counties, operating years and elevator count. We find that: (i) the percentage of counties with at least one elevator to entire counties in a state is highly correlated with elevator county. The correlation coefficient between "Percent" column of Panel A and "Mean" column of Panel C is 0.72; and (ii) the percentage ("Percent" column of Panel A) seems not correlated with the operating years ("Mean" column of Panel B) because their correlation coefficient is just 0.11. Spatial disparities in elevator count and operating years are shown in Figure 2A.1.

Land Capability

Land capability data are from National Resource Inventory (NRI). As in USDA (2015), there are eight Land Capability Classes (LCC). Specifically, Class I soils have no limitations restricting their use. However, Class VIII soils and miscellaneous areas have the limitation that precludes their use for commercial plant production and limits their use for recreation, wildlife, water supply, or esthetic purposes. As in the literature, we measure the soil quality by calculating the percentage of good land in total land. We employ Class I-II as the good land as in Goodwin et al. (2004), although Du *et al.* (2017) choose Class I-IV.

Corn
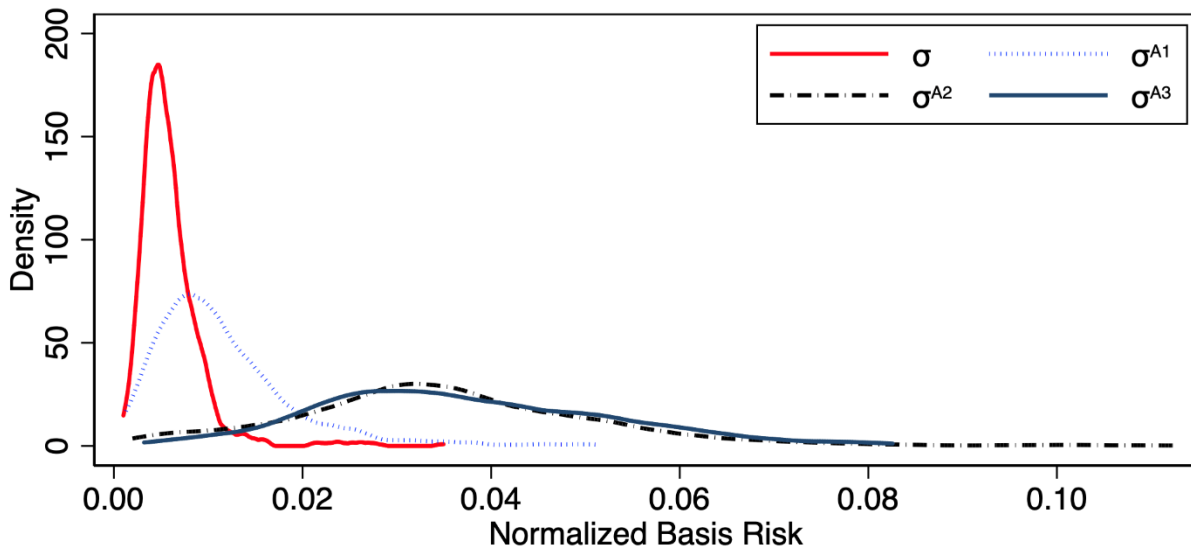
Soybeans
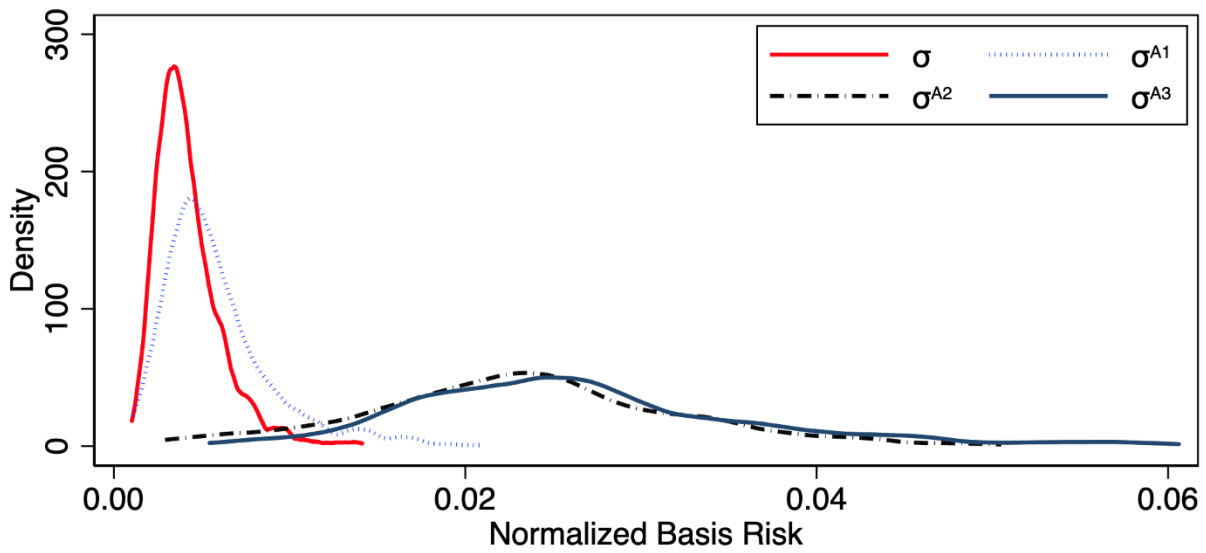
Figure 2.5 Kernel Density Estimation (KDE) for Four Basis Risks.
Notes: The Gaussian kernel is employed. Bandwidth is the optimal one that is selected to minimize the mean integrated square error. Distribution stands for spatial heterogeneity across 12 states.
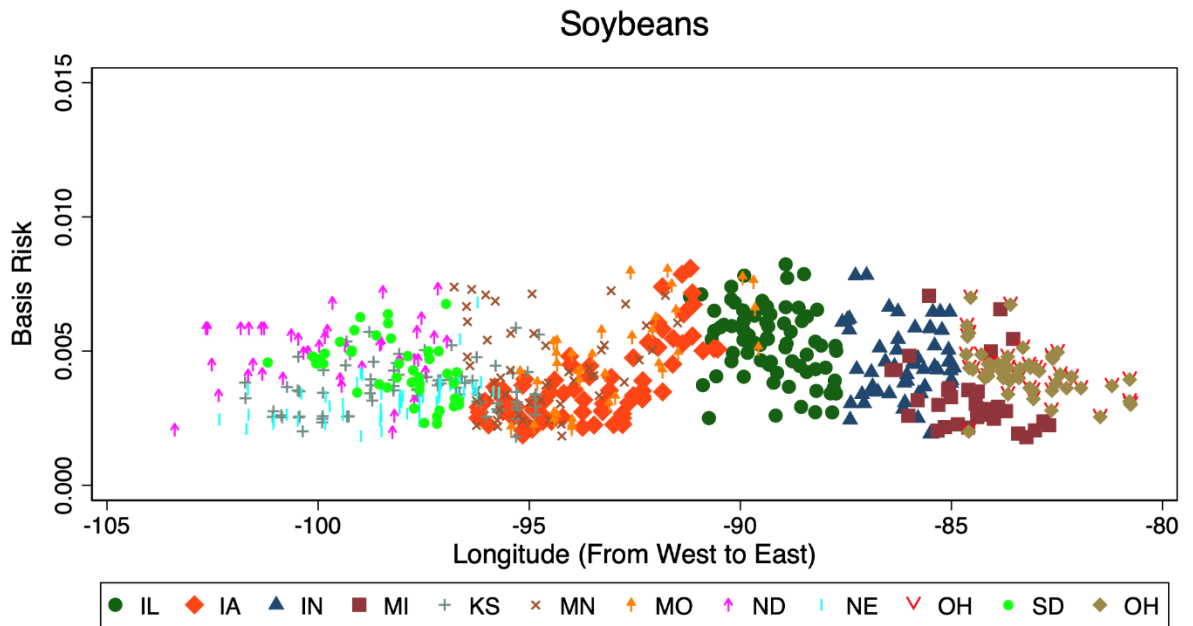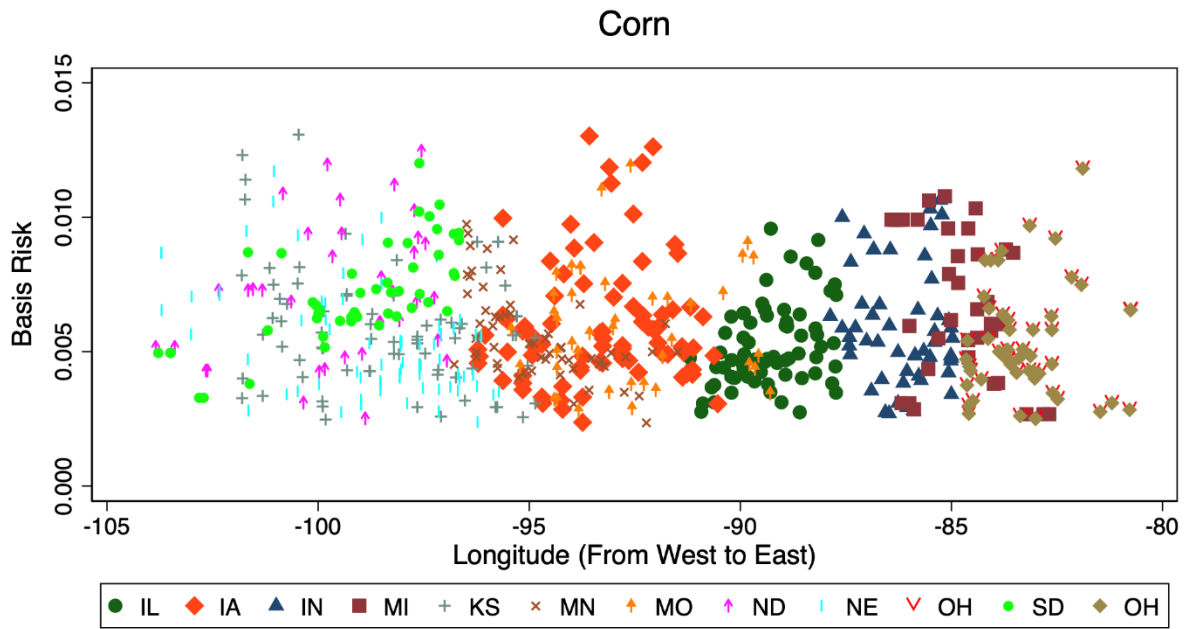
Figure 2.6 County-Level Basis Risks from West to East.
Note: Only points between 5th and 95th quantiles are included in this figure.
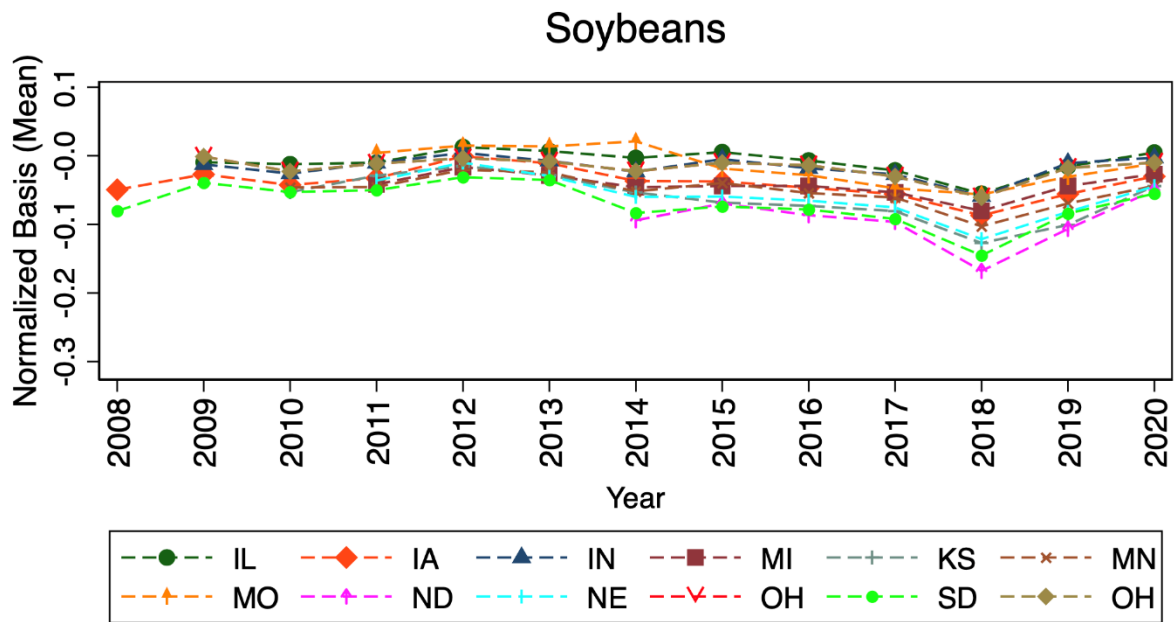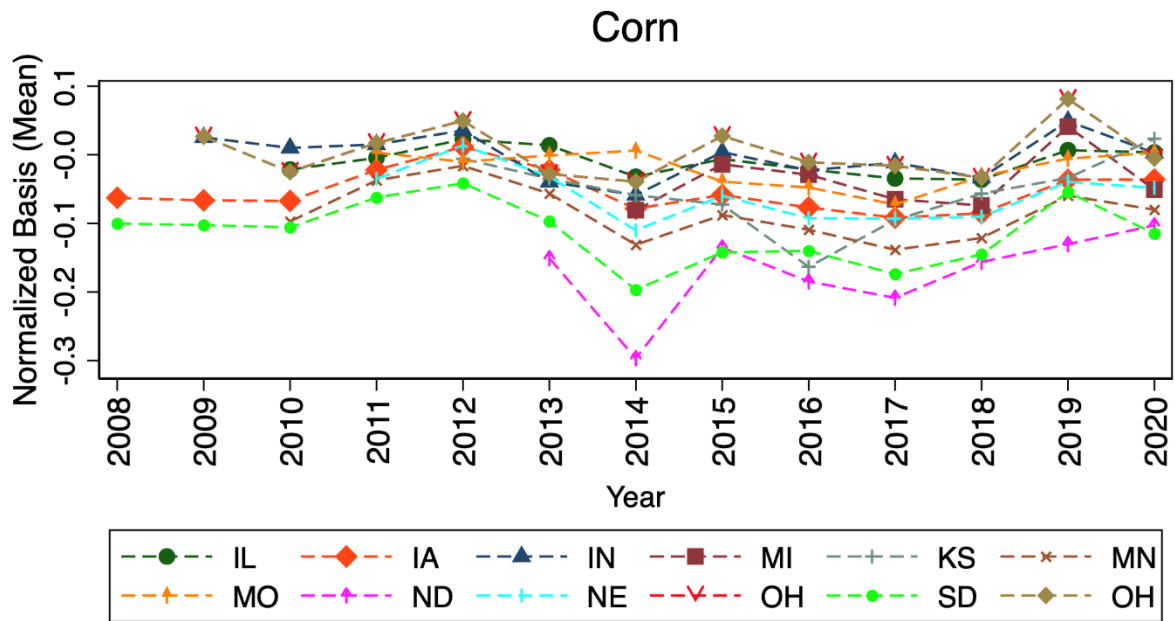
Figure 2.7 Patterns of Mean Normalized Basis for 12 States.
Note: Mean normalized basis is the mean value of all counties within a state per year. The goal here is to exhibit the heterogeneity in different states.

Corn Basis Risk

- ■ (0.008,0.127] □ (0.006,0.008] □ (0.005,0.006]
- □ (0.004,0.005] ■ [0.001,0.004] □ No data

Soybeans Basis Risk

- ■ (0.006,0.053] □ (0.004,0.006] □ (0.004,0.004]
- □ (0.003,0.004] ■ [0.001,0.003] □ No data

Corn Extensive

- ■ (0.93,1.25] □ (0.88,0.93] □ (0.83,0.88]
- □ (0.74,0.83] ■ [0.17,0.74] □ No data

Soybeans Extensive

- ■ (0.93,1.75] □ (0.87,0.93] □ (0.80,0.87]
- □ (0.72,0.80] ■ [0.30,0.72] □ No data

Corn Intensive

- ■ (0.80,0.86] □ (0.77,0.80] □ (0.74,0.77]
- □ (0.72,0.74] ■ [0.61,0.72] □ No data

Soybeans Intensive

- ■ (0.79,0.86] □ (0.77,0.79] □ (0.74,0.77]
- □ (0.72,0.74] ■ [0.64,0.72] □ No data

Figure 2.8 Basis Risk and Participation Rates.

Table 2.3 Elevator Information for 12 States.

State	Panel A: Number of Counties			Panel B: Operating Years Per County				Panel C: Elevator Count Per County			
	Total	Elevator≥1	Percent	Mean	Std. Dev	Min	Max	Mean	Std. Dev	Min	Max
Illinois	102	90	88.2%	3.71	1.07	1.5	6	6.22	5.22	1	24
Indiana	92	70	76.1%	4.84	1.64	1	8.5	2.7	1.74	1	8
Iowa	99	96	97.0%	4.43	1.43	1.5	8.64	9.52	5.24	1	23
Kansas	105	100	95.2%	3.46	1.63	1	8.33	5.09	4.09	1	23
Michigan	83	38	45.8%	3.98	1.43	1.5	8	2.45	1.27	1	6
Minnesota	87	64	73.6%	3.73	1.1	1.4	6.29	6.16	3.45	1	15
Missouri	115	58	50.4%	3.58	1.54	1	10	2.45	1.58	1	8
North Dakota	53	40	75.5%	3.3	1.18	1	5	3.53	2.63	1	13
Nebraska	93	81	87.1%	4.05	1.36	1	8.2	6.2	4.2	1	20
Ohio	88	64	72.7%	4	1.23	2	8.5	4.01	3.38	1	13
South Dakota	66	55	83.3%	4.06	2.1	1	10	3.75	2.16	1	8
Wisconsin	72	56	77.8%	3.34	1.24	1	6	3.34	2.27	1	10

Note: County-level mean operating years and total elevators are summarized at first. The intra-state distribution shows heterogeneities for those counties that have at least one elevator.

Weather Determinants

As in the literature (e.g., Schlenker and Roberts 2009; Xu et al. 2013), we construct two variables-Growing Degree Days (GDD) and Stress Degree Days (SDD)-to represent beneficial heat and heat stress during the growing season (April to September). The formulas for variable GDD and SDD in the year t are:

$$(2.26a)\ GDD_{c,t} = \sum_{d \in \Omega_t} \left[0.5 \left(\min(\max(T_{c,d,t}^{\max}, T^l), T^h) + \min(\max(T_{c,d,t}^{\min}, T^l), T^h) \right) - T^l \right]$$

$$(2.26b)\ SDD_{c,t} = \sum_{d \in \Omega_t} \left[0.5 \left(\max(T_{c,d,t}^{\max}, T^k) + \max(T_{c,d,t}^{\min}, T^k) \right) - T^k \right]$$

where c is county; d is day; Ω_t is the set of growing season days in year t. The thresholds are $T^l = 10°C$, $T^h = 30°C$ and $T^k = 32.2°C$. In this study, we use climatological normal around 30 years to control for weather effect on coverage level choice. The time length we selected is $\{1990, \dots, 2019\}$ and then $\bar{G}_c = (1/30) \sum_{j=1990}^{2019} GDD_{c,j}$, $\bar{S}_c = (1/19) \sum_{j=1990}^{2019} SDD_{c,j}$.

Yield Risk

We use coefficient of variance (CoV) to control for yield risk as in Goodwin *et al.* (2004). Denote y_{ct} be the actual county-level yield recorded by USDA NASS for county c and year t. As in Deng et al. (2007), we employ the trend equation $\log(y_{ct}) = \lambda_0 + \lambda_1(2020 - t) + u$, where $t \in T^{\det} = \{1971, \dots, 2020\}$ and u is the error term. We predict $\hat{y}_{ct}$ and $\hat{y}_{c,2020}$, and then construct detrended county-level yield as $y_{ct}^{\det} = y_{ct} \times \hat{y}_{c,2020}/\hat{y}_{ct}$. In final, yield risk can be measured as $YR_c = \sigma_{y\det}^c / \mu_{y\det}^c$, where $\sigma_{y\det}^c = \sqrt{\frac{1}{T^{\det}-1} \sum_{t \in T^{\det}} (y_{ct}^{\det} - \bar{y}_c^{\det})^2}$ and $\mu_{y\det}^c = (1/T^{\det}) \sum_{t \in T^{\det}} y_{ct}^{\det}$.

We may ask why the time length for detrending is 50-year ($T^{\det} \in \{1971, 2020\}$), but that for the empirical analysis is only 12 years ($T \in \{2009, \dots, 2020\}$). First, different time length does not affect our cross-sectional analysis. Recall that all variables in this study only show

spatial variations. Second, longer yield records can better capture the underlying yield risk because a more extended time length contains more information such as technological development, disaster, and policy adjustments.

Ratio of Ending Stock to Production

We construct the ratio of ending stock to production as an instrumental variable. The literature concludes that: (i) the relationship between supply and demand affects the basis level in each market (e.g., Jiang 1997); and (ii) there is an inverse relationship between basis and the stocks to storage capacity ratio (Karlson et al. 1993). For instance, when a massive volume of production puts pressure on storage in a market, the basis will weaken, and the farmer's revenue expectation will decrease.

To our best knowledge, county-level supply and demand data are hard to obtain. However, we are motivated to employ state-level data because USDA NASS provides state-level quarterly stock data (i.e., Mar, Jun, Sep, and Dec). Denote ES_{Ma}^{lst} as the ending stock of crop l in state s in March of year t and PD^{lst} as actual production. we construct $RaES_{ls} = \sum_{t \in T} ES_{Ma}^{lst} / \sum_{t \in T} PD^{lst}$ representing a ratio of ending stock to production. We assume that the effect of this ratio on the basis is negative, which indicates that the more the storage, the weak the basis (see Figure A2 for state-by-year variations over 2009-2020). We find: (i) the ratios for corn are higher than those for soybeans in general; and (ii) ratios in the main production area seem higher than in other regions.

Unit Choice Data for Simulation

We employ USDA RMA's 2009 farm-level contract choice data for our simulation. This dataset contains farm-level actual yield, approved historical yield, coverage level choice, premium payment, and location information (state and county). We may ask why such dated data

are still adopted in this study. The reasons are: (i) there are not many changes in participation rates after 2009, so we assume farmers' behavior does not have distinct changes as well; and (ii) observations are enough for simulation (corn: 687,274; soybeans 567,711). Therefore, the farmer's WTP will be implicitly estimated as in Eqn. (2.5).

Empirical Results

Estimation Method

We will employ Fractional Probit with Control Function (FPCF) for the empirical analysis. *First of all*, Fractional Probit is more appropriate than linear model because the participation rate is measured as a fraction. *Second*, potential endogeneity issues induced by omitted variables may exist. Some unobserved variables, e.g., farmer's subjective probability for future's loss, affect supply and demand in the local market, which then impact participation rate and basis risk.

Denote e_{lc} and ϕ_{lc} as extensive (the share of insured acres in total insurable acres) and intensive margin (acreage-weighted coverage choices) participation rates for crop l in county c; σ_{lc} as basis risk; $\mathbf{X_{lc}}$ as a vector of explanatory variables with $\mathbf{X_{lc}} = \{SQ, \bar{G}, \bar{S}, YR\}$; $\mathbf{Z_{lc}}$ as a vector of instrumental variables with $\mathbf{Z_{lc}} = \{\text{EleAmt}, \text{AveYear}, \text{RaES}\}$. Then the model specification is:

(2.27a) Control Function $\quad \sigma_{lc} = \mathbf{Z_{lc}}\Pi + \mathbf{X_{lc}}\beta^{(1)} + v_{lc};$

(2.27b) Fractional Probit $\quad E[\text{rate}_{lc}|\sigma_{lc}, X_{lc}, \eta_{lc}] = \Phi(\gamma\sigma_{lc} + \mathbf{X_{lc}}\beta^{(2)} + \eta_{lc}),$

Where $\text{rate}_{lc} \in \text{Pa} = \{e_{lc}, \phi_{lc}\}$, $E[\cdot]$ is the expectation operator, η_{lc} is an omitted variable with $\text{cov}(\sigma_{lc}, \eta_{lc}) \neq 0$. Explanatory variables are chosen as in the literature (e.g., LaFrance et al. 2002; Goodwin et al. 2004; Du et al. 2014; Yu and Sumner 2018; Che et al. 2020). As to the instrumental variables, we assume elevator operation and aggregate ending

stock are not correlated with the omitted variables such as farmer's subjective belief.

The estimation procedure for Fractional Probit with an Endogenous Explanatory Variable (EEV) has been discussed extensively in the literature (see Papke and Wooldridge 1996, 2008; Wooldridge 2015). If (v_{lc}, η_{lc}) is jointly normal then a two-step control function method can be processed as: (i) regress σ_{lc} on $\mathbf{Z_{lc}}, \mathbf{X_{lc}}$ and obtain the residuals $\hat{v}_{lc}$; (ii) use "Probit" of rate_{lc} on $\sigma_{lc}, \mathbf{X_{lc}}, \hat{v}_{lc}$ to estimate scaled coefficients, say $\hat{\gamma}_{scale}, \widehat{\boldsymbol{\beta}}_{\mathbf{scale}}^{(2)}, \hat{\delta}_{scale}$. Denote $\hat{\sigma}_{\eta_{lc}}^2$ as the variance of the omitted variable, then $\hat{\gamma}_{scale} = \hat{\gamma}/\left(1 + \hat{\sigma}_{\eta_{lc}}^2\right)^{1/2}$, which is also applied to $\widehat{\boldsymbol{\beta}}_{\mathbf{scale}}^{(2)}$ and $\hat{\delta}_{scale}$. The coefficient on $\hat{v}_{lc}$, labeled as $\hat{\gamma}_{scale}$, can provide a simple t test for the fixed effect with null hypothesis $H_0: \gamma = 0$ (exogeneous test).

Empirical Results

Table 2.4 reports semi-elasticities of insurance demand, which explains the change in the participation rate for a proportional change in basis risk. For extensive margin (e^1, e^2, e^3) participation rate, Fractional Probit with Control Function is an appropriate approach because all residuals are significant at a 1% significance level (p-value of the residual is less than 0.01). The semi-elasticities of insurance demand for corn are lower than -0.300, and for soybean are lower than -0.769, which implies that the share of insured acres in all insurable acres will increase by at least 3 and 7.7 percentage points for corn and soybean, respectively, when the basis risk decreases by 10%.

However, for the intensive margin participation rate, Fractional Probit with Control Function is just appropriate for the corn revenue contract, but Fractional Probit is appropriate for corn yield contract, soybean revenue, and yield contracts. The semi-elasticity of insurance demand for the corn revenue contract is -0.043, implying that the coverage level will increase by 0.4 percentage points if the basis risk decreases by 10%. The semi-elasticities for the soybean

Table 2.4 Semi-Elasticities for Corn and Soybeans.

Estimation Method	Fractional Probit					Fractional Probit with CF				
Dependent Variable	e^1	e^2	e^3	ϕ_{yi}	ϕ_{re}	e^1	e^2	e^3	ϕ_{yi}	ϕ_{re}
Corn										
Normalized Basis risk	0.003	0.003	0.003	-0.001	-0.001	-0.300+	-0.330+	-0.359+	-0.032	-0.043+
	(0.004)	(0.004)	(0.004)	(0.002)	(0.001)	(0.076)	(0.076)	(0.081)	(0.025)	(0.013)
Land Capability	0.061+	0.078+	0.067+	0.035+	0.023+	0.083+	0.102+	0.094+	0.037+	0.026+
	(0.009)	(0.009)	(0.009)	(0.005)	(0.002)	(0.011)	(0.011)	(0.012)	(0.005)	(0.002)
Growing Degree Days	-0.261+	-0.094*	-0.205+	0.083+	0.126+	-0.317+	-0.160+	-0.270+	0.078+	0.119+
	(0.050)	(0.050)	(0.049)	(0.023)	(0.009)	(0.050)	(0.052)	(0.049)	(0.022)	(0.009)
Stress Degree Days	0.032+	0.029+	0.026+	-0.016+	-0.020+	0.026+	0.021+	0.017+	-0.018+	-0.022+
	(0.005)	(0.006)	(0.006)	(0.002)	(0.001)	(0.006)	(0.006)	(0.006)	(0.003)	(0.001)
Yield Risk	0.132+	0.132+	0.129+	-0.040+	-0.027+	0.183+	0.189+	0.192+	-0.036+	-0.020+
	(0.014)	(0.017)	(0.016)	(0.005)	(0.003)	(0.021)	(0.023)	(0.023)	(0.007)	(0.004)
residual (p-value)						0.000	0.000	0.000	0.222	0.001
Obs.	747	747	747	747	747	747	747	747	747	747
Soybeans										
Normalized Basis risk	0.002	0.002		-0.004*	-0.002+	-0.769+	-0.807+		-0.042	-0.044
	(0.004)	(0.004)		(0.002)	(0.001)	(0.184)	(0.183)		(0.058)	(0.029)
Land Capability	0.067+	0.066+		0.050+	0.024+	-0.016	-0.020		0.046+	0.019+
	(0.010)	(0.010)		(0.006)	(0.002)	(0.022)	(0.022)		(0.009)	(0.004)
Growing Degree Days	-0.247+	-0.216+		0.032	0.058+	0.772+	0.854+		0.083	0.113+
	(0.050)	(0.050)		(0.030)	(0.011)	(0.251)	(0.250)		(0.081)	(0.040)
Stress Degree Days	0.019+	0.013**		-0.010+	-0.017+	-0.114+	-0.131+		-0.017	-0.024+
	(0.006)	(0.006)		(0.003)	(0.001)	(0.031)	(0.032)		(0.011)	(0.005)
Yield Risk	0.090+	0.087+		-0.056+	-0.021+	0.186+	0.188+		-0.051+	-0.016+
	(0.017)	(0.017)		(0.011)	(0.003)	(0.030)	(0.030)		(0.013)	(0.004)
residual (p-value)						0.000	0.000		0.501	0.143
Obs.	728	728		728	728	728	728		728	728

Note: (1) standard errors in parentheses; (2) * p<0.10, ** p<0.05, + p<0.01.

yield and revenue contracts are -0.004 and -0.002, respectively, implying that the coverage levels will increase by 0.04 and 0.02 percentage points if the basis risk decreases by 10%.

Conclusion

The basis is defined as the difference between spot and futures prices. Therefore, the basis risk measures the mismatch between cash and futures markets. Due to the crop insurance contract design, the basis and the basis risk exist, which might deter farmers' insurance uptake for both extensive and intensive margins. Understanding the effect of basis risk is crucial because crop revenue insurance, first introduced in the 1990s, has been the most popular product offered by Federal Crop Insurance Corporation.

In this study, we first develop a conceptual model to analyze farmers' acreage response to basis risk. Given a coverage level, a farmer's willingness to pay for insurance products increases when basis risk decreases, which implies that a farmer is inclined to pay a higher premium per acre or insure more acres when the basis risk is lower. Next, we employed Fractional Probit with Control Function for the empirical analysis. Regression results show that when the basis risk decreases by 10%: (i) the share of insured acres in all insurable acres will increase by at least 3 and 7.7 percentage points for corn and soybean respectively; (ii) coverage level choice will increase by 0.4 and 0.02 percentage points for corn revenue and soybean revenue contracts, respectively.

The effect of basis risk on participation rate might shed light on adjusting federal policies. If the revenue contract is revised to remove the basis, basis risk will decrease by 100%. Given the current subsidy structure, the shares of insured acres in total insurable acres for corn and soybean will be more significant than 1 (theoretically, the maximum is 1); coverage level choice for the corn revenue contract increases from 0.76 to 0.803; coverage level choice for the

soybean revenue contract rises from 0.76 to 0.762.

Besides removing the basis, policymakers might adjust the subsidy structure to be consistent with the underlying basis risk. The regression results show that the lower the basis risk, the higher the intensive margin participation rate. Therefore, the subsidy rate might be reduced in the areas with low basis risks but increased in the areas with high basis risks.

REFERENCES

Babcock, B.A., E.K. Choi and E. Feinerman. 1993. Risk and probability premiums for CARA utility functions. *Journal of Agricultural and Resource Economics:* 17-24.

Babcock, B.A. and D.A. Hennessy. 1996. Input demand under yield and revenue insurance. *Amer. J. of Agr. Econ*. 78(2): 416-427.

Calvin, L. 1992. Participation in the US federal crop insurance program (No. 1800). US Department of Agriculture, Economic Research Service.

Carter, M., G. Elabed, and E. Serfilippi. 2015. Behavioral economic insights on index insurance design. *Agricultural Finance Review*.

Che, Y., H. Feng and D.A. Hennessy. 2020. Recency effects and participation at the extensive and intensive margins in the US Federal Crop Insurance Program. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 45(1): 52-85.

Chite, R. 1988. Federal crop insurance: background and current issues. In CRS report for Congress (USA). Congressional Research Service.

Clarke, D.J. 2016. A theory of rational demand for index insurance. *American Economic Journal: Microeconomics*, 8(1): 283-306.

Coble, K.H., T.O. Knight, R.D. Pope and J.R. Williams. 1997. An expected-indemnity approach to the measurement of moral hazard in crop insurance. *Amer. J. of Agr. Econ*. 79(1): 216-226.

Coble, K.H., T.O. Knight, B.K. Goodwin, M.F. Miller, R.M. Rejesus and G. Duffield. 2010. A comprehensive review of the rma aph and combo rating methodology: Final report. prepared by Sumaria systems for the Risk Management agency.

Deng, X., B.J. Barnett and D.V. Vedenov. 2007. Is there a viable market for area-based crop insurance?. *Amer. J. of Agr. Econ*. 89(2): 508-519.

Deng, X., B.J. Barnett, D.V. Vedenov, and J.W. West. 2007. Hedging dairy production losses using weather-based index insurance. *Agricultural Economics,* 36(2): 271-280.

Doherty, N. A., and A. Richter. 2002. Moral hazard, basis risk, and gap insurance. *Journal of Risk and Insurance,* 69(1): 9-24.

Du, X. and D.A. Hennessy. 2012. The planting real option in cash rent valuation. *Applied Economics,* 44(6): 765-776.

Du, X., D.A. Hennessy and H. Feng. 2014. A natural resource theory of US crop insurance contract choice. *Amer. J. of Agr. Econ*. 96(1): 232-252.

Du, X., H. Feng and D.A. Hennessy. 2017. Rationality of choices in subsidized crop insurance

markets. *Amer. J. of Agr. Econ*. 99(3): 732-756.

Feng, H., X. Du and D.A. Hennessy. 2020. Depressed demand for crop insurance contracts, and a rationale based on third generation Prospect Theory. *Agricultural Economics*, 51(1): 59-73.

Figlewski, S. (1984). Hedging performance and basis risk in stock index futures. *The Journal of Finance,* 39(3): 657-669.

Firpo, S., A.F. M. Kobus Galvao, T. Parker and P. Rosa-Dias. 2020. Loss aversion and the welfare ranking of policy interventions. arXiv preprint arXiv:2004.08468.

Gallagher, J., 2014. Learning about an infrequent event: evidence from flood insurance take-up in the United States. *American Economic Journal: Applied Economics*: 206-233.

Gardner, B.L. and R.A. Kramer. 1986. Experience with crop insurance programs in the United States. Pub. for Internatl. Food Policy Research Inst. by Johns Hopkins Univ. Press.

Glauber, J.W., 2004. Crop insurance reconsidered. *Amer. J. of Agr. Econ*. 86(5): 1179-1195.

Gollier, C. and J.W. Pratt. 1996. Risk vulnerability and the tempering effect of background risk. *Econometrica: Journal of the Econometric Society*: 1109-1123.

Gollier, C. 2001. The economics of risk and time. MIT press.

Goodwin, B.K. 1994. Premium rate determination in the federal crop insurance program: what do averages have to say about risk?. *Journal of Agricultural and Resource Economics*: 382-395.

Goodwin, B.K. 2001. Problems with market insurance in agriculture. *Amer. J. of Agr. Econ*. 83(3): 643-649.

Goodwin, B.K. and A. Hungerford. 2015. Copula-Based Models of Systemic Risk in U.S. Agriculture: Implications for Crop Insurance and Reinsurance Contracts. *Amer. J. of Agr. Econ*. 97: 879-96.

Haushalter, G. D. 2000. Financing policy, basis risk, and corporate hedging: Evidence from oil and gas producers. *The Journal of Finance*, 55(1): 107-152.

Hojjati, B. and N.E. Bockstael. 1988. Modeling the demand for crop insurance. International Food Policy Research Institute.

Jiang, B. 1997. Corn and soybean basis behavior and forecasting: fundamental and alternative approaches. Iowa State University.

Joe, H. and J.J. Xu. 1996. The estimation method of inference functions for margins for multivariate models.

Karlson, N., B. Anderson, and R.P. Dahl. 1993. Cash-Futures Price Relationships: Guides to

Corn Marketing (No. 1701-2016-139145).

Kousky, C. 2017. Disasters as learning experiences or disasters as policy opportunities? Examining flood insurance purchases after hurricanes. *Risk analysis*, 37(3): 517-530.

Meyer, J. 1987. Two-moment decision models and expected utility maximization. *The American economic review*: 421-430.

Meyer, J., and R.H. Rasche. 1992. Sufficient conditions for expected utility to imply mean-standard deviation rankings: empirical evidence concerning the location and scale condition. *The Economic Journal*, 102(410) 91-106.

Naik, G. and R.M. Leuthold. 1991. A Note on the Factors Affecting Corn Basis Relationships. *Journal of Agricultural and Applied Economics*, 23(1): 147-153.

Nayak, S. C., B.B. Misra, and H.S. Behera. 2014. Impact of data normalization on stock index forecasting. *International Journal of Computer Information Systems and Industrial Management Applications*, 6(2014): 257-269.

O'Donoghue, E. 2014. The effects of premium subsidies on demand for crop insurance. USDA-ERS economic research report, (169).

Papke, L.E. and J.M. Wooldridge. 1996. Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. *Journal of applied econometrics*, 11(6): 619-632.

Papke, L.E. and J.M. Wooldridge. 2008. Panel data methods for fractional response variables with an application to test pass rates. *Journal of econometrics*, 145(1-2): 121-133.

Ramirez, O.A. and J.S. Shonkwiler. 2017. A probabilistic model of the crop insurance purchase decision. *Journal of Agricultural and Resource Economics*: 10-26.

Rosch, Stephanie. 2021. "Federal Crop Insurance: A Primer", Report No. R46686.

Rosa, I. 2018a. Federal crop insurance: Program overview for the 115th congress. Report R45193.

Rosa, I. 2018b. Farm bill primer: Federal crop insurance. Congressional Research Service.

Sandmo, A. 1971. On the theory of the competitive firm under price uncertainty. *The American Economic Review*, 61(1): 65-73.

Schmidt, U., C. Starmer and R. Sugden. 2008. Third-generation prospect theory. *Journal of Risk and Uncertainty*, 36(3): 203-223.

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

Sheather, S.J. and M.C. Jones. 1991. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(3): 683-690.

Skees, J.R., 1987. Future research needs on federal multiple peril crop insurance (No. 2096-2018-3251).

Smith, V.H. and B.K. Goodwin. 1996. Crop insurance, moral hazard, and agricultural chemical use. *Amer. J. of Agr. Econ*. 78(2): 428-438.

Stein, D. 2018. Dynamics of demand for rainfall index insurance: evidence from a commercial product in India. The World Bank Economic Review, 32(3): 692-708.

Turvey, C.G., A. Weersink, and S.H.C. Chiang .2006: Pricing Weather Insurance with a Random Strike Price: The Ontario Ice-Wine Harvest *Amer. J. of Agr. Econ*. *88(August): 696-709*

U.S. Department of Agriculture. 2015. Summary Report: 2012 National Resources Inventory, Natural Resources Conservation Service, Washington, DC, and Center for Survey Statistics and Methodology, Iowa State University, Ames, Iowa. http://www.nrcs.usda.gov/technical/nri/12summary

US Risk Management Agency. 2022. Crop Insurance Handbook. FCIC 18010-1.

US GAO.2014. Considerations in Reducing Federal Premium Subsidies. Report GAO-14-700. US Gov. Account. Off., Washington DC.

US GAO. 2015. In areas with higher crop production risks, costs are greater, and premiums may not cover expected losses. Report GAO-15-215. US Gov. Account. Off., Washington DC.

Wooldridge, J. M. 2015. Control function methods in applied econometrics. *Journal of Human Resources,* 50(2): 420-445.

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

Yan, J., 2007. Enjoy the joy of copulas: with a package copula. *Journal of Statistical Software,* 21(4): 1-21.

Yu, J. and D.A. Sumner. 2018. Effects of subsidized crop insurance on crop choices. *Agricultural Economics,* 49(4): 533-545.

Zulauf, C., J. Coppess, G. Schnitkey and N. Paulson. 2018. Premium Subsidy and Insured

US Acres: Differential Impact by Crop. *farmdoc daily*, 8.

APPENDIX A: Monte Carlo Simulation

1. Estimation for Farmer's Willingness to Pay

We employ Monte Carlo simulation associated with Gaussian Copula Method to measure farmers' WTP (\$/acre) at all coverage levels. Given a coverage level, a high WTP indicates that a farmer would like to pay a high premium per acre or insure more acres. The main hypothesis we will test is that farmers' WTP decreases when basis risk increases (H1).

First of all, we use the Copula method to draw a unit-level subsample. As in the literature (Yan 2007; Du and Hennessy 2012; Goodwin and Hungerford 2015), we assume that a dependence structure exists among yield, spot, and futures prices. Our motivation is to estimate crop-specific farmers' WTP without considering other effects such as crop rotation. Therefore, the three-dimensional Multivariate Gaussian Copula (MGC) will be used, which indicates that the dimension of variance-covariance matrix is 3×3 (see more details in Appendix B).

Denote y_{st}, P_{st} and F_{st} as yield, cash and futures prices with $s \in \{\text{IL}, \text{IN}, \dots, \text{WI}\}$ and $t \in T^{\text{det}} = \{1971, \dots, 2020\}$. Yield and cash price data are from USDA NASS; futures price data are from Chicago Board of Trade (CBOT). Three main steps are used in the simulation: (i) assume yield model specification $y_{st} = \beta_0 + \beta_1(2020 - t) + \varepsilon_{st}$ where ε_{st} is the random variation. After estimating $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\varepsilon}_{st}$ via OLS, we can normalize yield variation by $\hat{\varepsilon}_{st} = y_{st}/\hat{y}_{st}$ where $\hat{y}_{st} = y_{st} - \hat{\varepsilon}_{st}$. Denote by $\bar{\varepsilon}_s$ and σ_{ε}, respectively, the mean and standard deviation of $\{\varepsilon_{st}\}_{t \in T^{\text{det}}}$. Then the upper support of $\hat{\varepsilon}_{st}$ can be constructed as $\hat{\varepsilon}_{s,\max} = \bar{\varepsilon}_s + 3\sigma_{\varepsilon}$ and the lower support $\tilde{\varepsilon}_{s,\min}$ can be imposed as 0. As such, a standard beta random variable $\tilde{\xi}$ can be constructed as $\tilde{\xi}_{st} = (\varepsilon_{st} - \tilde{\varepsilon}_{s,\min})/(\tilde{\varepsilon}_{s,\max} - \tilde{\varepsilon}_{s,\min})$ with $\tilde{\xi}_s = 1$ whenever $\varepsilon_{st} > \tilde{\varepsilon}_{s,\max}$; (ii) assume difference of logarithm of prices in 2 consecutive years, $\tilde{\zeta}_{st} = log(P_{s,t}) - log(P_{s,t-1})$ and $\tilde{\eta}_{st} = log(F_{st}) - log(F_{s,t-1})$ to be normally distributed; and (iii) with $\tilde{\xi}_{st}$, $\tilde{\zeta}_{st}$ and $\tilde{\eta}_{st}$, we

draw 5,000 records from the 2008 unit-level NASS/RMA yield dataset. Five input variables

contain state-level spot price, futures price in October, futures price in February, actual yield and

approved historical yield (APH).

Second, we employ Constant Absolute Risk Aversion (CARA) utility function $u(w) = 1 - e^{-Aw}$ where w is farmer's income and A is a risk-averse coefficient. Farmer's income can be

set as $400/acre, which requires that price and yield for corn (soybeans) are $4/bushel

($10/bushel) and 100 bushel/acre (40 bushel/acre). Then we assume that a farmer encounters a

normal year in which the farmer's gamble size h is $200. As the Eqn.(4) in Babcock et al.

(1993), we obtain the risk premium θ which reflects how much money a farmer can give up for

the potential risk. Babcock and Hennessy (1996) use risk premiums of 20% and 40%. In our

case, we set the risk premium as 0.20 and then the farmer's WTP as $40/acre, which is consistent

with the mean value of premiums farmers paid for revenue contracts over 2009-2020. For

robustness checks, we also set other three risk premiums 0.10, 0.06 and 0.02, which correspond

to farmer's WTP $20, $12 and $4. As such, we will employ four risk-averse coefficients – 0.008,

0.004, 0.002 and 0.0008- to correspond the four risk premiums 0.20, 0.10, 0.06 and 0.02.

Figure 2A.3 shows farmer's WTP when the risk-averse coefficient A is set as 0.004. We

show the results with a format "rank of basis risk - state name – rank of extensive margin

participation rate" on the y-axis. The range of ranking is from "01" to "12" where "01" and "12"

represent the lowest and highest orders respectively. In our case, the ideal result is "01-state

name-12" ("12-state name-01") which means that a state with the lowest (highest) basis risk has

the highest (lowest) extensive margin participation rate. Furthermore, the x-axis reports farmers'

WTP measured as $ per acre. The ideal result is that given a coverage level, a farmer's WTP

decreases when the ranking of basis risk increases.

Results in Figure 2A.3 seem not to support our argument robustly. Figures 2A.4 and 2A.5 show farmers' WTP for all four risk-averse coefficients for all states. We can then find: (i) farmer's WTP increases as coverage level increases; (ii) given a coverage level, farmer's WTP at four risk coefficients may intersect; and (iii) given a coverage level and a risk-averse coefficient, differential of farmer's WTP in different states seem not distinct.

Reasons for failure in the simulation may contain: (i) risk-averse coefficients are heterogenous in 12 states; (ii) county-level simulation should be employed rather than state-level simulation; (iii) basis risk cannot be effectively controlled because there are other confounding factors such as yield risk or subsidy effect. Hence, a regression analysis is needed to further identify the effect of basis risk on participation rates.

2. Copula Method

Yield, cash price and future price are three continuous random variables in our simulation, which can be denoted as X_Y, X_C, X_F. Suppose the joint distribution function is $F(x_Y, x_C, x_F)$ and marginal distribution functions are F_Y, F_C, F_F, then a 3-copula $C(\cdot)$ satisfies:

(A2.1) $F(x_Y, x_C, x_F) = C\big(F_Y(x_Y), F_C(x_C), F_F(x_F)\big)$ for $(x_Y, x_C, x_F) \in (-\infty, +\infty)^3$

Where $F_i = u_i \in [0,1]$ with $i \in \{Y, C, F\}$. Let F_i^{-1} be the inverse distribution function with $F_i^{-1} = sup\{x_i | F_i(X_i) \le u_i\}$, the Multivariate Gaussian Copula (MGC) can be constructed as

(A2.2) $C^{MGC}(u_Y, u_c, u_F; R) = \Phi(F_Y^{-1}(u_Y), F_C^{-1}(u_C), F_F^{-1}(u_F); R)$,

Where R is a symmetric, positive definite matrix with

(A2.3) $R = \begin{bmatrix} 1 & \rho_{CY} & \rho_{FY} \\ \rho_{CY} & 1 & \rho_{FC} \\ \rho_{FY} & \rho_{FC} & 1 \end{bmatrix}$

Where ρ_m with $m \in \{CY, FY, FC\}$ is the dispersion parameter. Let β and ρ be the vectors

of marginal distribution and the copula's dispersion parameters, then estimated parameter vector

is $\theta = (\beta^T, \rho^T)^T$. Given N observations, the corresponding log-likelihood function can be

specified as (Yan, 2007):

(A2.4) $l(\theta) = \sum_{i=1}^{N} \log c\{F_Y(X_{Y,i}; \beta), F_C(X_{C,i}; \beta), F_F(X_{F,i}; \beta); \rho\}$

$$+ \sum_{i=1}^{N} \sum_{j \in \{Y,C,F\}} \log f_i(X_{j,i}; \beta)$$

where $c(u_Y, u_C, u_F) = f[F_Y^{-1}(u_Y), F_C^{-1}(u_C), F_F^{-1}(u_F)] / \prod_{j \in \{Y,C,F\}} f_j[F_j^{-1}(u_j)]$ is the

copula density function; while $f(\cdot)$ and $f_i(\cdot)$ represent the multivariate and the univariate

marginal density functions, respectively. The ML estimator of θ is $\hat{\theta}_{ML} = \arg\max_{\theta \in \Theta} l(\theta)$. We

apply the two-step estimation method called Inference Functions for Margins (IFM) as in Joe and

Xu (1996):

(A2.5a) $\qquad \hat{\beta}_{IFM} = \arg\max_{\beta} \sum_{i=1}^{N} \sum_{j \in \{Y,C,F\}} \log f_i(X_{j,i}; \beta)$

(A2.5b) $\qquad \hat{\rho}_{IFM} = \arg\max_{\rho} \sum_{i=1}^{N} \log c\{F_Y(X_{Y,i}; \hat{\beta}_{IFM}), F_C(X_{C,i}; \hat{\beta}_{IFM}), F_F(X_{F,i}; \hat{\beta}_{IFM}); \rho\}$

The first step (A.5.a) consists of an ML estimation for each marginal distribution $j \in$

$\{Y, C, F\}$:

(A2.5c) $\qquad\qquad \hat{\beta}_{j,IFM} = \arg\max_{\beta_j} \sum_{i=1}^{N} \log f_i(X_{j,i}; \beta_j)$

3. Monte Carlo Method

The procedure of simulation is employed as in Du and Hennessy (2012):

(a) we first simulate 3 independent random variates $x = (x_Y, x_C, x_F)^T$ from the standard

distribution $N(0,1)$.

(b) Generate $v = Ax$ where A is the Cholesky decomposition of the estimated MGC

dispersion matrix $\hat{R}$, then $v = (v_Y, v_C, v_F)^T$.

(c) Set $u_j = \Phi(v_j)$, $j \in \{Y, C, F\}$ where Φ denotes the univariate $N(0,1)$ distribution

function.

(d) Set $\tilde{x}_j = F_j^{-1}(u_j)$, $j \in \{Y, C, F\}$ where F_j^{-1} denotes the inverse of the marginal cumulative density function, then a realization is $\tilde{x} = (\tilde{x}_Y, \tilde{x}_C, \tilde{x}_F)^T$. We repeat this procedure to obtain 5,000 realizations for each state S, $S \in \{IL, IN, \dots, WI\}$.

Corn Elevator Amount

- ■ (8.0,24.0]
- ■ (5.0,8.0]
- □ (3.0,5.0]
- □ (2.0,3.0]
- ■ [1.0,2.0]
- □ No data

Corn Elevator Year

- ■ (5.0,10.0]
- ■ (4.0,5.0]
- □ (3.5,4.0]
- □ (2.7,3.5]
- ■ [1.0,2.7]
- □ No data

Soybeans Elevator Amount

- ■ (7.0,27.0]
- ■ (5.0,7.0]
- □ (3.0,5.0]
- □ (2.0,3.0]
- ■ [1.0,2.0]
- □ No data

Soybeans Elevator Year

- ■ (5.0,10.0]
- ■ (4.0,5.0]
- □ (3.2,4.0]
- □ (2.6,3.2]
- ■ [1.0,2.6]
- □ No data
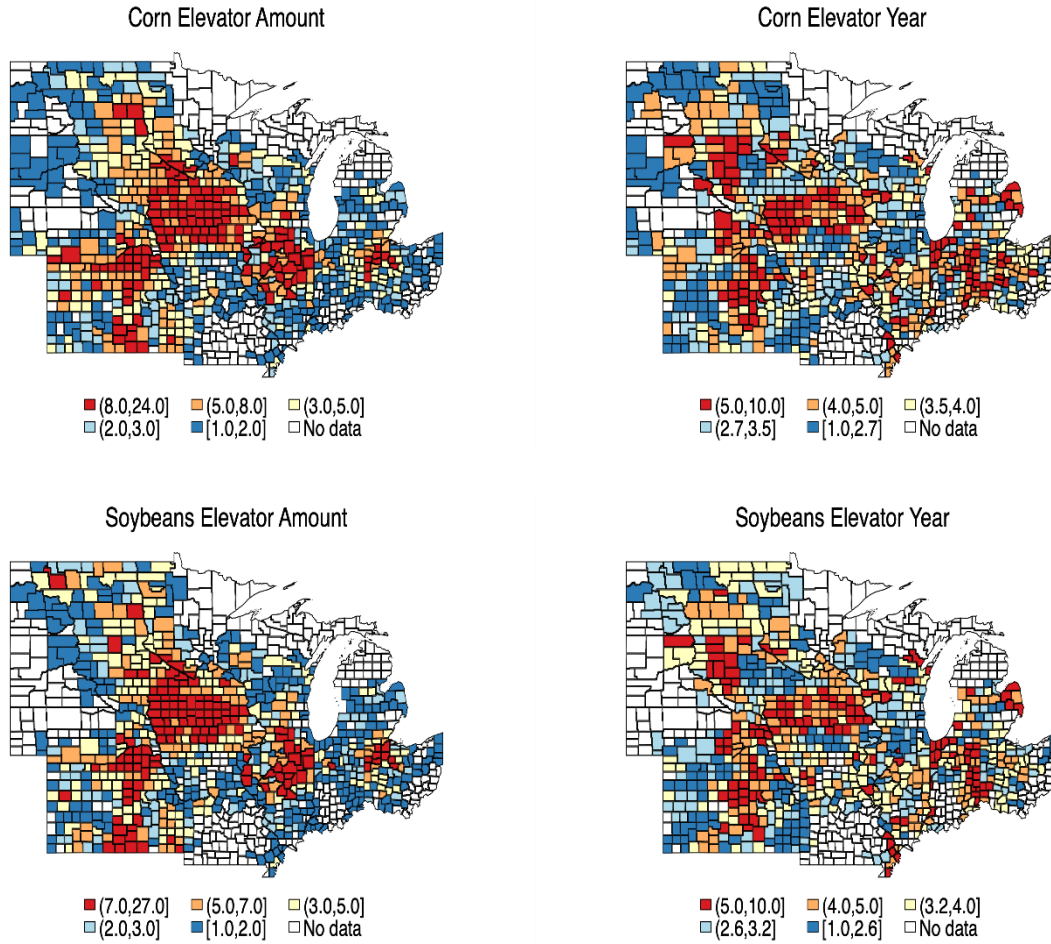
Figure 2A.1 Elevator Count and Operating Years.

Note: Elevator Year = $(1/N_c) \sum_{i=1}^{N_c} year_{i,c}$ where $year_{i,c}$ is the number of years for elevator i in county c; N_c is the elevator count. Both variables represent the elevator history in each county.
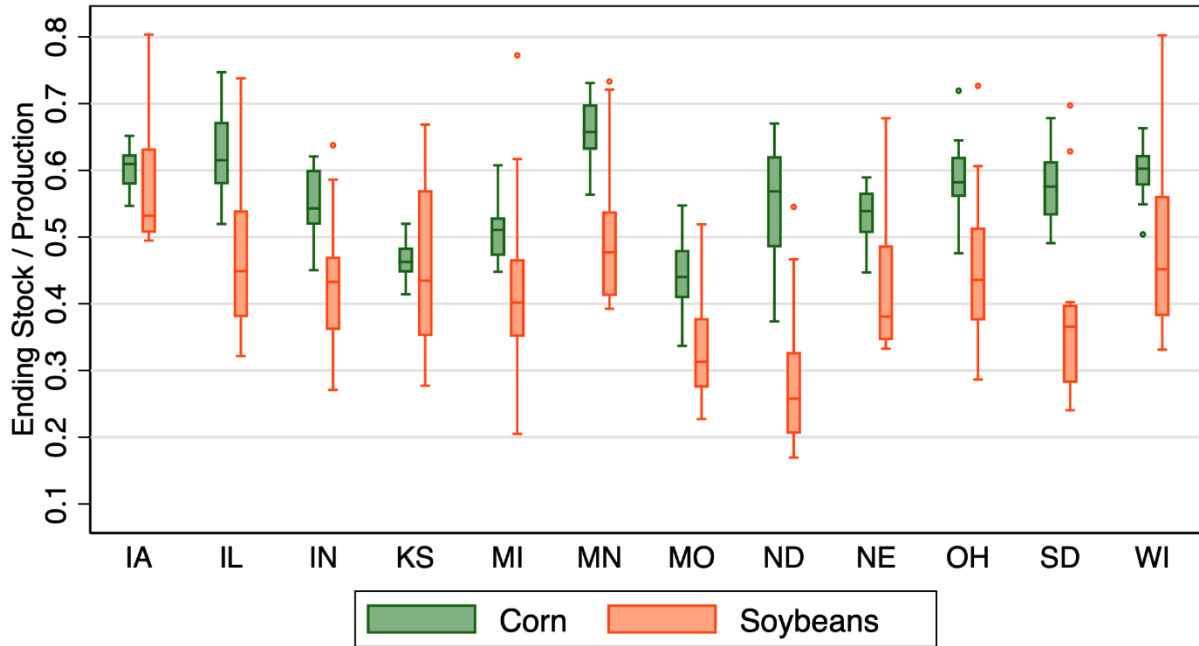
Figure 2A.2 Ratio of Ending Stocks to Production.
Note: Ratio of ending stock to production here is a state-by-year variable. The distribution in a state shows state-level temporal variation over 2009-2020.

Corn
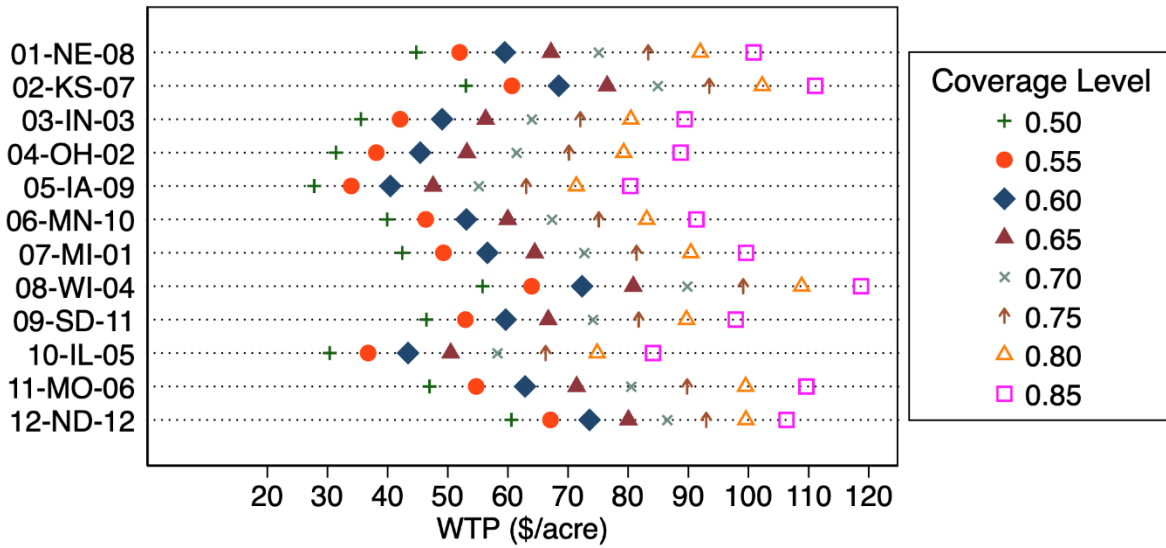
Soybeans
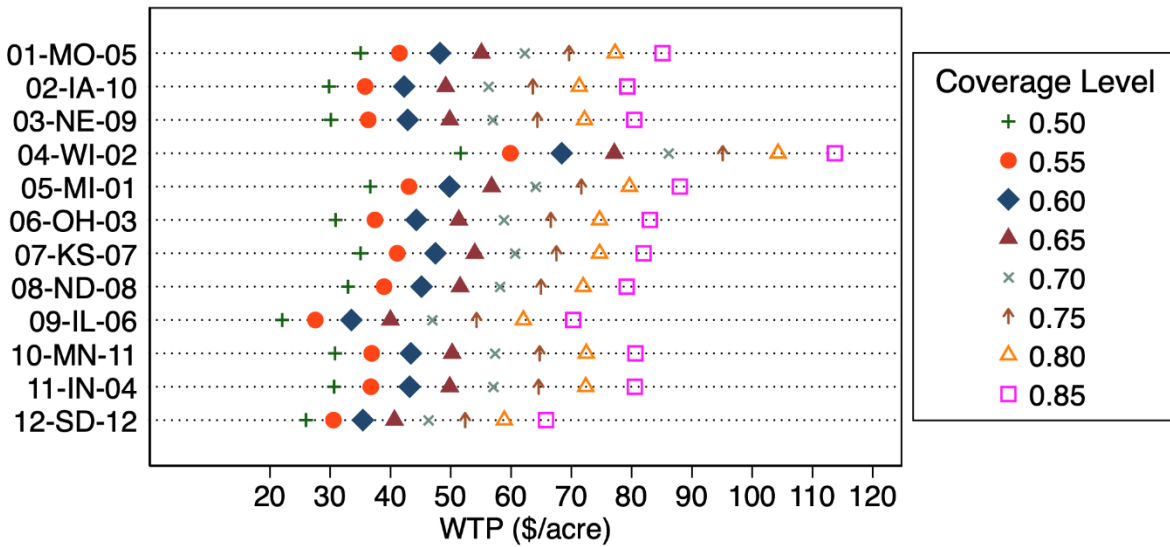
Figure 2A.3 Willingness-to-Pay for A = 0.004 (i.e., risk premium = 0.1).
Note: Number before and after state's name represent ranks of basis risk and extensive margin participation rates. The "01" and "12" indicate the lowest and highest values respectively.
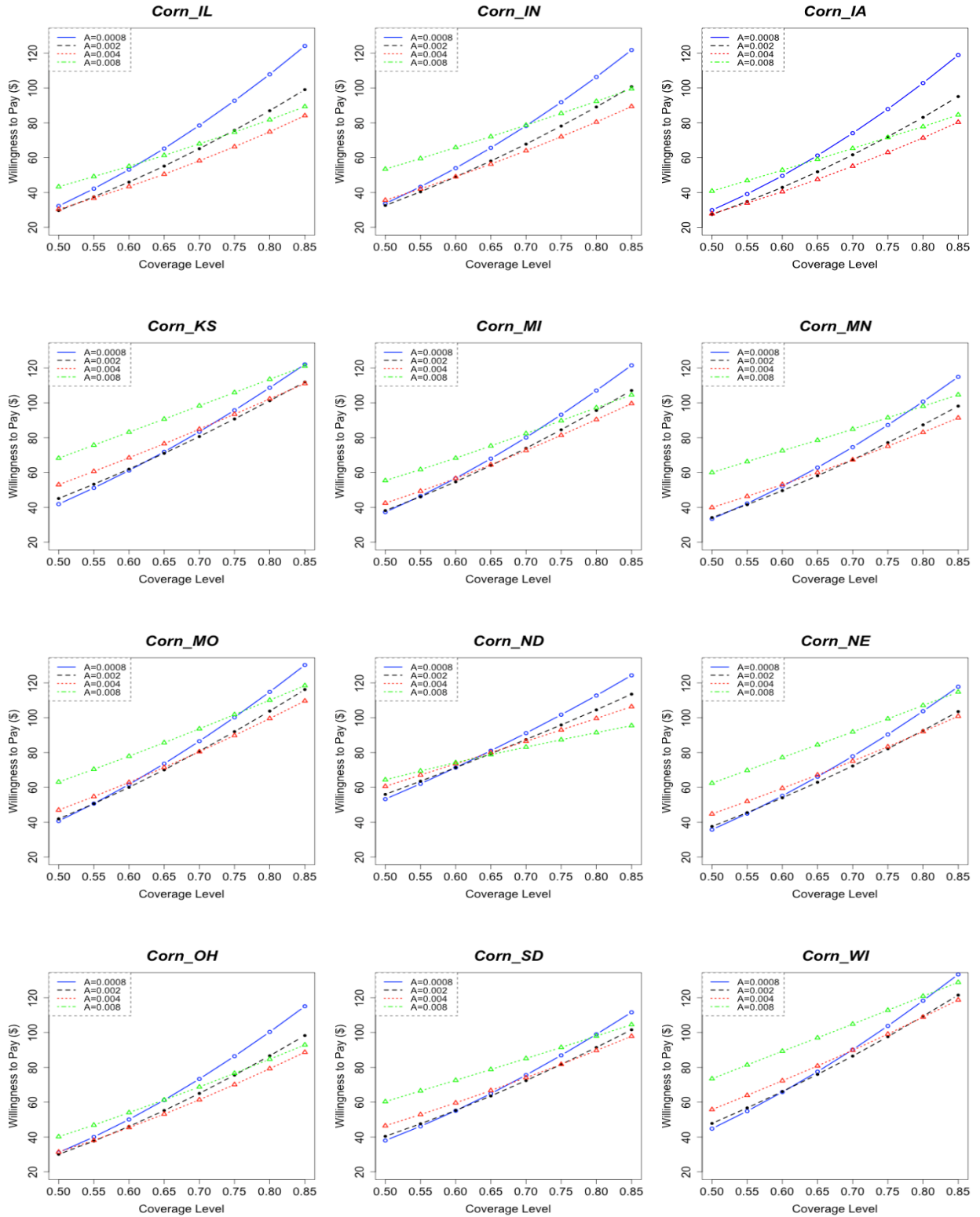
Figure 2A.4 Farmers' Willingness-to-Pay for Corn.
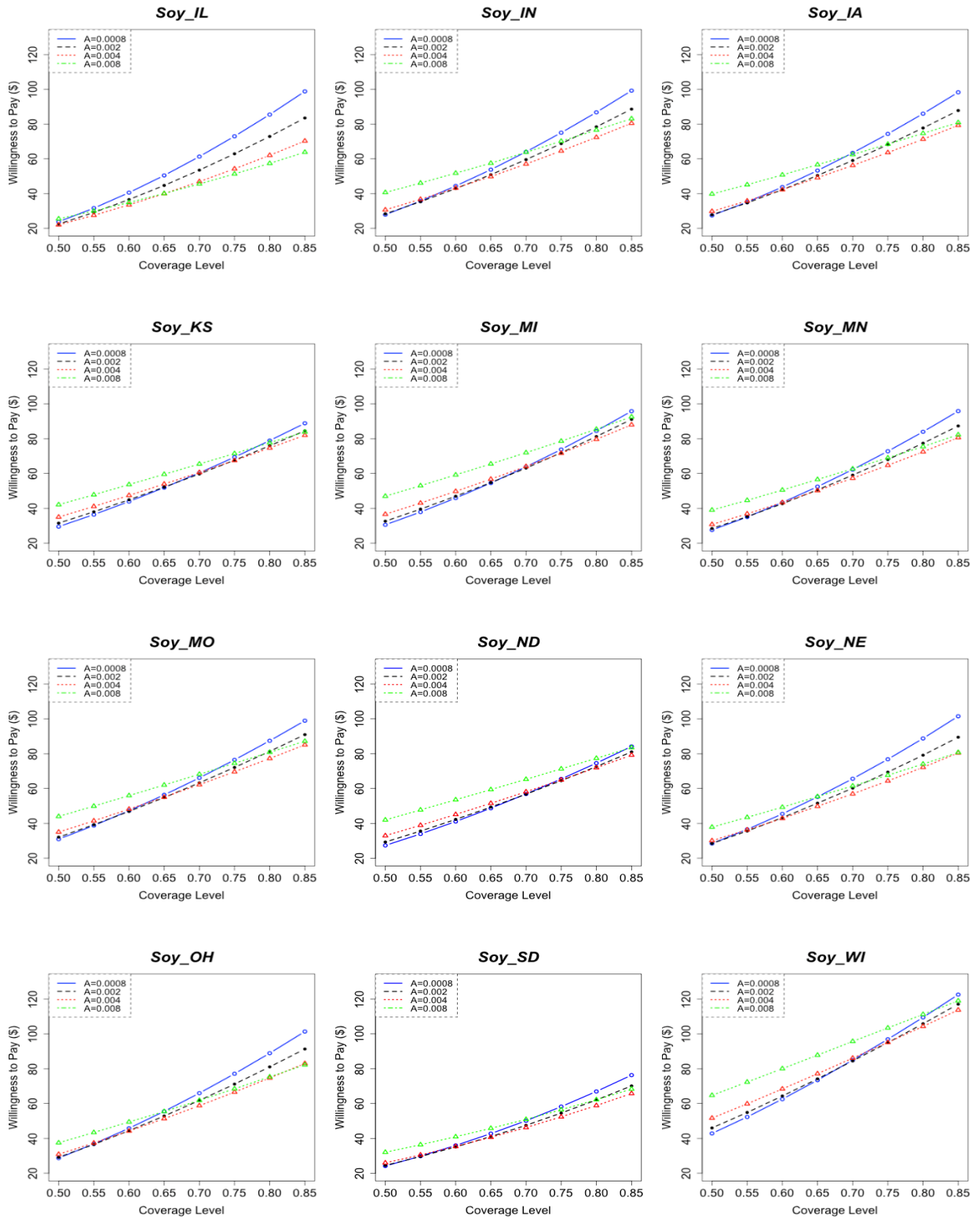
Figure 2A.5 Farmers' Willingness-to-Pay for Soybeans.

CHAPTER 3 Extensive and Intensive Margins of Irrigation Water Demand In the Great Lakes Region

Abstract

The literature has explored water irrigation issues in arid regions; however, few studies investigate irrigation water usage in relatively water-rich areas such as the Great Lakes. Although the water conservation policy was implemented in recent years, there has been an upward trend in irrigation water demand from 2003 to 2018, including irrigated acres, number of pumped wells, and average depth to water. We employ firm-level data from USDA Farm and Ranch Irrigation Survey (FRIS) and Irrigation and Water Management Survey (IWMS) to examine what factors impact farmers' response to irrigation water usage. The findings are: (i) price elasticities of irrigation water usage vary significantly according to model specifications and water costs; (ii) the water usage at both extensive (irrigated acres) and intensive (water application per acre) margins are input price inelastic; and (iii) price elasticities of water usage are homogeneous across crops but heterogeneous across states. Our estimates of price elasticities of water usage can be used to calculate the cost of irrigation water reduction through water pricing, which might shed light on long-term water sustainability.

Introduction

The literature has explored irrigation water issues in the arid regions (U.S. western states), including water conservation (Neigri and Hanchar 1989; Huffaker et al. 1998), the economic impacts of climate change (Deschênes and Greenstone 2007), the effect of electricity subsidies on groundwater extraction (Badiani and Jessoe 2013), the impact of irrigation technology on reduction of groundwater extraction (Pfeiffer and Lin 2014), precision irrigation (Adeyemi et al. 2017), and price elasticities for irrigation water (Howitt et al. 1980; Moore et al. 1994; Hooker and Alexander 1998; Hendricks and Peterson 2012; Mieno and Brozović, N. 2017). However, few studies investigated the issues regarding irrigation water demand in a relatively water-rich area, e.g., the Great Lakes region.

In 2008, the Great Lakes-St. Lawrence River Basin Water Resources Compact became law, requiring the eight U.S. states (IL, IN, MI, MN, NY, OH, PA and WI) to work together to protect water resources in this region. Per this inter-state Compact, all eight states have implemented many regulations for future water management (Lautenberger and Norris 2016). Although the goal of the rules is water conservation, there has been an upward trend in irrigation water demand from 2003 to 2018, including irrigated acres, the number of pumped wells, and the average water depth (see Table 3.1). Investigating why farmers increase water demand is crucial because the depressed water levels might induce costly treatment of a bottomless aquifer groundwater supply.

In this study, we investigate irrigation water usage using farm-level survey data from USDA National Agricultural Statistics Service (NASS) overlying the Great Lakes region. Our motivations mainly come from that: (i) the increasing water demand may induce a concern regarding water scarcity; (ii) current policies to restrict water usage do not seem to be efficient,

so it is necessary to examine the own-price elasticity of irrigation water demand since it is the

critical parameter in determining the impacts of change in water-related policies (Hendricks and

Peterson 2012); (iii) the literature commonly adopts marginal irrigation cost (MIC) (Casewell

and Zilberman 1986; Gonzalez-Alvarez et al. 2006; Mieno and Brozović 2017), and many recent

studies use average energy cost (AEC) (Ito 2014; Kornelis and Norris 2020); however, no

research investigates which water cost is appropriate in the Great Lakes region; (iv) what

underlying factors affect farmer's behavior on groundwater extraction.

Table 3.1 Summary of Water Usage in the Great Lakes Region.

State	Acres Irrigated (thousands)				Acre-feet Water Applied (thousands)			
	2003	2008	2013	2018	2003	2008	2013	2018
Illinois	375	457	542	566	237	231	368	333
Indiana	276	404	455	583	132	199	235	248
Michigan	433	532	588	827	218	298	313	458
Minnesota	435	504	517	555	294	334	324	249
Ohio	14	19	42	39	16	9	26	17
Wisconsin	392	396	473	518	357	322	388	294
State	Number of Pumped Wells				Average Depth to Water (feet)			
	2003	2008	2013	2018	2003	2008	2013	2018
Illinois	3,204	3,857	5,252	5,095	111	117	117	129
Indiana	2,318	2,746	4,133	5,056	83	85	95	96
Michigan	4,031	4,097	7,554	9,691	121	117	105	121
Minnesota	3,797	4,312	5,385	5,905	130	122	131	136
Ohio	975	215	1,064	1,203	123	135	110	103
Wisconsin	3,002	3,192	6,169	5,166	155	147	149	166

Data Source: USDA Farm and Ranch Irrigation Survey (2003,2008,2013); Irrigation and Water
Management Survey (2018).

There is an increasing research interest in the issues regarding water withdrawal and

induced water scarcity in the water-abundant region. Grannemann et al. (2000) note that the

available freshwater supply is more than adequate to meet the Great Lakes region's agricultural,

industrial, residential, and ecological needs. However, some findings and anecdotes exist to presage a potential water scarcity in this region. De Loe and Kreutzwiser (2000) note that climate changes are likely to increase the frequency of shallow levels in the Great Lakes region. Gronewold and Stow (2014) remind that the federal agencies from the United States and Canada documented the lowest water levels recorded on Lakes Michigan and Huron in January 2013. For the Kalamazoo River watershed in Southwest Michigan, Mubako et al. (2013) noted that irrigation water withdrawals in water-abundant regions can lead to surface water scarcity due to seasonal and spatial concentration during low-flow summer months. Kornelis and Norris (2020) investigate price elasticities and climatic determinants of irrigation water demand in the Great Lakes region.

We build on Kornelis and Norris (2020) but further contribute to three primary pieces of literature. First, our analysis provides evidence to understand which kind of water cost is appropriate in the Great Lakes region. Marginal irrigation cost is commonly adopted in studies that analyze water demand in the U.S western states. However, many studies argue that consumers respond to average rather than marginal costs due to the complexity of electricity pricing schedules (Foster and Beattie 1979; Foster and Beattie 1981; Borenstein 2009; Ito 2014). Kornelis and Norris (2020) adopt average energy cost directly but do not provide how marginal irrigation cost performs in the empirical analysis. Since energy employed in western states and the Great Lakes region are the same (e.g., electricity, diesel, and natural gas), we may ask whether marginal irrigation cost is a good predictor for irrigation water usage.

Second, we contribute to estimating price elasticity in the Great Lakes region with multiple model specifications. Price elasticity depends on the water cost construction and model specifications adoption. Our analysis provides evidence that policymakers may overestimate the

price elasticities if they only employ the Tobit Model, which may induce ineffective policies. We compare price elasticities from Heckman Selection Model, Tobit Model and of extensive (irrigated acres/irrigated share) and intensive (water application per acre) margins of irrigation water usage (see, for example, Moore et al. 1994; Mullen et al. 2009). Per estimated price elasticities, we further calculate the cost of reducing irrigation water usage through water pricing.

Third, our analysis provides the estimations of price elasticity with different tolerances of measurement error. The first measurement error is the misreporting of energy expenditure and water application, which generates extremely high or low water costs (outliers). Kornelis and Norris (2020) address the first measurement error by dropping firms with average energy costs above the 95th percentile and below the 5th percentile. The second measurement error exists in extensive and intensive margins of water usage. Farmers report total and crop-specific values for irrigated acres and water application in the questionnaire, but two values may be inconsistent. Kornelis and Norris (2020) address the second measurement error by dropping observations with a difference between two values less than 5 percent. We find that extensive margin is more precise than intensive margin. The reason is that extensive margin can be more readily measured than intensive margin. To obtain robust estimations, we adopt many tolerances of measurement error for the empirical analysis. The third measurement error exists when we impute marginal irrigation cost- as in Mieno and Brozović (2017), omitting the pressure head in the engineering equation will underestimate marginal irrigation cost. Therefore, we first impute Total Dynamic Head (TDH) and then drop outliers.

The rest of this paper is organized as follows. Section 2 reports all data and variables which are employed in this study. Empirical results are in section 3, and then section 4 concludes.

Data and Main Variables

Six states (IL, IN, MI, MN, OH, WI) were selected as our targeted area. The most

important data used in this study are from USDA Farm and Ranch Irrigation Survey (FRIS:

2003/2008/2013) and Irrigation and Water Management Survey (IWMS:2018), which follow the

latest Census of Agriculture (COA: 2002/2007/2012/2017). Specifically, all farms with irrigation

behavior can be identified in COA, and a sample was drawn with a targeted size. Farms are

different in each survey, so our sample is not a panel dataset.

Table 3.2 Summary of Sample Size.

	2018	2013	2008§	2003
Panel A: National Level				
Targeted sample size	35,000	35,000	35,000	NA
Final sample size	34,783	34,966	33,085	25,014
certainty	1,340	2,095	2,738	1,823
noncertainty	33,443	32,871	30,347	23,191
Panel B: Great Lakes Region				
Illinois	329	913	599	569
Indiana	359	886	585	580
Michigan	541	1,103	613	584
Minnesota	412	873	685	695
Ohio	382	697	241	314
Wisconsin	403	905	513	513

Note: "NA" indicates the missing value. "§" means that the sample design in 2008 contains two
parts: (1) for the general sample, the targeted sample size is 25,000 and the final sample size is
23,089; (2) for horticulture sample, the targeted sample size is 10,000 and the final sample size is
9,996. Hence, in Panel B of 2008, state-level values represent farms from the final sample with
23,089 farms.

FRIS/IWMS issued a targeted sample size every survey year, e.g., 35,000 in 2018 (first

column of Table 3.2). All targeted farms were selected in terms of a stratification strategy: (i) a

certainty stratum (with probability 1); and (ii) the remaining strata (with probability less than 1).

Table 3.2 reports the information of the final sample and the farms in the Great Lakes region.

One noteworthy thing is that farms in 2018 in each state of the Great Lakes region declined to

around 50% in 2013 (see Panel B of Table 3.2). However, farms in some western states, e.g., CA, CO, NE, and TX, increased significantly (see Figure 3A.2). FRIS/IWMS did not clearly explain why sample sizes in the Great Lakes region have been changed. In the empirical part, we employ weights provided by USDA FRIS/IWMS to estimate price elasticities.

Sample Selection

There were 13,465 farms in the 4 survey years in total. We dropped farms: (i) without planting acres because there is no irrigation behavior on these farms; and (ii) without crop-specific records because these records can be used to address measurement errors. In final, there are 7,936 farms left for our analysis.

Kornelis and Norris (2020) compare total water usage and the sum of crop-specific water usage, then select farms with a difference of less than 5 percent. Per a similar method, we compare the total and the sum of crop-specific water usage for extensive (irrigated acres) and intensive (water application per acre) margins, respectively. Let TE_i denote the total irrigated acres for the farm i; TI_i denote the total water use (acre-inch). Let CE_i and CI_i denote the sum of crop-specific irrigated acres and water usage, respectively. We then construct two indexes - $DE_i = |CE_i/TE_i - 1|$ and $DI_i = |CI_i/TI_i - 1|$ - for our subsample selection. These two indexes measure the difference between the total value and the sum of crop-specific values. For example, when $DE_i = 0$, that means farm i precisely reported irrigated acres; when $DE_i = 0.1$, the sum of crop-specific irrigated acres may be higher or lower than the total irrigated acres by 10 percent. We can call DE and DI as tolerances for extensive and intensive margin water usages, respectively. It is noteworthy that total water usage record for each farm contains groundwater, on-farm surface water and off-farm water.

Table 3.3 reports the observations for 4 error tolerance criteria - 0, 0.01, 0.05 and 0.1. We

find that measurement for extensive margin (irrigated acres) is more accurate than intensive

margin (water application per acre). For example, farms satisfying the condition "$DE = 0$" are

6,837, accounting for 86.2% of the sample; however, those satisfying the condition "$DI = 0$"

only account for 29.1%. For each error tolerance, the farms for extensive margin are more than

those for intensive margin. One possible explanation is that measurement of water application is

not precise due to a lack of tools such as flow meters. We will do empirical analysis for

subsamples from all four tolerances to avoid biased estimations.

Table 3.3 Summary for Different Precision Conditions.

Irrigated Acres		Water Application		Obs. of Intersection
Condition	Obs.	Condition	Obs.	
$DE = 0$	6,837	$DI = 0$	2,309	2,281
$DE \leq 0.01$	6,916	$DI \leq 0.01$	2,795	2,753
$DE \leq 0.05$	7,063	$DI \leq 0.05$	3,903	3,827
$DE \leq 0.1$	7,177	$DI \leq 0.1$	4,695	4,605
Total			7,936	

Note: 0, 0.01, 0.05 and 0.10 represent error tolerances. For example, "0" indicates that each farm
in this sub-sample has a perfect match between the total value and sum of crop-specific values.
Water application is measured as acre-inch/acre. "Obs. of Intersection" shows both conditions
are satisfied at the same time. For example, the 2,281 (the first row and last column) indicates
how many farms report the exact records for both extensive and intensive margins.

Water Use - Extensive and Intensive Margins

  This study analyses two types of water use: extensive margin (irrigated acres) and the

intensive margin (water application per acre). Extensive margin reflects the cropland allocation

decisions, and intensive margin reflects short-run water application decisions (Moore et al.

1994).

  Figure 3.1 shows state-by-year irrigated acres for five selected states. Ohio is omitted

because irrigated acres are much smaller than in other states. Panel A of Figure 3.1 said that

Panel A

Panel B
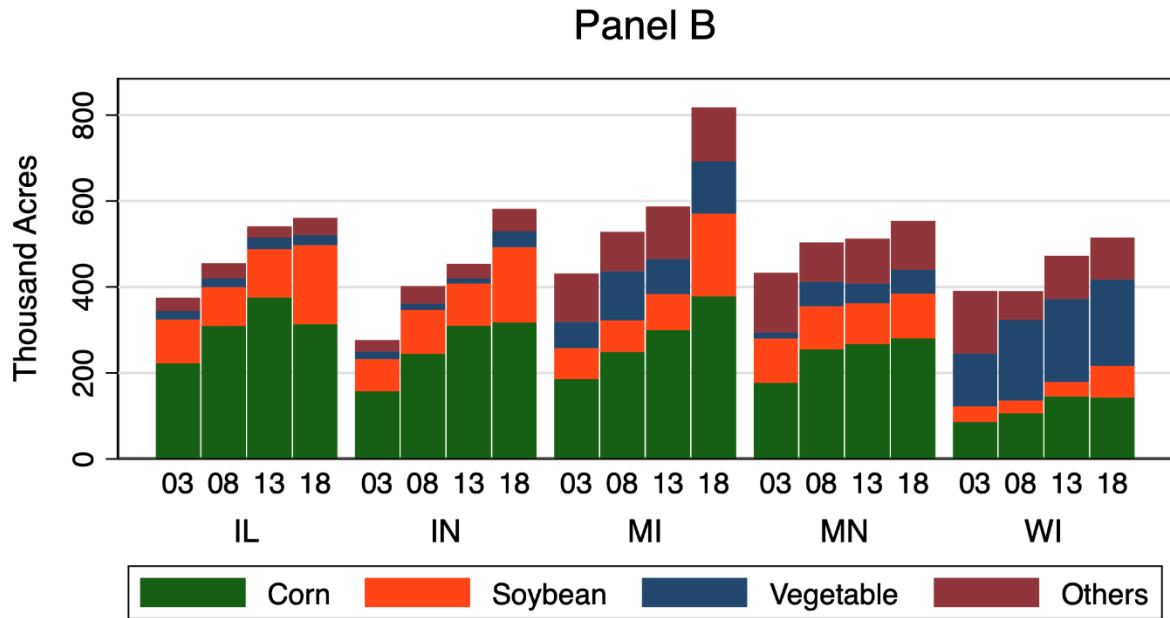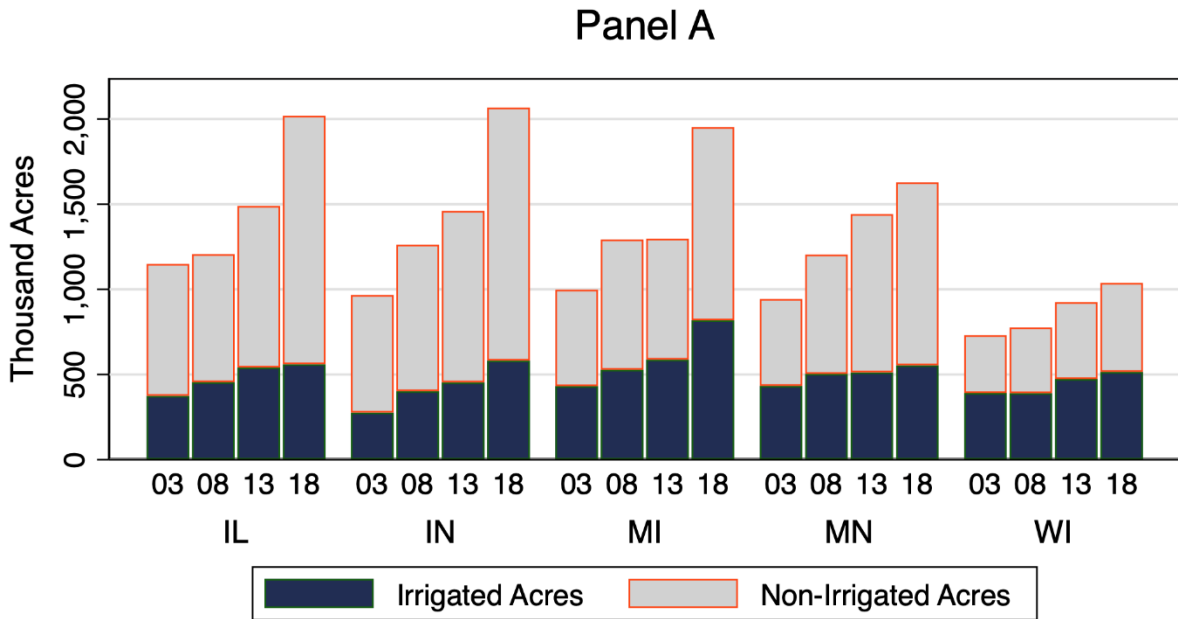
Figure 3.1 Irrigated Acres by Crop, State and Year.
Data Source: FRIS (2003, 2008, 2013) and IWMS (2018) summary report.
Note: Ohio is not included due to the small value of irrigated acres, but it will be in the empirical analysis.
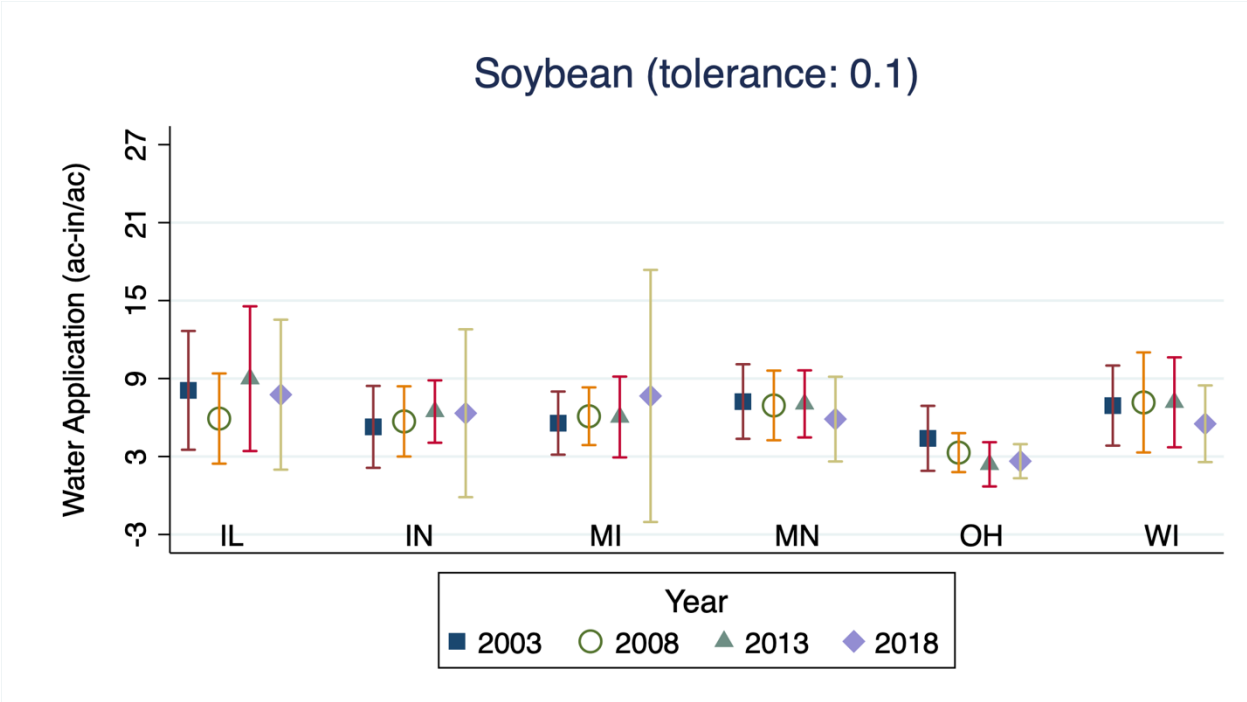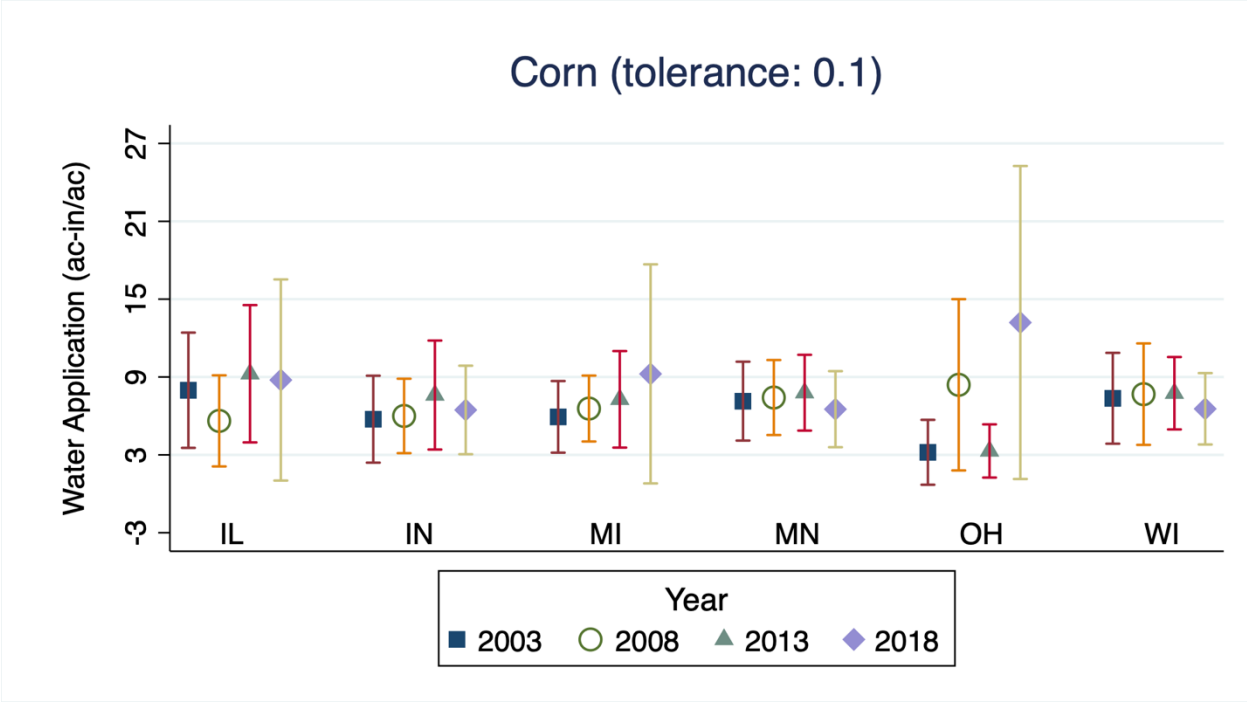
Figure 3.2 Water Application by Crop, State, and Year.

Data Source: Farm-level records in FRIS (2003, 2008, 2013) and IWMS (2018).

Note: The middle point indicates the mean value of firm-level water applications in each (state, year) pair. The range indicates 1 standard deviation. This sample includes all observations with an error tolerance 0.1.

irrigated and non-irrigated acres have an increasing trend in each state. The crop-specific

irrigated acres have a similar pattern (see Panel B of Figure 3.1). In this study, we select corn and

soybeans as the targeted crops for two reasons: (i) both crops account for a large percentage of

the total irrigated acres; and (ii) amounts of farms planting corn and soybeans are far beyond that

of other crops (see Table 3A.1).

Figure 3.2 shows state-by-year water applications (acre inch per acre) with a tolerance of

0.1. More error tolerances will be employed in the empirical analysis. Different from irrigated

acres, water applications are variable across years. There are two interesting points: (1) from

2003 to 2013, water usage has an increasing trend in nearly all states except Illinois and Ohio;

(2) between 2013 and 2018, a decreasing trend exists except for Michigan and Ohio. We may

argue that: (i) Illinois improved irrigation efficiency or had high energy cost in 2008; (ii) after

2013, irrigation efficiency has improved quite a lot.

Irrigation Water Cost

In the literature, there are two irrigation water costs for estimating price elasticities. The

first water cost is the marginal irrigation cost calculated from an engineering equation; the

second water cost is the average energy cost calculated based on energy expenditure and water

usage. Kornelis and Norris (2020) adopt the average cost approach by assuming that the average

water cost may approximate the marginal cost when there are no significant changes in energy

prices during the irrigation season. Farmers report energy expenses and parameters for irrigation

in the FRIS/IWMS, such as well depth and pumping capacity. Therefore, we are motivated to

compare average energy cost and the imputed marginal irrigation cost because: (i) marginal

irrigation cost is commonly used in the studies for U.S. western states (Moore et al. 1994;

Gonzalez-Alvarez et al., 2006; Mullen et al., 2009; Hendricks and Peterson 2012; Pfeiffer and

Lin 2014); and (ii) price elasticity based on different irrigation water cost may help policymakers better understand the potential consequences of tax structure.

Let AEC_i denote average energy cost for farm i; $TotExpen_i$ denote total energy expenditure; TI_i denote the total water usage defined as above. The formula for average energy cost is $AEC_i = TotExpen_i/TI_i$, which is measured as dollars per acre inch. Now we turn to construct marginal irrigation costs based on Static Water Level and Total Dynamic Head, respectively. Why are two irrigation costs imputed? Mieno and Brozović (2017) note that the total dynamic head contains static water level, draw-down, and pressure head, so omitting components in the Total Dynamic Head will induce an underestimation of marginal cost.

Let MIC^S denote marginal irrigation cost based on static water level; MIC^T denote the marginal irrigation cost based on Total Dynamic Head; SW denote the static water level; η denote the pumping efficiency, EP denote the energy price. WHP denote water horsepower (i.e., size of engine Horse-Power); GPM is pumping capacity measured as gallon/min. As in Rogers and Alam (2006), we employ $TDH = WHP \times 3,960/GPM$. Then the two imputed marginal irrigation costs are

(3.1a) $\qquad MIC^S = SW \times \eta \times EP,$

(3.1b) $\qquad MIC^T = (WHP \times 3,960/GPM) \times \eta \times EP$

Table 3.4 The 100% Nebraska Performance Criteria (NPC) for Pumping Plant.

Fuel Type	Fuel Unit	NPC for Pumping Plants	Fuel Units for Lifting 1 acre-foot of water 1 foot
Electricity	kWh	0.885 whp-hr/kWh	1.551
Natural Gas (925 BTU/cf)	mcf	61.7 whp-hr/MCF	0.0223
Diesel	gal	12.50 whp-hr/gal	0.1098
Propane	gal	6.89 whp-hr/gal	0.1993

Note: kWh represents the kilowatt hour. whp-hr represents water horsepower hours. 1 mcf = 1,032 cubic feet. gal represents gallon.

Table 3.5 Average Electricity and Diesel Prices (May-September).

Electricity (cents/kWh)	IL	IN	MI	MN	OH	WI	Diesel ($/gal)	GGE (kWh/gal)	Diesel (cents/kWh)
2003	8.08	6.44	8.11	7.49	7.94	7.93	2.36	37.95	6.22
2008	6.49	4.77	5.96	5.2	5.39	5.77	3.72	37.95	9.8
2013	5.53	6.14	7.23	6.51	5.74	6.9	3.54	37.95	9.34
2018	6.66	7.3	7.09	7.67	6.99	7.43	3.14	37.95	8.27

Note: All price values are adjusted to USD 2018 with the irrigation-seasonally (May to September) average Consumer Price Index (CPI) for all urban consumers. The CPI data are from U.S. Bureau of Labor Statistics. GGE represents Gasoline Gallon Equivalent.

Table 3.6 Farms by Energy Sources and Error Tolerance.

Energy Source	Obs. With Only One	All Obs.	Percentage (All Obs./Total)
No Condition (No constraint; Total=7,936)			
Electricity	3,720	5,971	75.2
Natural Gas	16	64	0.8
LP Gas, Propane	39	220	2.8
Diesel Fuel	1,448	3,667	46.2
Gasoline	170	264	3.3
Condition 1 ($DE = 0; DI = 0$; Total=2,281)			
Electricity	1,043	1,604	70.3
Natural Gas	7	13	0.6
LP Gas, Propane	14	71	3.1
Diesel Fuel	533	1,099	48.2
Gasoline	34	50	2.2
Condition 2 ($DE \leq 0.01; DI \leq 0.01$; Total=2,753)			
Electricity	1,243	1,951	70.9
Natural Gas	10	23	0.8
LP Gas, Propane	15	88	3.2
Diesel Fuel	642	1,350	49.0
Gasoline	39	59	2.1
Condition 3 ($DE \leq 0.05; DI \leq 0.05$; Total=3,827)			
Electricity	1,720	2,830	73.9
Natural Gas	14	41	1.1
LP Gas, Propane	20	119	3.1
Diesel Fuel	805	1,914	50.0
Gasoline	46	76	2.0
Condition 4 ($DE \leq 0.1; DI \leq 0.1$; Total=4,605)			
Electricity	2,070	3,476	75.5
Natural Gas	15	45	1.0
LP Gas, Propane	24	140	3.0
Diesel Fuel	916	2,321	50.4
Gasoline	47	82	1.8

The parameters for irrigation, i.e., SW, WHP, GPM, can be obtained from farm-level

FRIS/IWMS datasets. Pumping efficiency is Nebraska Performance Criteria (NPC) for Pumping

Plants with 100% efficiency (see Table 3.4), commonly used in the literature. State-by-year

energy price data are from the U.S. Energy Information Administration (EIA) (see Table 3.5), all

of which are adjusted to the 2018 dollars. The electricity price measured as cents/kWh is

monthly average retail price in the industrial sector. The diesel price is weekly Midwest No.2

diesel retail price measured as \$/gal. We impute the marginal irrigation cost for farms which: (i)

have at least one well; and (ii) use electricity, diesel, or both. The reasons are that farms using

other energy sources account for a low percentage in the sample (see Table 3.6). It is noteworthy

that both imputed irrigation water costs are the costs for groundwater extraction and that surface

water is not considered.
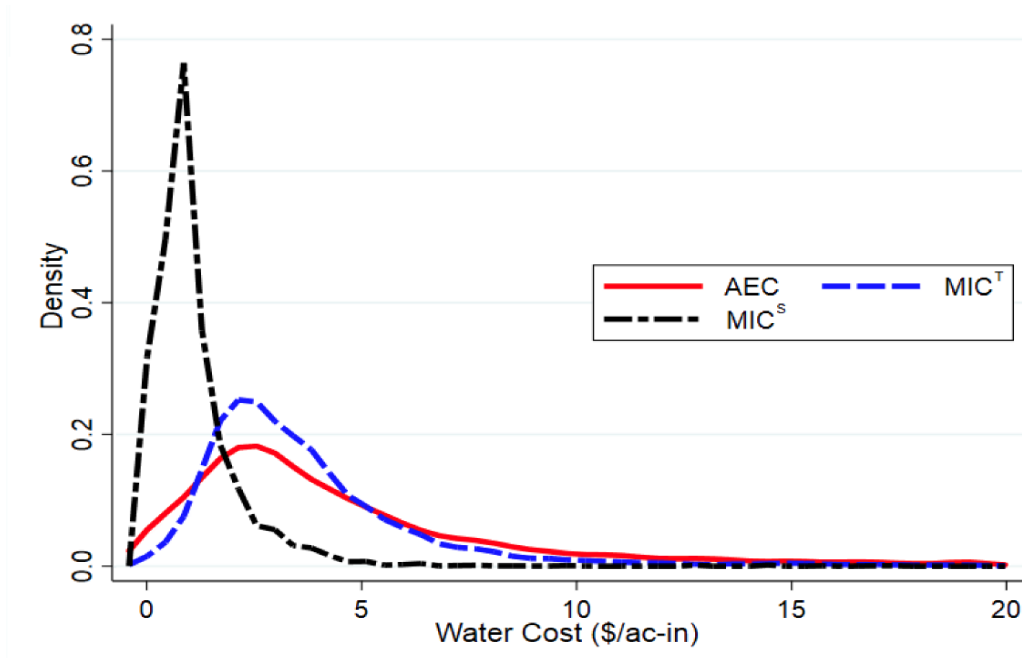
Figure 3.3 Kernel Density Estimation (KDE) for Three Water Costs.
Note: The kernel function is Gaussian function. Bandwidth is calculated by minimizing the mean integrated squared error. Observations with values larger than 20 are excluded.

Figure 3.3 shows the kernel density estimations (KDE) for three irrigation water costs - AEC, MIC^T, MIC^s. We find that: (i) AEC and MIC^T are flatter than MIC^s, indicating irrigation water cost based on static water level is underestimated; (ii) the distribution difference between AEC and MIC^T is not distinct, although MIC^T seems more concentrated around its mode. Figure 3.4 provides the comparison between AEC and MIC^T by state and year. The difference between both water costs is not distinct as well.

Since there are farm-level observations for both water costs, we are motivated to have a paired t-test to examine whether the mean difference between AEC and MIC^T is different from 0. The null hypothesis is that $E[AEC_i - MIC_i^T] = 0$ where i represents farms and $E[\cdot]$ is expectation operation. The t-statistic is 12.15 with 5,972 degrees of freedom. The corresponding two-tailed p-value is 0.00 that is less than 0.01. Therefore, AEC is statistically different from

116

MIC^T, although their distributions are much close.

Figure 3.4 Irrigation Water Costs by State and Year.
Note: The middle point indicates the mean value of firm-level water costs in each (state, year) pair. The range indicates 1 standard deviation.
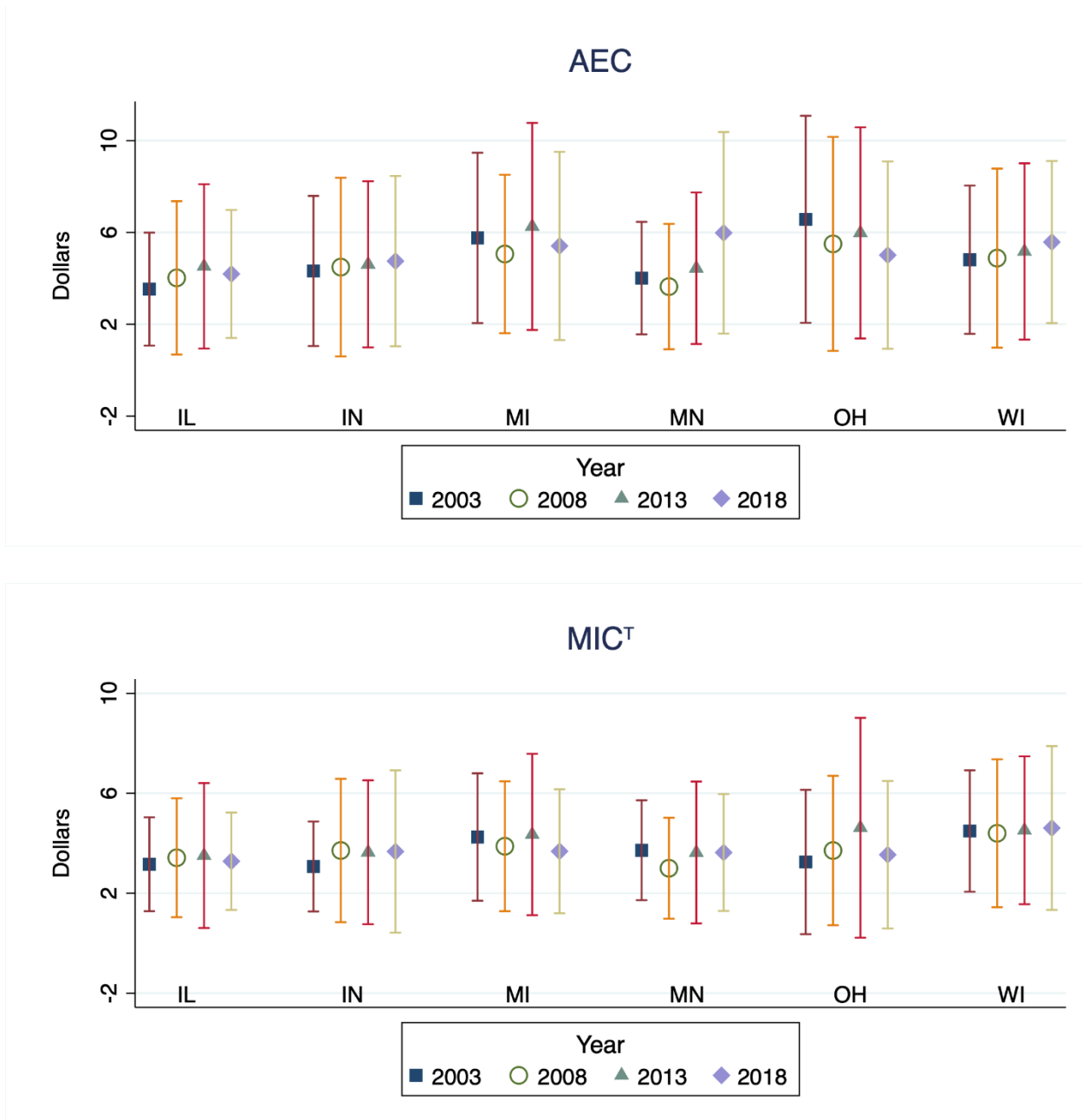
Crop Price

High crop prices are expected to incentivize farmers' irrigation behavior. As a result, farmers might irrigate more acres or increase water application (ac-in/ac) as a profit maximizer. Since the farmer's decision depends on the expected crop prices before the planting time, we employ state-level February price received data published by USDA National Agricultural Statistics Service (NASS). All data are adjusted to 2018 dollars with urban consumers' annual average Consumer Price Index (CPI) from the U.S. Bureau of Labor Statistics.

Weather Determinants

In the literature, many studies have investigated the effects of climate determinants on cropland allocation (Moore et al. 1994; Schlenker and Roberts 2006; Schlenker and Roberts 2009), water quality (Fausey et al. 1995), water management (De Loe and Kreutzwiser 2000), agricultural production (Deschênes and Greenstone 2007), irrigation technology adoption (Loe and Kreutzwiser 2000; Pfeiffer and Lin 2014), and irrigation water demand (Hendricks and Peterson 2012; Kornelis and Norris 2020).

Temperature data employed in this study are from National Oceanic and Atmospheric Administration (NOAA). As in the literature (e.g., Xu et al. 2013; Che et al. 2020), we construct two variables-Growing Degree Days (GDD) and Stress Degree Days (SDD)-to represent beneficial heat and heat stress during the growing season (April to September) respectively. The formulas for variable GDD and SDD in year t are:

(3.2a) $\quad GDD_{c,t} = \sum_{d \in \Omega_t} \left[0.5 \left(\min\left(\max\left(\text{T}^{\max}_{c,d,t}, \text{T}^l\right), \text{T}^h\right) + \min\left(\max\left(\text{T}^{\min}_{c,d,t}, \text{T}^l\right), \text{T}^h\right) \right) - T^l \right]$

(3.2b) $\quad SDD_{c,t} = \sum_{d \in \Omega_t} \left[0.5 \left(\max\left(T^{max}_{c,d,t}, T^k\right) + \max\left(T^{min}_{c,d,t}, T^k\right) \right) - T^k \right]$

Where c is county; d is day; Ω_t is the set of growing season days in year t. The

thresholds are $T^l = 10°C$ (lower bound), $T^h = 30°C$ (upper bound) and $T^k = 32.2°C$. Let $\bar{G}_c = (1/14)\sum_{j=1989}^{2002} GDD_{c,j}$ and $\bar{S}_c = (1/14)\sum_{j=1990}^{2002} SDD_{c,j}$ denote the climatological normals before 2003. Two variables can be constructed for the climate variablity:

(3.3a) $\qquad GD_{c,t} = GDD_{c,t} - \bar{G}_c$

(3.3b) $\qquad SD_{c,t} = SDD_{c,t} - \bar{S}_c$

Land Capability

Land capability data are from National Resource Inventory (NRI). There are eight Land Capability Classes (LCC), where Class I represents the best land and Class VIII represents the worst land. Therefore, land capability can be measured as the percentage of the first two Classes in all Classes (acres for Class I-II/total acres for Class I-VIII).

Farm Size

Hendricks and Peterson (2012) note that the short-run response to water use is smaller than the long-run response because irrigators can respond to price changes by adjusting irrigation capital. We select two indexes - the value of sales or total acres - to control the farm size. The effect of the scale of farm size on the extensive margin (irrigated acres) is expected to be positive because irrigation practice helps increase yields. On the other hand, the effect of the scale of farm size on the intensive margin (water application per acre) is expected to be negative because larger farms are inclined to improve irrigation efficiency by adopting better equipment. Considering the potential multicollinearity of both variables, we will employ total acres as the primary variable controlling for the farm size and use the sales value for the robustness check.

Empirical Analysis

In this empirical part, the dependent variables of most interest are extensive margin (irrigated acres) and the intensive margin (water application per acre). The primary independent

variables are average energy cost and imputed marginal irrigation cost. We will estimate state-level price elasticities of irrigation water usage with different model specifications.

Extensive Margins

Table 3.7 shows that 2,676 farms plant corn without irrigation and 4,423 farms plant soybean without irrigation, accounting for 33.7% and 55.7%, respectively. Because of the censored data, we will start from the Tobit model and then employ a generalized Tobit model-the Cragg hurdle regression.

Table 3.7 Summary of Main Variables.

Variable	Obs.	Mean	Std.Dev
I. Dependent Variable			
Total Irrigated Acres	7,936	606	1,142
Total Water Application (ac-in/ac)	7,818	6.90	4.56
Crop-Specific Irrigation Acres			
Corn (≥ 0)	7,936	290.37	574.56
Corn (> 0)	5,260	438.10	658.30
Soybean (≥ 0)	7,936	106.36	252.94
Soybean (> 0)	3,513	240.27	335.22
Crop-Specific Water Application (ac-in/ac)			
Corn (≥ 0)	7,936	4.55	4.58
Corn (> 0)	5,260	6.87	3.98
Soybean (≥ 0)	7,936	2.77	4.22
Soybean (> 0)	3,513	6.26	4.29
II. Water Irrigation Cost ($/ac-in)			
Average cost	7,434	4.44	3.70
Marginal Cost	6,282	3.77	2.70
III. Control Variable			
Crop Price (Feb, $/bu)			
Corn	7,936	4.74	1.30
Soybean	7,936	10.89	1.77
Weather Determinant			
GD	7,063	-12.59	116.99
SD	7,063	-2.27	4.78
Total Planted Acres	7,936	1466.92	2154.68
Land Capability	7,932	0.40	0.21

Let n_{itl}^* denote the latent irrigated acres for farm i for crop l in year t; Cost_{it} denote irrigation water costs, including average energy cost and imputed marginal irrigation cost; $P_{stl,\text{Fe}}$ denote the crop price in state s in February (projected price before the planting time); $\mathbf{Z_{it}}$ denote a set of explanation variables including county-level weather determinants (GD and SD), county-level land capability and farm-level total acres; ε_{it} denote the error term with $\varepsilon_{it} \sim N[0, \sigma_\varepsilon^2]$. Then the Tobit model specification is:

(3.4) $n_{itl}^* = \alpha_0 + \alpha_1 P_{stl,\text{Fe}} + \alpha_2 \text{Cost}_{it} + \mathbf{Z_{it}'}\widetilde{\alpha} + \varepsilon_{it}$ with $n_{itl} = \begin{cases} n_{itl}^*, & \text{if } n_{itl}^* > 0 \\ 0, & \text{if } n_{itl}^* > 0 \end{cases}$

For the Cragg Hurdle (Double-Hurdle) model, we let s_{itl}^* denote a selection variable for farm i for crop l in year t; $u_{1,itl}$ and $u_{2,itl}$ are error terms with $\begin{pmatrix} u_{1,itl} \\ u_{2,itl} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \sigma^2 \end{pmatrix} \right]$, then the hurdle model specification is as:

(3.5a) first hurdle: $s_{itl}^* = \gamma_0 + \gamma_1 P_{stl,\text{Fe}} + \mathbf{Z_{it}'}\gamma + u_{1,it}$ with $s_{itl} = \begin{cases} 1, & \text{if } s_{itl}^* > 0 \\ 0, & \text{if } s_{itl}^* < 0 \end{cases}$,

(3.5b) second hurdle: $n_{itl}^* = \beta_0 + \beta_1 \text{Cost}_{it} + \mathbf{Z_{it}'}\beta + u_{2,it}$,

Where $\mathbf{Z_{it}}$ is the same as in the Tobit model. The Eqn.(3.5a) and (3.b) generate probit estimator (first hurdle) and Tobit estimator (second hurdle), respectivley. In the Eqn. (3.5a), we assume that farmers' irrigation decisions (whether to irrigate) depend on the projected crop prices. Notably, the correlation of error terms in two hurdle equations is assumed as 0.

Intensive Margin

As discussed, misreporting for the intensive margin (water application) is not uncommon (see Table 3.4). As in the literature, the OLS estimator would be unbiased when the dependent variable is mismeasured (Bound and Krueger 1991; Bound et al. 1994; Hausman 2001). Therefore, we employ the OLS estimation for water application per acre w_{itl} for farm i for crop

l in year t. The reduced form equation is as

$$(3.6) \quad w_{itl} = \kappa_0 + \kappa_1 \text{Cost}_{it} + \kappa_2 P_{Stl,\text{Fe}} + \tilde{\mathbf{Z}}'_{\mathbf{it}}\boldsymbol{\kappa} + c_i + \gamma_s + \kappa_t + u_{itl}, \; w_{itl} > 0$$

The OLS estimation is unbiased if $\text{cov}(\text{Cost}_{it}, c_i) = 0$ which satisfies the exogenous condition. However, the condition that the pumping cost is exogenous to the farmer is too restrictive because some firm-level characteristics such as well depth indeed affect the pumping costs (see, for example, Hendricks and Peterson 2012). For robustness checks, we also provide regression results from Tobit model.

Empirical Results

In the sample selection, we consider farmers' misreporting for irrigated acres and water application. Another measurement error is from farmers' misreporting for energy expenditure. Kornelis and Norris (2020) dropped the firms above the 95th percentile and below the 5th percentile for the average energy cost of water. In our data, the 5th and 95th percentiles for average energy cost are \$0.53 and \$25, respectively; those for imputed marginal irrigation cost based on Total Dynamic Head are \$1.11 and \$19.1. Therefore, we employ average energy cost and marginal irrigation cost between the 5th and 95th percentiles for our analysis, and then use other percentiles for the robustness checks.

Based on FRIS/IWMS survey weights, we first use the Tobit model for the sum of crop-specific irrigated acres, and the Pooled OLS for total water application. Table 3.8 reports regression results for all four error tolerances (0, 0.01, 0.05, 0.1) for two water costs while controlling for other determinants. We find that: (i) elasticities from AEC are a little larger than those from MIC in general, no matter for extensive or intensive margins; (ii) AEC has a better performance than MIC for irrigated acres (see F-value of regression), but there is a contrary case for water application (see R^2 of regression); and (iii) estimation results are consistent when the

error tolerances change.

Now we turn to examine crop-specific price elasticities. For the extensive margin of irrigation water usage for corn, as in Table 3.9, the price elasticities of the MIC are lower than those of the AEC. The underestimation of price elasticities of the MIC might induce a high tax structure when the policymakers' target for the reduction of irrigated acres is pre-determined. The price elasticities of the Cregg Hurdle model (around 0.163) are half that of the Tobit model (about 0.29), which implies that policies based on estimation results in the Tobit model might overestimate the outcome of water reduction.

The upper panel of Figure 3.5 shows the price elasticities for the extensive margin by state, crop, and model specification based on average energy cost. The findings are: (i) price elasticities of OH are higher than that of other states; (ii) for corn, price elasticities of the Tobit model are higher than that of the Cregg Hurdle model in IN, MI, MN, and WI, but the contrary case happens in IL, and OH; and (iii) for soybean, price elasticities of the Tobit model are higher than that of the Cregg Hurdle model in all states except OH.

Kornelis and Norris (2020) report that the price elasticities of water application (ac-in/ac) for corn and soybeans are -0.3 and -0.3, respectively, consistent with our results (see Table 3.10). However, we find that interstate heterogeneity exists. As in the lower panel of Figure 3.5, OH has the maximum value of price elasticity (-0.08), and MI has the minimum value (-0.28), which implies that water reduction will differ among states if a uniform tax structure on average energy cost is provided.

The Effect of Potential Tax Structure on Water Reduction

Estimated price elasticities can be used for predicting water reduction under different tax structures. Price elasticities for extensive margin (irrigated acres) are from

the Cregg Hurdle model, and those for intensive margin (water application) are from

Pooled OLS estimation. We set a hypothetical scenario that there is a 10% tax for average

energy cost. All else being equal (e.g., equipment count and irrigation efficiency), the tax

on average energy cost is equivalent to that on energy price. The irrigated acres, water

application, and total water usage in 2018 are employed as the benchmark. Total water

usage is the multiplication of irrigated acres and water application.

Table 3.11 reports that: (i) average water reductions for corn and soybean are

4.2% and 4.1%, respectively; (ii) The corn change is stabler than soybean because

changes in total water usage for corn are around 4% in all states except OH, but those for

soybean vary over 2.6%-8.3%. The 10% tax on average energy cost is arbitrarily

selected. When a higher tax is chosen by policymakers, e.g., 20%, average water

reductions for corn and soybean will be 7.2% and 8.1%, respectively.

Table 3.8 Price Elasticities for Extensive and Intensive Margins.

Dependent Variable	Total Irrigated Acres							
Estimation Method	Tobit (lower bound=1)							
Tolerance	0	0.01	0.05	0.1	0	0.01	0.05	0.1
Average Energy Cost	-0.053***	-0.053***	-0.051***	-0.049***				
Marginal Irrigation Cost					-0.039***	-0.039***	-0.038***	-0.037***
Land Capability	-0.06***	-0.06***	-0.06***	-0.06***	-0.07***	-0.08***	-0.08***	-0.08***
Total Acres (,000)	0.24***	0.24***	0.25***	0.25***	0.26***	0.27***	0.27***	0.27***
GD	-0.0005	-0.0005	-0.0006	-0.0006	-0.002	-0.0002	-0.0003	-0.0003
SD	0.008***	0.007***	0.007**	0.007**	0.008***	0.007**	0.007**	0.007**
Obs.	5,410	5,476	5,578	5,666	4,588	4,648	4,745	4820
Population Size	23,306	23,537	24,093	24,653	17,870	18,110	18,596	19,073
F-value of Regression	40.72	40.85	42.05	43.88	17.95	18.07	19.29	20.16
Dependent Variable	Water Application (ac-in/ac)							
Estimation Method	Pooled OLS							
Tolerance	0	0.01	0.05	0.1	0	0.01	0.05	0.1
Average Energy Cost	-0.26***	-0.26***	-0.23***	-0.23***				
Marginal Irrigation Cost					-0.21***	-0.19***	-0.18***	-0.17***
Land Capability	0.03	0.005	-0.06	-0.07*	-0.33***	-0.38***	-0.51***	-0.49***
Total Acres (,000)	-0.03***	-0.03***	-0.03***	-0.02***	0.65***	0.65***	0.81***	0.78***
GD	-0.003	-0.001	-0.0001	-0.0001	0.05**	0.04**	0.01*	0.004**
SD	-0.03**	-0.02**	-0.014**	-0.014**	-0.0002	-0.01	-0.001	0.006
Obs.	1,803	2,186	3,097	3,748	1,515	1,855	2,661	3,252
Population Size	7,308	8,802	11,970	14,488	5,419	6,629	9,285	11,434
R^2 of Regression	0.17	0.17	0.15	0.15	0.23	0.28	0.35	0.33

Note:*,** and *** represent 10%, 5% and 1% significance levels, respectively. All regressions include state fixed effect controlling for macro-level shock and year fixed effect controlling for the yearly differential.

Table 3.9 Price Elasticities of Extensive Margins for Corn.

Dependent Variable	Irrigated Acres							
Crop	Corn							
Estimation Method	Tobit (lower bound=0)							
Tolerance	0	0.01	0.05	0.1	0	0.01	0.05	0.1
Average Energy Cost	-0.293***	-0.297***	-0.29***	-0.29***				
Marginal Irrigation Cost					-0.178***	-0.179***	-0.179***	-0.171***
Land Capability	0.17***	0.17***	0.17***	0.18***	0.17***	0.17***	0.17***	0.17***
Total Acres (,000)	0.49***	0.49***	0.51***	0.51***	0.50***	0.50***	0.53***	0.53***
GD	-0.002	-0.002	-0.002	-0.002	0.0002	0.0003	-0.0002	-0.0002
SD	0.02**	0.02**	0.02**	0.02**	0.02**	0.02**	0.02**	0.02**
Obs.	5,410	5,476	5,578	5,666	4,588	4,648	4,745	4,820
F-value	44.4	44.97	44.7	43.82	32.36	32.8	33.98	34.52
Estimation Method	Cregg Hurdle (lower bound=0)							
Tolerance	0	0.01	0.05	0.1	0	0.01	0.05	0.1
Average Energy Cost	-0.164***	-0.164***	-0.163***	-0.162***				
Marginal Irrigation Cost					-0.117***	-0.125***	-0.132***	-0.132***
Land Capability	0.19***	0.19***	0.19***	0.19***	0.20***	0.20***	0.20***	0.20***
Total Acres (,000)	0.51***	0.51***	0.51***	0.51***	0.5***	0.5***	0.5***	0.5***
GD	0.009***	0.009**	0.009***	0.009***	0.01**	0.01**	0.01**	0.01**
SD	0.03**	0.03***	0.03***	0.03***	0.02**	0.02**	0.02**	0.02**
Obs.	5,410	5,476	5,578	5,666	4,588	4,648	4,746	4,820
Pseudo R^2	0.034	0.034	0.034	0.034	0.032	0.032	0.032	0.032
p-value of lnsigma	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note: *, ** and *** represent 10%, 5% and 1% significance levels respectively. Linearized standard errors are in the parenthesis.

Table 3.10 Price Elasticities of Intensive Margins for Corn.

Dependent Variable	Water Application (ac-in/ac)							
Crop	Corn							
Estimation Method	Tobit (lower bound=0)							
Tolerance	0	0.01	0.05	0.1	0	0.01	0.05	0.1
Average Energy Cost	-0.27***	-0.28***	-0.27***	-0.27***				
Marginal Irrigation Cost					-0.10***	-0.11***	-0.056	-0.051
Land Capability	-0.11**	-0.13**	-0.16**	-0.16**	-0.1*	-0.17***	-0.21***	-0.19***
Total Acres (,000)	-0.003	-0.009	-0.001	-0.004	-0.016	-0.024	-0.011	-0.01
GD	0.011*	0.005*	0.002*	-0.001*	0.0103	0.002	0.0049	0.00007
SD	-0.03***	-0.03***	-0.02***	-0.02***	-0.05***	-0.014	-0.19*	-0.02*
Obs.	1,328	1,621	2,345	2,888	1,196	1,467	2,143	2,647
F-value	19.46	17.38	24.89	30.16	3.03	3.34	5.89	4.73
Estimation Method	Pooled OLS							
Tolerance	0	0.01	0.05	0.1	0	0.01	0.05	0.1
Average Energy Cost	-0.27***	-0.29***	-0.27***	-0.27***				
Marginal Irrigation Cost					-0.11***	-0.12***	-0.06	-0.06
Land Capability	-0.11**	-0.13***	-0.16***	-0.16***	-0.1*	-0.18***	-0.23***	-0.21***
Total Acres (,000)	-0.003	-0.01	-0.002	-0.004	-0.017	-0.025	-0.012	-0.01
GD	0.01*	0.01	0.002	-0.001	0.01	0.003	0.005	0.0001
SD	-0.04***	-0.03***	-0.02***	-0.02***	-0.05***	-0.02	-0.02*	-0.02*
Obs.	1,328	1,621	2,345	2,888	1,196	1,467	2,143	2,647
R^2	0.26	0.24	0.22	0.21	0.08	0.07	0.05	0.04

Note: *, ** and *** represent 10%, 5% and 1% significance levels respectively. Linearized standard errors are in the parenthesis.
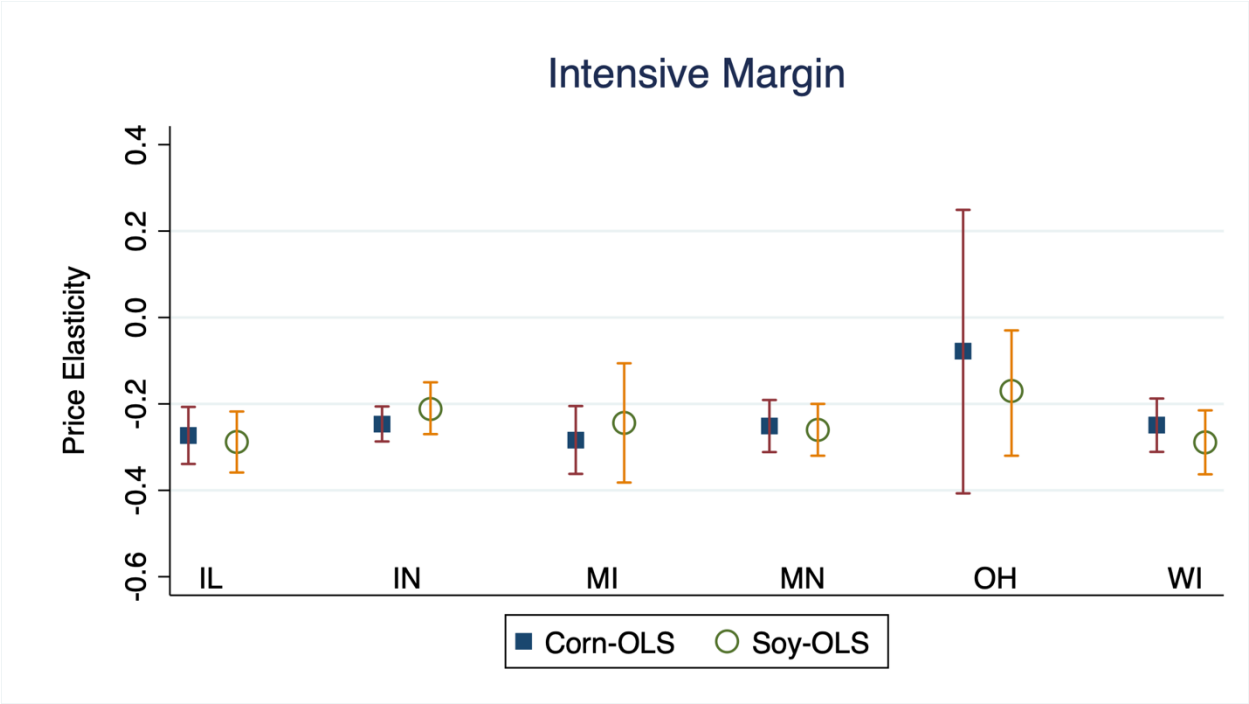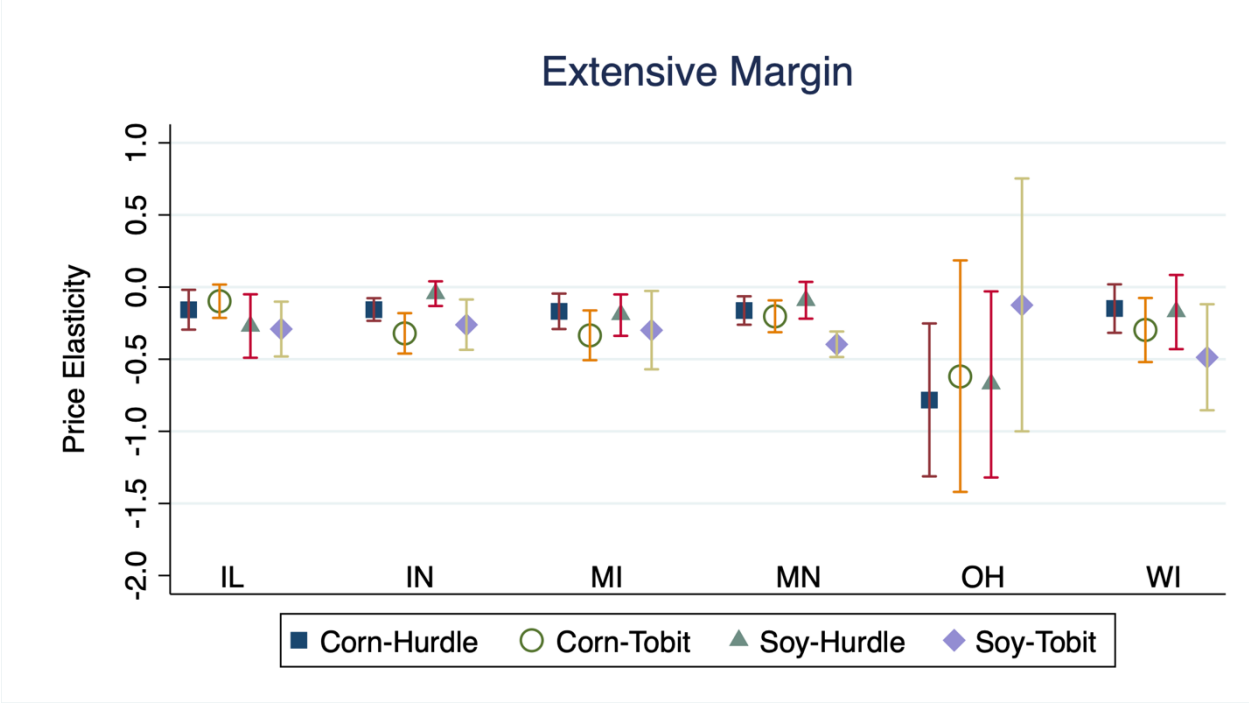
Figure 3.5 Price Elasticities for Extensive and Intensive Margins.
Note: Water cost used is average energy cost. Standard errors are estimated by Taylor linearization with weights from USDA NASS. The range of each estimation represents 95% confidence interval. Average Energy Cost are employed for the estimation. Price elasticities are estimated when other variables are at their mean values.

Table 3.11 Predicted Water Reduction with Estimated Price Elasticities based on Average Energy Cost.

			IL	IN	MI	MN	OH	WI
Corn								
	Year 2018	Irrigated Acres (,000)	313.4	317.2	378.7	280.4	8.0	142.9
		Water Application (ac-in/ac)	7.2	4.8	6.0	6.0	8.4	6.0
		Total Water Usage (ac-in)	2256.5	1522.6	2272.2	1682.4	67.2	857.4
	10% Tax for Average Energy Cost	Irrigated Acres (,000)	308.5	312.3	372.3	275.8	7.4	140.8
		Water Application (ac-in/ac)	7.0	4.7	5.8	5.8	8.3	5.9
		Total Water Usage (ac-in)	2160.4	1461.8	2170.6	1613.4	61.5	823.7
		Change of Total Water Usage (%)	-4.3	-4.0	-4.5	-4.1	-8.5	-3.9
Soybean								
	Year 2018	Irrigated Acres (,000)	184.4	175.3	192.5	104.2	4.7	73.8
		Water Application (ac-in/ac)	6.0	6.0	6.0	4.8	2.4	4.8
		Total Water Usage (ac-in)	1106.4	1051.8	1155.0	500.2	11.3	354.2
	10% Tax for Average Energy Cost	Irrigated Acres (,000)	179.4	174.5	188.8	103.2	4.4	72.5
		Water Application (ac-in/ac)	5.8	5.9	5.9	4.7	2.4	4.7
		Total Water Usage (ac-in)	1045.5	1024.9	1105.4	482.7	10.3	338.2
		Change of Total Water Usage (%)	-5.5	-2.6	-4.3	-3.5	-8.3	-4.5

Data Source: USDA Farm and Ranch Irrigation Survey (2003,2008,2013); Irrigation and Water Management Survey (2018). Estimated price elasticities are from empirical analysis.

Conclusion

Few studies focus on the potential water scarcity issue in a water-rich region. As such, more research is needed better to understand irrigation water issues in the Great Lakes region. By employing USDA FRIS/IWMS data from 2003 to 2018, we investigate what kind of water cost farmers respond to, estimate multiple price elasticities of irrigation water usage, and predict the effect of a potential tax on irrigation water reduction.

Our findings contain that: (i) AEC and MIC have similar performance for values (extensive and intensive margins) which include all crops, although price elasticities based on MIC are slightly smaller; (ii) for crop-specific values, price elasticities based on MIC might be underestimated and be sensitive to error tolerance; (iii) if a 10% tax structure on average energy cost were employed by policymakers, the total water usage will decrease by about 4% for both corn and soybean; and (iv) when average energy cost increases by 20% due to a tax policy, the total water usage will decrease by 7.2% and 8.1% for corn and soybean, respectively.

The external validities of our findings are that: (i) price elasticities in previous studies investigating U.S. western states might be underestimated because of the adoption of a uniform irrigation pumping efficiency; (ii) model specification impacts the estimations of price elasticity of extensive margin water usage, and (iii) interstate price elasticities are heterogeneous, so a uniform tax structure will induce a differential of water reduction when all else is equal.

REFERENCES

Baum, C.F. 2008. Modeling proportions. *Stata Journal* 8: 299–303.

Borenstein, S. 2009. To what electricity price do consumers respond? Residential demand elasticity under increasing-block pricing. Preliminary Draft April, 30, 95.

Bound, J., and A.B. Krueger. 1991. The extent of measurement error in longitudinal earnings data: Do two wrongs make a right?. *Journal of labor economics*, 9(1), 1-24.

Bound, J., C. Brown, G.J. Duncan, and W.L. Rodgers. 1994. Evidence on the validity of cross-sectional and longitudinal labor market data. *Journal of Labor Economics*, 12(3), 345-368.

Burton, I.: 1996, The growth of adaptation capacity: practice and policy'. In Adapting to Climate Change: An International Perspective, eds. J.B. Smith, N. Bhatti, G.V. Menzhulin, R. Benioff, M. Campos, B. Jallow, F. Rijsberman, M.1. Budyko, and RK Dixon, 55-67, Springer, New York.

Caswell, M. F., and D. Zilberman. 1986. The effects of well depth and land quality on the choice of irrigation technology. *Amer. J. of Agr. Econ*. 68(4): 798-811.

Che, Y., H. Feng, and D.A. Hennessy. 2020. Recency effects and participation at the extensive and intensive margins in the US Federal Crop Insurance Program. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 45(1), 52-85.

Chiu, Y. W., B. Walseth, and S. Suh. Water embodied in bioethanol in the United States. *Environmental Science and Technology* Vol. 43, No.8: 2688-92, 2009.

Condon, N., H. Klemick, and A. Wolverton. 2015. Impacts of ethanol policy on corn prices: A review and meta-analysis of recent evidence. *Food Policy*, 51: 63-73.

Croley, T.I. II.: 1991. 'CCC GCM 2xC02 hydrological impacts on the Great Lakes'. Working Committee 3, International Joint Commission Levels Reference Study.

De Loe, R. C., and R.D. Kreutzwiser. 2000. Climate variability, climate change and water resource management in the Great Lakes. *Climatic Change*, 45(1): 163-179.

Deschênes, O., and M. Greenstone. 2007. The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather. *American Economic Review*, 97(1): 354-385.

Döll, P. 2002. Impact of climate change and variability on irrigation requirements: a global perspective. *Climatic change*, 54(3): 269-293.

Foster, H. S., and B.R. Beattie. 1979. Urban residential demand for water in the United States. *Land Economics*, 55(1): 43-58.

Foster, H. S., and B.R. Beattie. 1981. Urban residential demand for water in the United States: reply. *Land Economics*, 57(2): 257-265.

Gonzalez-Alvarez, Y., A.G. Keeler, and J.D. Mullen. 2006. Farm-level irrigation and the marginal cost of water use: Evidence from Georgia. *Journal of Environmental Management*, 80(4): 311-317.

Grannemann, N. G., et al. 2000. The importance of ground water in the Great Lakes Region (No. 2000-4008). US Geological Survey.

Hausman, J. 2001. Mismeasured variables in econometric analysis: problems from the right and problems from the left. *Journal of Economic perspectives*, 15(4): 57-67.

Hendricks, N.P., and J.M. Peterson. 2012. "Fixed Effects Estimation of the Intensive and Extensive Margins of Irrigation Water Demand." *Journal of Agricultural and Resource Economics*, 37(1): 1–19.

Hendricks, N. P. 2018. Potential benefits from innovations to reduce heat and water stress in agriculture. *Journal of the Association of Environmental and Resource Economists*, 5(3): 545-576.

Hrozencik, R. A., D.T. Manning, J.F. Suter, and C. Goemans. (2022). Impacts of Block-Rate Energy Pricing on Groundwater Demand in Irrigated Agriculture. *Amer J. of Agri Econ*, 104(1): 404-427.

Ito, K. 2014. Do consumers respond to marginal or average price? Evidence from nonlinear electricity pricing. *American Economic Review*, 104(2): 537-63.

Kornelis, A., and P. Norris. 2020. Irrigation Water Demand: Price Elasticities and Climatic Determinants in the Great Lakes Region. *Agric. & Resour. Econ. Rev*. 49(3): 437-464.

Lautenberger, M. C., & Norris, P. E. 2016. Private rights, public interests and water use conflicts: evolving water law and policy in Michigan. Water Policy, 18(4), 903-917.

Mieno, T., and N. Brozović. 2016. Price elasticity of groundwater demand: attenuation and amplification bias due to incomplete Information. *Amer J. of Agri Econ,* https://doi.org/10.1093/ajae/aaw089.

Moore, M. R., N.R. Gollehon, and M.B. Carey. 1994. Multicrop production decisions in western irrigated agriculture: the role of water price. *Amer J. of Agri Econ*. 76(4): 859-874.

Mubako, S. T., B.L. Ruddell, and A.S. Mayer (2013). Relationship between water withdrawals and freshwater ecosystem water scarcity quantified at multiple scales for a Great Lakes watershed. *J. Water Resour. Plan. & Manage*, 139(6): 671-681.

Nataraj, S., and W.M. Hanemann. 2011. Does marginal price matter? A regression discontinuity approach to estimating water demand. *Journal of Environmental Economics and Management*, 61(2): 198-212.

Papke, L. E., and J.M. Wooldridge. 1996. Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. Journal of applied econometrics, 11(6): 619-632.

Papke, L. E., and J.M. Wooldridge. 2008. Panel data methods for fractional response variables with an application to test pass rates. *Journal of econometrics*, 145(1-2): 121-133.

Pfeiffer, L., and C.Y.C. Lin. 2014. Does efficient irrigation technology lead to reduced groundwater extraction? Empirical evidence. *Journal of Environmental Economics and Management*, 67(2): 189-208.

Rogers, D. H. 2006. Pumping plant efficiency, fuel options and costs. In Proceedings for 2006 Central Plains irrigation conference, Colby, Kansas, February 21-22. Colorado State University. Libraries.

Rogers, D. and M. Alam. 2006. "Comparing Irrigation Energy Costs." Irrigation Management Series MF-2360, Kansas State University, 2006. Available online at http://www.ksre.ksu.edu/library/ ageng2/mf2360.pdf.

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

USDA, N. 2018. Irrigation and Water Management Survey. Summary available at: https://www.nass.usda. gov/Publications/Highlights/2019/2017Census_Irrigation_and_ WaterManagement. pdf.

Wooldridge, J. M. 2005. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of applied econometrics*, 20(1): 39-54.

Wooldridge, J.M., 2010. Econometric Analysis of Cross Section and Panel Data, second ed. MIT Press, Cambridge, MA.

Wooldridge, J. M. 2019. Correlated random effects models with unbalanced panels. *Journal of Econometrics*, 211(1): 137-150.

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

Table 3A.1 Crop-Specific Farm Amounts (N=7,936).

	Crop	Obs.		Crop	Obs.
01	Corn for grain or seed	5,260	07	Wheat for grain or seed	347
02	Soybeans for beans	3,513	08	Beans, dry edible	267
03	Alfalfa and alfalfa mixtures	750	09	Sorghum for grain or seed	21
04	Potatoes	679	10	Rice	6
05	All berries	661	11	Cotton	3
06	Corn for silage or greenchop	512	12	Peanut	0

Note: Observations in our final sample are 7,936. 12 crop categories are selected in this table with the rank of farm amounts.
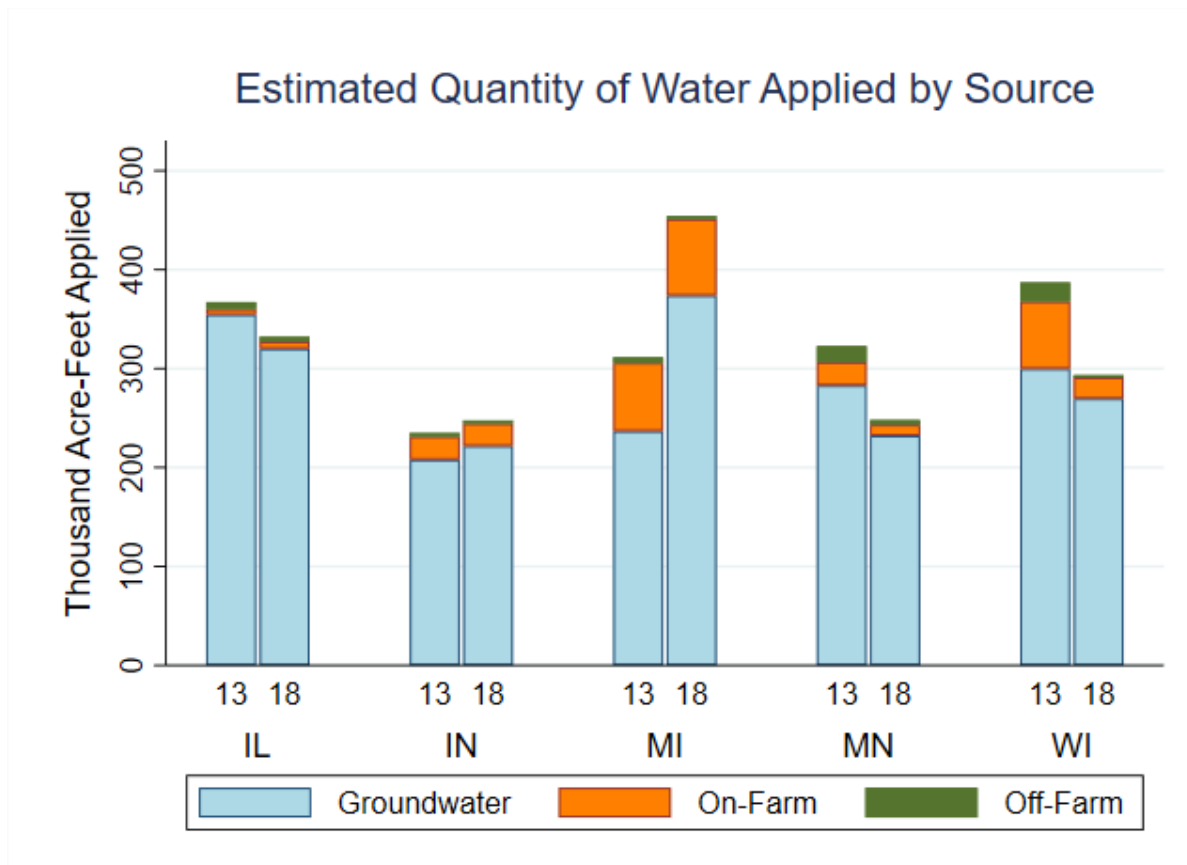

Figure 3A.1 Water Usage by Source in 2013 and 2018.
Note: Data are from 2013 FRIS and 2018 IWMS summary report.

2003

2008

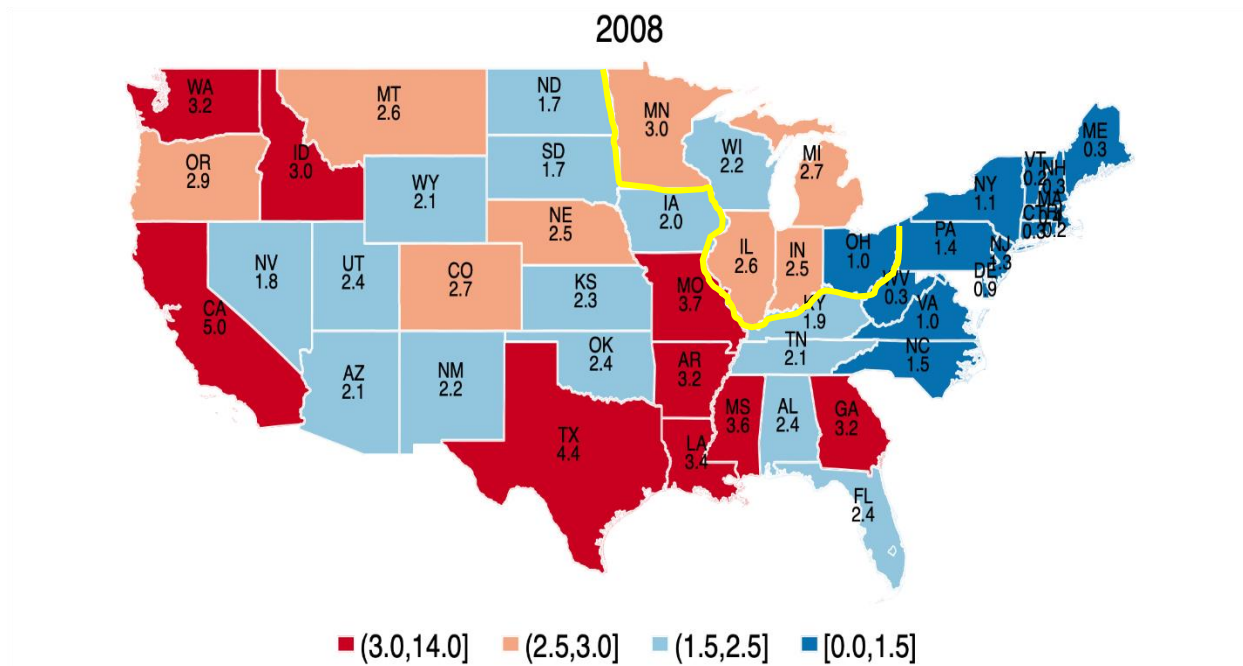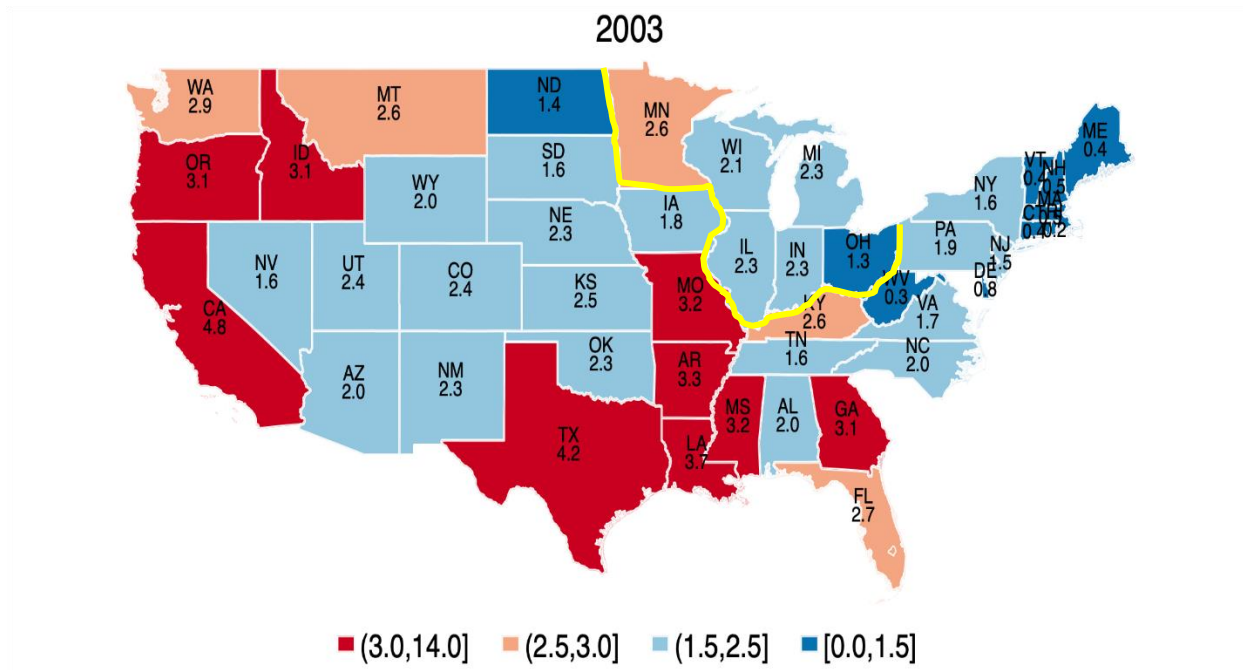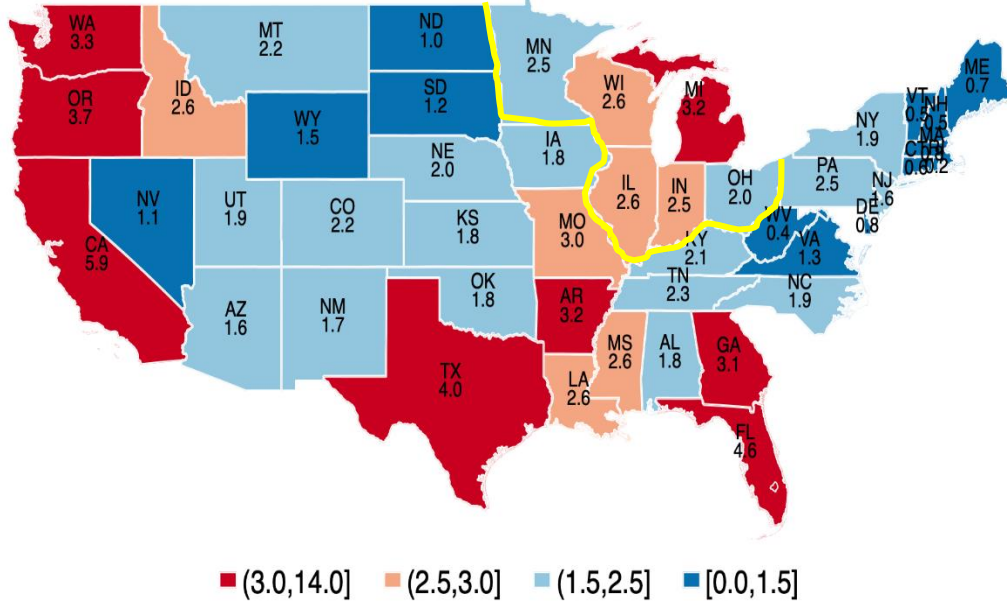(3.0,14.0]   (2.5,3.0]   (1.5,2.5]   [0.0,1.5]

Figure 3A.2 Percentage of State-Level Sample Size over 2003-2018.
Note: The yellow curve indicates the targeted area in this study. For example, "MI 1.6" means farm size of Michigan accounts for around 1.6% in the national farm size (=541/34,783).
Source: USDA FRIS/IMWS.

Figure 3A.2 (cont'd).

2013

Legend: (3.0,14.0] (2.5,3.0] (1.5,2.5] [0.0,1.5]

WA 3.3, OR 3.7, CA 5.9, NV 1.1, ID 2.6, MT 2.2, WY 1.5, UT 1.9, AZ 1.6, CO 2.2, NM 1.7, ND 1.0, SD 1.2, NE 2.0, KS 1.8, OK 1.8, TX 4.0, MN 2.5, IA 1.8, MO 3.0, AR 3.2, LA 2.6, WI 2.6, IL 2.6, MS 2.6, MI 3.2, IN 2.5, KY 2.1, TN 2.3, AL 1.8, GA 3.1, OH 2.0, WV 0.4, VA 1.3, NC 1.9, FL 4.6, PA 2.5, NY 1.9, NJ 1.6, DE 0.8, VT 0.5, NH 1.5, MA 1.2, CT 0.6, RI 0.2, ME 0.7

2018

Legend: (3.0,14.0] (2.5,3.0] (1.5,2.5] [0.0,1.5]

WA 4.7, OR 4.8, CA 13.4, NV 0.8, ID 5.2, MT 2.9, WY 1.8, UT 2.7, AZ 1.6, CO 4.6, NM 2.0, ND 0.3, SD 0.5, NE 6.9, KS 2.2, OK 1.1, TX 6.7, MN 1.2, IA 0.6, MO 2.3, AR 5.3, LA 1.6, WI 1.2, IL 0.9, MS 2.4, MI 1.6, IN 1.0, KY 1.0, TN 0.8, AL 0.7, GA 2.0, OH 1.1, WV 0.3, VA 0.7, NC 1.3, FL 2.7, PA 1.3, NY 1.0, NJ 0.8, DE 0.4, VT 0.2, NH 0.3, MA 0.3, CT 0.3, ME 0.7