Crop Insurance Subsidies, Production Incentives, and Optimal Policy Design: Evidence from the U.S. Midwest

Job Market Paper of Homa Taheri*

Abstract

Agricultural insurance subsidies are a major component of farm policy in the United States. Subsidies encourage farmer participation, facilitate risk management, and stabilize rural incomes. However, farmers often share in conversations that subsidies distort their production decisions and affect crop diversity, and alter sensitivity of yields to extreme weather events. In contrast, existing empirical research on the U.S. Federal Crop Insurance Program (FCIP) generally documents limited effects of subsidies on crop yields. This apparent mismatch between practitioner experience and empirical findings motivates this paper.

In the first part of the paper, I re-examine the production effects of crop insurance subsidies by distinguishing between two margins of enrollment: the external margin, measured by the percentage of planted acreage insured, and the internal margin, measured by insurance liability per insured acre. To identify causal effects, I develop instrumental variables that capture (i) variation in the ease of purchasing insurance, (ii) the historical reliability of weather-related news consumed by local farmers, and (iii) the past performance of farmers. I further account for spatial proximity across counties to capture peer effects in insurance adoption. To examine dynamic responses to policy shocks, I employ the local projection method to trace the medium- and long-run effects of subsidy changes on production outcomes. [Results to be added]

In the second part of the paper, I move from positive analysis to a normative question: what is the welfare-maximizing level of crop insurance subsidies? I address this question using a sufficient statistics framework commonly applied in public finance and international trade. This approach allows me to quantify the trade-off between risk-sharing benefits and production distortions implied by the estimated behavioral responses. [Results to be added]

Together, the results contribute to understanding how agricultural insurance design affects production behavior and inform ongoing policy debates about the appropriate level of federal support for climate risk management in agriculture.

Keywords: Insurance economics, Public Finance, Agricultural Risk, Welfare evaluation

JEL Codes: G22, Q14, H22, H54

*Homa Taheri is a Ph.D. candidate in Economics at Indiana University Bloomington, a researcher at the International Food Policy Research Institute (IFPRI), and a visiting Scholar at the Center for Spatial Data Science (CSDS) at the University of Chicago. Contact: htaheri@iu.edu. I am deeply grateful to Ahmad Lashkaripour for his guidance and support throughout this project. I thank Jisang Yu and Francis Tsiboe for generously sharing replication materials. I also thank the participants at the 89th Annual Meeting of the Midwest Economics Association, the Hoosiers Conference, and my colleagues at CSDS for valuable comments and feedback. Numerous staff members at the U.S. Department of Agriculture, especially at the Risk Management Agency, provided essential data assistance and institutional guidance. This project would not have been possible without their continued support. All errors are my own.

"*Crop Insurance and a system of storage reserves would help to protect the income of individual farmers against the hazards of crop failure; it would help to protect consumers against shortages of food supply and against extremes of prices; and finally, it would assist in providing a more nearly even flow of farm supplies, thus stabilizing farm buying power and contributing to the security of business and employment.*" - Franklin D. Roosevelt, February 18, 1937, Message to Congress on Crop Insurance (Roosevelt 1937).

"*Agriculture ... is our wisest pursuit, because it will in the end contribute most to real wealth, good morals & happiness.* - Thomas Jefferson, Letter to George Washington, August 14, 1787 (Jefferson 1787)".

1 Introduction

Crop insurance transfers climate risk from farmers to the public sector through a heavily subsidized program administered by the U.S. Department of Agriculture (USDA). Under the Federal Crop Insurance Program (FCIP), farmers pay a subsidized premium. In return, the USDA provides indemnity payments when realized yields (or revenue, depending on the policy) fall below guaranteed levels. From the farmer's perspective, the expected net return from insurance depends on two factors: the subsidy rate and the actuarial performance of the policy. Because subsidies change both expected returns and the distribution of risk, they can affect production decisions. In particular, subsidies may influence crop choice, land allocation, and investment in adaptation.

In my conversations with the members of the Chicago Farmers[1], I often hear them sharing their experience with the FCIP: how it discourages growing produce other than the covered produce, and how they know of some farmers who have benefited from the program for financing their risky investment in some form of practice they did not have any prior experience with; examples of untended effects of policies, and moral hazard of the beneficiaries. Back to my office, and in the literature, I have come across a large body of literature that has studied the incidents of moral hazard and adverse selection of farmers from the point of view of use of inputs such as irrigation, fertilizers and pesticides, as well as decisions such as the area planted and harvested, and crop yield. In all cases except the very last one, other scholars have found evidence on negative effect of enrollment in crop insurance on farming practices. When it comes to the effect on crop yield in the US, Roberts, E. J. O'Donoghue, and Key (2007) reports little effect on yield distributions. This contrast between effects on input use and outcomes, raises the central question of my paper.

The first part of this paper addresses this question that if crop insurance alters production incentives, why do we observe limited effects on yields? Are traditional yield-based analyses missing important behavioral responses operating through crop allocation, risk exposure, or adaptation margins? In order to solve the puzzle of yield effects of enrollment in crop insurance, I first divide the task into extensive and intensive margins. At the extensive margin, I study the effect of an extra unit of insured acre per planted acre on the yield outcome of farmers. At the intensive margin, I study the effect of an extra US Dollar liability of farmers to the FCIP per insured acre on the yield outcome of farmers. To identify causal effects, I develop instrumental variables that capture (i) variation in

[1] https://chicagofarmers.org/

the ease of purchasing insurance, (ii) the historical reliability of weather-related news consumed by local farmers, and (iii) the past performance of farmers. I further account for spatial proximity across counties to capture peer effects in insurance adoption. To examine dynamic responses to policy shocks, I employ the local projection method to trace the medium- and long-run effects of subsidy changes on production outcomes. This is the first empirical analysis of long term effects of Farm Bills in literature. [Results to be added]

In the second part of the paper, I move from positive analysis to a normative question: what is the welfare-maximizing level of crop insurance subsidies? I address this question using a sufficient statistics framework commonly applied in public finance and international trade. This approach allows me to quantify the trade-off between risk-sharing benefits and production distortions implied by the estimated behavioral responses. I model the optimal level of subsidies for two cases of no land relocation across crops (Model A), and with relocation of acreage across crops (Model B) and recover the welfare effects of premium subsidies using estimable demand elasticities and fiscal parameters, without fully modeling production or consumption. The closest prior work to my analysis is the Ph.D. dissertation of Jialing Yu (2017, ch. 4), that characterizes the marginal welfare effect consequences and optimal policy design for the FCIP. Jialing Yu (2017) reports that under current policy settings, the net marginal welfare effects of coverage levels and subsidy rates are negative, and both are greater than their optimal levels. [Results to be added]

Prior research on yield effects of the FCIP very extensively focuses on the differential performance of insured and unsured farmers when in some form or another exposed to the consequences climate change. One exception is Roberts, E. J. O'Donoghue, and Key (2007) who studied the yield effects of crop insurance by by comparing how crop yields change for those cycling into insurance relative to those in the same county who were insured in both periods, uninsured in both periods, or who cycled out of insurance. The authors found that both buy-up and CAT insurance have little effect on yield distributions.

Instances of the predominant line of work that focuses on the responsiveness of insured crops to the changing climate, are studies such as Annan and Schlenker (2015) that show the highly subsidized crop insurance program of the USDA makes farmers more sensitive to fluctuations in extreme heat. The authors find that the sensitivity to extreme heat is 67% larger for insured corn than for uninsured corn, and 43% higher for soybeans than for uninsured soybeans. They interpret this finding as evidence that crop insurance subsidies can discourage costly adaptive investments, thereby increasing vulnerability to heat stress. Their results highlight a central tension in agricultural policy. Although crop insurance is designed to mitigate risk, it can also inadvertently reduce resilience to extreme weather. Subsequent work reinforces this concern. Regmi, Tack, and Featherstone (2023) finds that insured dryland corn is approximately 38% more sensitive to extreme heat than uninsured corn. These findings suggest that subsidies on the intensive margin not only reduce incentives to adapt but can also systematically increase the exposure of insured crops to weather-induced yield losses. Other studies emphasize the ways in which crop insurance subsidies shape production behavior and, in turn, exposure to weather risk. Jisang Yu, A. Smith, and Sumner (2018) estimate that a 10% increase in premium subsidies raises insured crop acreage by about 0.43%. By incentivizing greater insured acres, subsidies expand the area of production that is directly exposed to extreme weather events. This suggests that on the extensive margin, subsidies affect not only adaptation decisions but also the spatial scale of exposure to risk.

Some scholars have identified the role of Insurance subsidies on resilience through innovation and the adoption of technology. Chemeris, Liu, and Alan P. Ker (2022) find that while subsidies can encourage innovation in some contexts, they may simultaneously weaken crop yield resiliency under climate change. Their analysis indicates that the relationship between insurance, innovation, and weather risk is complex: subsidies provide stability in the short run but may undermine incentives for longer-term resilience strategies.

In this dissertation I use the case of American farmers for this goal, but implications of this study are by no means limited to the US, since subsidy provision for farmers is a common practice worldwide (see Appendix E) for more information on subsidies in other countries.

This paper contributes to a few lines of research. 1) The effect of premium subsidies on demand for crop insurance is a well-studied question in the literature (E. O'Donoghue 2014; Babcock and Hart 2005). 2) Sensitivity of crop yield to extreme weather in the Midwest. 3) Social welfare effects of crop insurance (Chung 2014), 4) crop insurance as a strategy for adapting to climate change(Falco et al. 2014) and 5) Method for decreasing income risk and income hedging of farmers (Walters and Preston 2018).

The remainder of the paper proceeds as follows. Section 2 describes the institutional background of the FCIP. Section 3 introduces the data. Section 4 discusses the methodology. Sections on empirical framework, results, and conclusion are currently available as an unpolished format on page 48 and onward.

2 Background

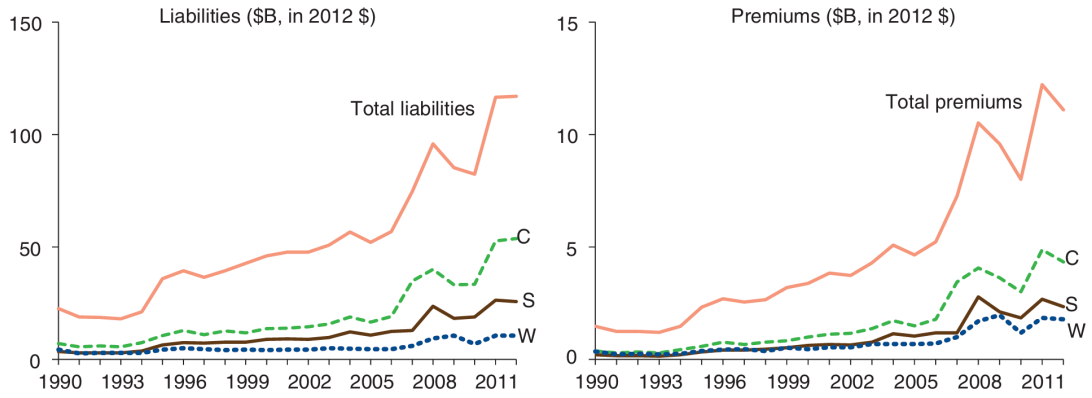
2.1 Institutional Background

Federal Crop Insurance Program (FCIP) is the largest protective program of the government of United States for the farmers. Initiated in 1937, it is also the oldest of the world's public crop insurance schemes. Since 1990s, the Federal crop insurance program has raised to a prominent Government support for U.S. crop farmers. The Federal Crop Insurance was first developed under the Agricultural Adjustment Act of 1938, as a part of the New Deal. Its introduction was a response to the Great Depression and large declines in crop yields resulting from the Dust Bowl. Before the passage of Federal Crop Insurance Act of 1980, only a small percentage of farmers enrolled in the program. But once the 1980 act authorized premium subsidies and included private insurers into the program for marketing purposes, the participation rate increased significantly. The goal of the 1980 Act was to create an insurance program that would replace disaster relief measures while operating on an actuarially sound basis with limited government interference (Barry K Goodwin 1993). Since then, the Federal Crop Insurance has operated through a shared public private arrangement funded by tax payers and producers and offers three types of insurance coverage: yield-based, revenue-based and area-based. In its current design, it is a heavily subsidized program - from 2007 to 2016, in aggregate, Government paid 61% of total premium, and producers paid 39% of total premium (Rosa 2018)- and is one of the largest public insurance programs in the United States. With nearly $15 billion in annual subsidies. Like health, unemployment, and disability insurance, crop insurance embodies the fundamental trade-off in social

insurance design: public subsidies provide risk protection, but they also shape behavior in ways that can amplify or dampen aggregate risks. This paper examines how subsidy design in crop insurance interacts with extreme weather, offering evidence on both the risk-sharing benefits and the incentive distortions of one of the most consequential U.S. social insurance programs.

A major structural shift occurred with the Federal Crop Insurance Act of 1980, which introduced Multiple-Peril Crop Insurance (MPCI) as the core federal product. MPCI expanded coverage to nearly all unavoidable natural causes of yield loss and substantially broadened the set of eligible crops and counties (Vincent H Smith and Glauber 2012). Subsequent reforms—most notably the Crop Insurance Reform Act of 1994—further consolidated MPCI as the dominant risk management instrument by introducing heavily subsidized catastrophic (CAT) coverage and linking access to ad hoc disaster assistance to insurance participation (Glauber 2013). By the mid-1990s, the FCIP had effectively transitioned from a narrowly targeted, named-peril system to a nationwide, multi-peril income stabilization program.

Figure 2
Growth in participation in crop insurance: alternate measures, normalized to 2012 dollars



Note the scale of the y-axes—a factor of 10 difference between liabilities and premiums.
C = corn; S = soybeans; W = wheat. Premiums and liabilities were normalized for these charts using the Consumer Price Index.
Source: USDA, Economic Research Service using data from USDA, Risk Management Agency, Summary of Business files, 1990-2012.

Figure 1: Growth in participation in crop insurance:alternate measures, normalized to 2012 dollars

2.2 The Premiums, Subsidies and Types of Protection

Ain't no one gonna want to read that. Radically cut down on this and integrate in your data section, or put this in an appendix. The US government has played a historically active role in protecting producers against yield and revenue risks by developing, promoting, and subsidizing crop insurance. By reducing the costs of crop yield and crop revenue insurance to farmers, USDA encourages insurance participation (Young, Vandeveer, and Schnepf, 2001). USDA subsidizes insurance premiums based on the riskiness of insured crops at rates that vary by type of insurance, coverage levels, and unit types; although for a given combination of coverage level, insurance product, and unit type, the subsidy rates of all crops are the same Jisang Yu, A. Smith, and Sumner (2018). This makes the premium rates and

subsidy rates that participating farms face endogenous to their production strategies.

Since insurance payouts are a convex function of farm yield, the increase in variability adds to the cost of providing aggregate insurance.

In terms of type of insurance, the two most common products for major field crops are yield and revenue protection. For yield protection, the indemnity is triggered when actual yield is smaller than historical average yield. For revenue protection, the indemnity is triggered when actual yield times harvest price is less than historical average yield times the larger of either projected price or harvest price. Revenue protection generally has a higher premium rate per dollar of liability than yield protection. to be completed: add a history summary of policy changes in USDA using Rosa (2018) and Edwards (2025).

Table 1: Placeholder Caption

Year	Legislation	Major Changes t
1938	Agricultural Adjustment Act of 1938	Established the F
1941	55 Stat. 255	Authorized insur
1944	Agricultural Appropriation Act of 1944	Prohibited FCIC
1944	58 Stat. 918	Reauthorized FC
1947	61 Stat. 718	Restricted FCIC
1980	Federal Crop Insurance Act of 1980	Authorized subsi
1985	Food Security Act of 1985	Introduced conse
1994	Federal Crop Insurance Reform and Department of Agriculture Reauthorization Act of 1994	Implemented revo
1996	Federal Agriculture Improvement and Reform Act of 1996	Removed require

2.3 Stylized Facts

To add: Statistics and map on adoption of different types of insurance coverage for corn and soybeans in different places, to justify the differences in experienced insurance subsidy.

For corn:

	(1)	(2)	(3)	(4)	(5)	(6)
coverageLevel	0.491	0.346	0.440	0.672	0.674	0.678
	(0.277)	(0.292)	(0.311)	(0.113)	(0.115)	(0.118)
policiesSoldCount	41.55	37.26	22.54	19.38	20.19	22.80
	(99.41)	(84.88)	(48.75)	(36.83)	(45.56)	(58.14)
policiesEarningPremiumCount	36.55	30.74	18.14	14.42	14.39	15.25
	(99.74)	(84.59)	(48.09)	(30.15)	(37.06)	(45.82)
policiesIndemnifiedCount	7.500	4.755	2.762	3.939	5.116	3.166
	(29.12)	(22.00)	(10.09)	(11.70)	(20.28)	(12.62)
unitsEarningPremium	76.45	57.65	40.44	36.56	28.74	27.99
	(213.9)	(179.1)	(112.4)	(85.46)	(78.55)	(87.26)
unitsIndemnifiedCount	12.02	7.779	4.897	7.867	8.673	4.751
	(52.94)	(39.88)	(20.51)	(26.66)	(36.35)	(19.48)
netReportedQuantity	4025.5	3589.1	2572.1	2541.1	2987.6	3227.0
	(11445.2)	(10183.3)	(6825.9)	(6081.7)	(8709.4)	(10570.1)
endorsedCompanionAcres	0	0	0	0	0	17.06
	(0)	(0)	(0)	(0)	(0)	(238.2)
liabilityAmount_1989	587641.7	361465.6	325086.2	469877.8	932008.1	827882.5
	(1901243.7)	(1337407.4)	(969280.0)	(1540095.4)	(3220053.1)	(3087080.7)
N	35582	29608	80975	198749	156548	97988

mean coefficients; sd in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

	(1)	(2)	(3)	(4)	(5)	(6)
liabilityAmount_1989	587641.7	361465.6	325086.2	469877.8	932008.1	827882.5
	(1901243.7)	(1337407.4)	(969280.0)	(1540095.4)	(3220053.1)	(3087080.7)
totalPremiumAmount_1989	31469.3	20996.0	21418.9	43600.8	82345.0	71745.1
	(89301.1)	(70699.1)	(60181.6)	(151968.5)	(271121.1)	(258985.2)
subsidyAmount_1989	8023.8	11373.5	7649.0	24427.3	50072.2	44373.1
	(22820.3)	(31465.9)	(21383.3)	(84324.4)	(165571.3)	(164829.4)
statePrivateSubsidy_1989	0	0	0	5.806	0	2.169
	(0)	(0)	(0)	(221.0)	(0)	(75.37)
additionalSubsidy_1989	0	0	0	14.19	0	0
	(0)	(0)	(0)	(335.9)	(0)	(0)
EFApremiumDiscount_1989	0	0	2343.8	0	0	0
	(0)	(0)	(9550.5)	(0)	(0)	(0)
indemnityAmount_1989	33152.7	15323.2	11539.1	28142.9	95122.3	27212.1
	(201541.3)	(89097.5)	(64082.7)	(136409.1)	(573611.8)	(159469.3)
premium_per_acre_1989	8.581	7.684	9.646	15.33	22.40	17.32
	(5.230)	(5.634)	(6.231)	(11.84)	(14.94)	(26.75)
premium_per_liability_1989	0.0800	0.0834	0.0676	0.0703	0.0572	0.0513
	(0.0545)	(0.0668)	(0.0508)	(0.0527)	(0.0467)	(0.0456)
subsidy_per_acre_1989	2.039	3.804	3.190	8.629	13.63	10.67
	(1.023)	(1.619)	(1.472)	(5.994)	(8.593)	(12.83)
subsidy_per_liability_1989	0.0199	0.0512	0.0280	0.0418	0.0363	0.0328
	(0.0129)	(0.0465)	(0.0267)	(0.0315)	(0.0305)	(0.0299)
indemnity_per_acre_1989	10.43	6.462	7.310	11.99	23.12	9.587
	(19.93)	(14.62)	(17.10)	(24.39)	(49.46)	(26.65)
indemnity_per_liability_1989	0.102	0.0650	0.0549	0.0595	0.0545	0.0292
	(0.194)	(0.144)	(0.133)	(0.123)	(0.112)	(0.0764)
N	35582	29608	80975	198749	156548	97988

mean coefficients; sd in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

For soybeans:

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
coverageLevel	0.494	0.343	0.444	0.678	0.676	0.680	0.701
	(0.280)	(0.295)	(0.311)	(0.113)	(0.116)	(0.118)	(0.126)
policiesSoldCount	38.69	40.94	23.57	20.19	20.85	23.84	22.14
	(86.93)	(91.61)	(50.58)	(38.24)	(46.05)	(58.70)	(57.37)
policiesEarningPremiumCount	32.69	33.07	19.03	14.97	14.65	16.46	14.93
	(86.92)	(90.56)	(49.63)	(30.43)	(35.90)	(45.61)	(43.54)
policiesIndemnifiedCount	6.449	3.481	3.295	4.317	3.988	3.594	4.280
	(23.59)	(15.52)	(10.68)	(13.03)	(13.68)	(12.15)	(15.90)
unitsEarningPremium	69.58	63.62	42.96	38.13	29.76	31.20	29.44
	(186.8)	(189.6)	(113.0)	(84.75)	(76.51)	(86.59)	(85.18)
unitsIndemnifiedCount	10.26	5.459	6.025	8.889	6.604	5.720	6.924
	(42.88)	(26.75)	(21.76)	(30.86)	(24.53)	(20.10)	(29.46)
netReportedQuantity	3136.9	3725.8	2743.0	2539.6	2796.0	3457.1	3161.7
	(8781.9)	(10450.9)	(7367.4)	(6287.7)	(7925.0)	(10693.4)	(10415.5)
endorsedCompanionAcres	0	0	0	0	0	14.76	237.9
	(0)	(0)	(0)	(0)	(0)	(194.1)	(1643.1)
N	30267	24326	70702	180334	142595	88253	96700

mean coefficients; sd in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

	(1)	(2)	(3)	(4)	(5)	(6)
liabilityAmount_1989	353402.3	280457.6	239544.6	296427.4	540008.3	595843.7
	(1091944.1)	(963936.3)	(638826.3)	(838080.2)	(1650699.7)	(1946333.2)
totalPremiumAmount_1989	19285.1	15594.2	15038.4	27965.0	51278.4	50223.1
	(51560.7)	(50303.7)	(39900.2)	(92900.9)	(160152.2)	(175173.2)
subsidyAmount_1989	4873.3	9261.8	6243.0	15823.3	31484.7	31559.9
	(13604.1)	(28350.5)	(20057.0)	(52645.2)	(99580.4)	(114453.8)
statePrivateSubsidy_1989	0	0	0	3.774	0	1.862
	(0)	(0)	(0)	(79.94)	(0)	(51.46)
additionalSubsidy_1989	0	0	0	6.154	0	0
	(0)	(0)	(0)	(164.4)	(0)	(0)
EFApremiumDiscount_1989	0	0	1522.9	0	0	0
	(0)	(0)	(5626.1)	(0)	(0)	(0)
indemnityAmount_1989	18890.9	8330.0	10162.4	22162.8	28105.0	20389.9
	(99597.7)	(49187.9)	(42463.4)	(115571.6)	(144952.6)	(96143.0)
premium_per_acre_1989	7.024	5.849	6.901	11.08	15.99	11.77
	(4.541)	(4.390)	(4.459)	(9.318)	(11.02)	(53.14)
premium_per_liability_1989	0.0867	0.0809	0.0648	0.0714	0.0593	0.0471
	(0.0765)	(0.0729)	(0.0533)	(0.0572)	(0.0498)	(0.0428)
subsidy_per_acre_1989	1.651	2.909	2.358	6.176	9.620	7.181
	(0.909)	(1.418)	(1.201)	(4.744)	(6.104)	(23.60)
subsidy_per_liability_1989	0.0213	0.0508	0.0275	0.0421	0.0373	0.0299
	(0.0177)	(0.0524)	(0.0299)	(0.0342)	(0.0318)	(0.0274)
indemnity_per_acre_1989	6.955	4.082	6.127	9.339	9.562	6.802
	(13.39)	(9.823)	(12.80)	(18.90)	(20.91)	(16.41)
indemnity_per_liability_1989	0.0830	0.0492	0.0612	0.0591	0.0373	0.0305
	(0.162)	(0.121)	(0.132)	(0.117)	(0.0863)	(0.0790)
N	30267	24326	70702	180334	142595	88253

mean coefficients; sd in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

During the 1989-2013 period, the weather has been relatively more conducive for producing soybeans, as captured by number of moderate days and extremely hot days (defined in Section 4) in Figures 2 and 3.

1) Soybeans replenish soil nitrogen, and that is typically the reason why they are grown along with (or in rotation with) corn. We discuss this in our forthcoming JHR article (Bellemare, Liu, and Hadrich

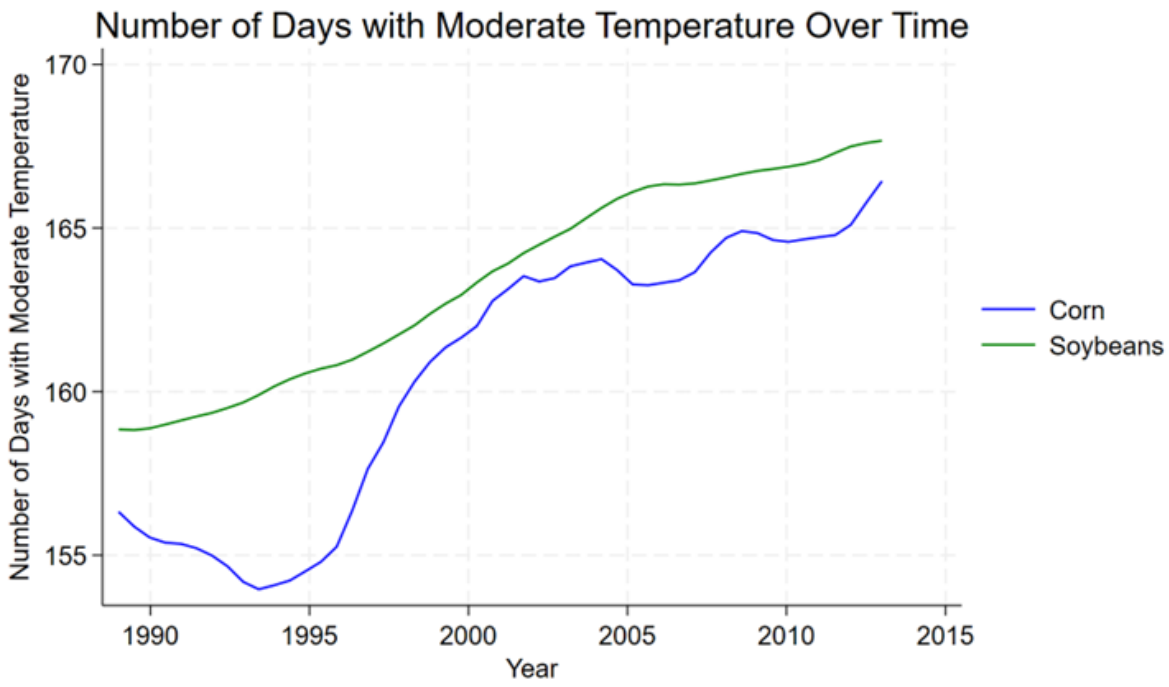


Figure 2: Number of days with moderate temperature (1989-2013)

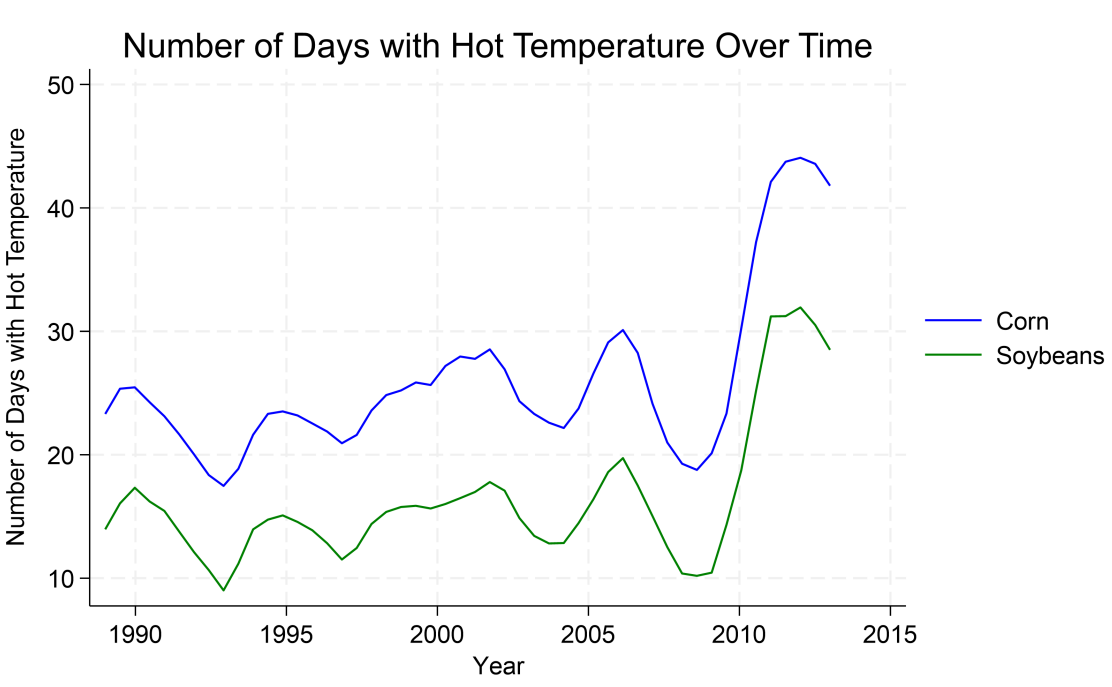


Figure 3: Number of days with extreme temperature (1989-2013)

2025) 2) Footnotes go after punctuation. Shifting from weather variables to production behavior of farmers, we know that a rule of thumb for farmers to decide on whether to grow corn or soybeans is to check the relative size of soybeans price to corn price and grow soybeans when this ratio is higher than 2.5 LaPorte (2025). Whereas Figure 3 that although in early 1990s we expected the farmers to have higher relative soybeans-to-corn acreage, they did not follow their rule of planting decisions. Figure 4 has a more detailed picture of relative prices of two crops over our study period.

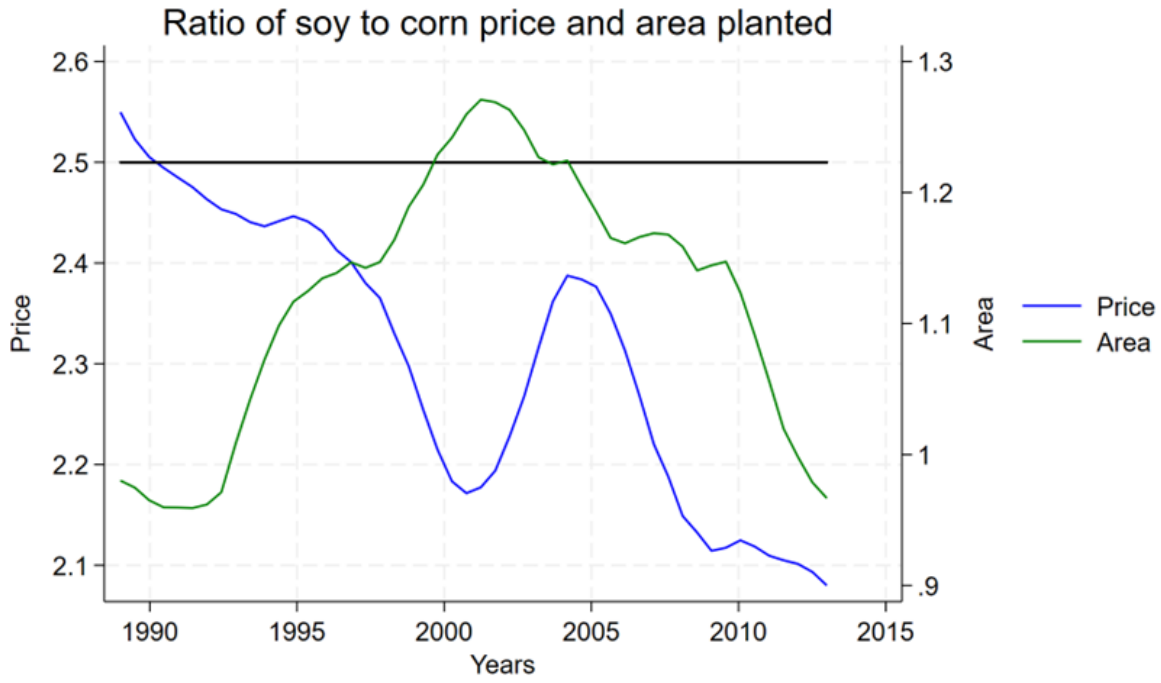


Figure 4: Left axis: Received price of farmers (blue) for soybeans relative to corn. Right axis: Relative area planted for soybeans relative to corn (green). The black horizontal line shows the rule of thumb threshold of switching from corn to soybeans, as known in the farming community.

Another observation is that corn yield has been higher in levels and variability compared to soybeans, suggesting that corn is a higher risk-higher return crop compared to soybeans.

Type of Insurance Enrollment To do: Have graphs of same type to have same color codes, and decrease the empty gap between figures and their caption. I need to clean them in paint.

Education Level of Study Area

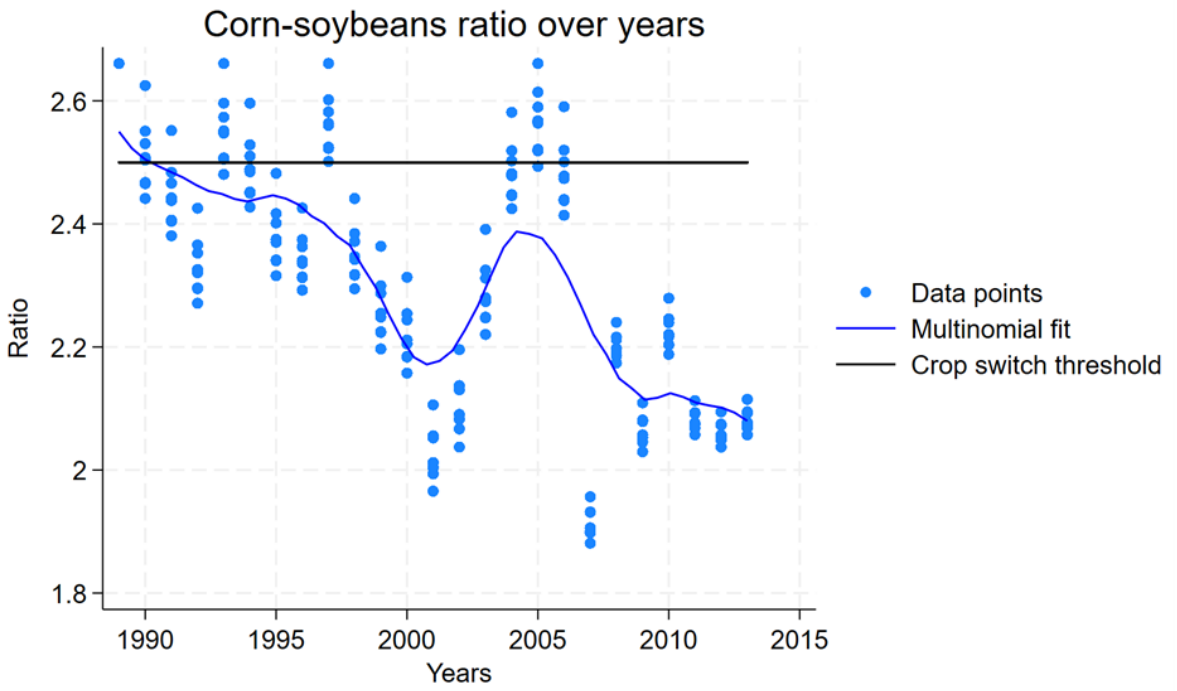


Figure 5: Relative price of soybeans to corn in eight Midwest states and the crop switching threshold.

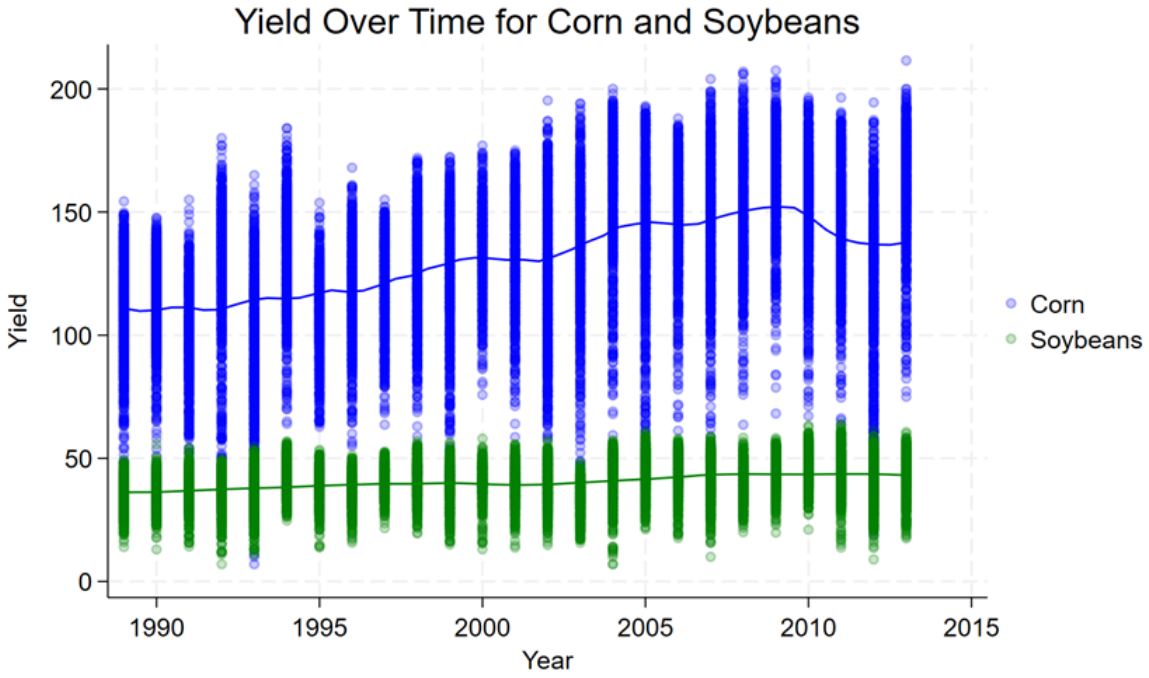


Figure 6: Corn yield is higher and has higher variance compared to soybeans.

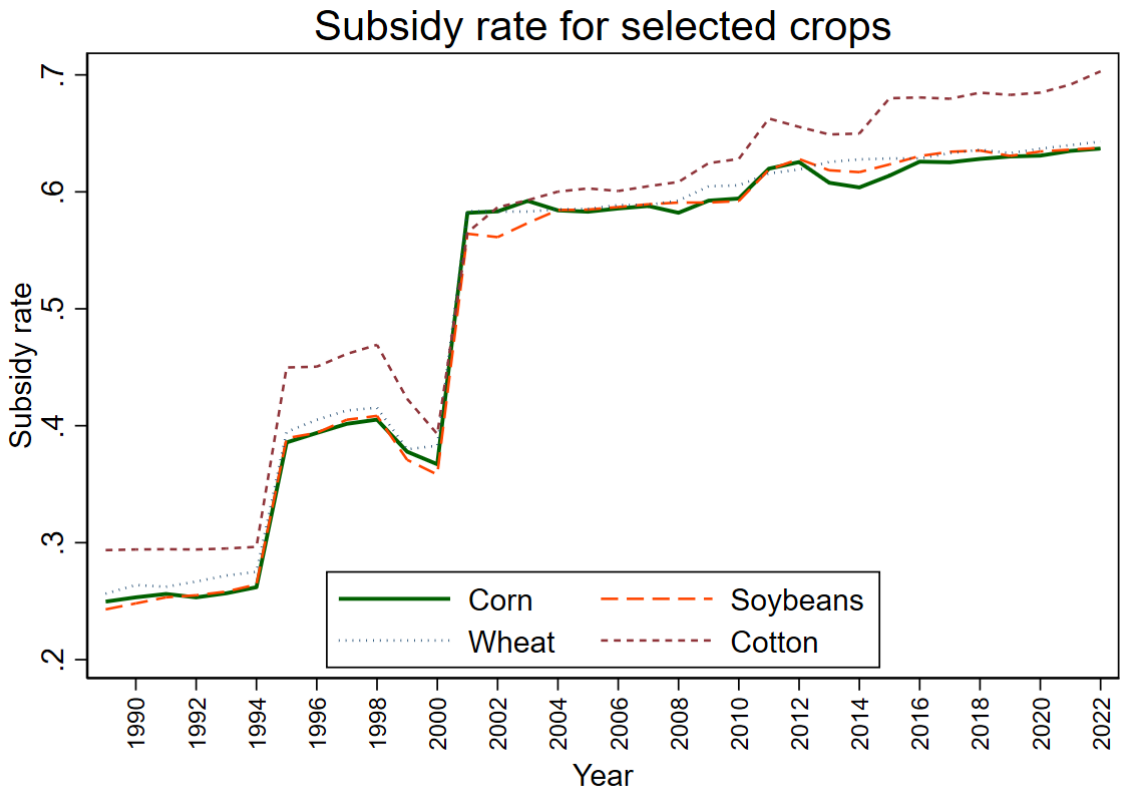
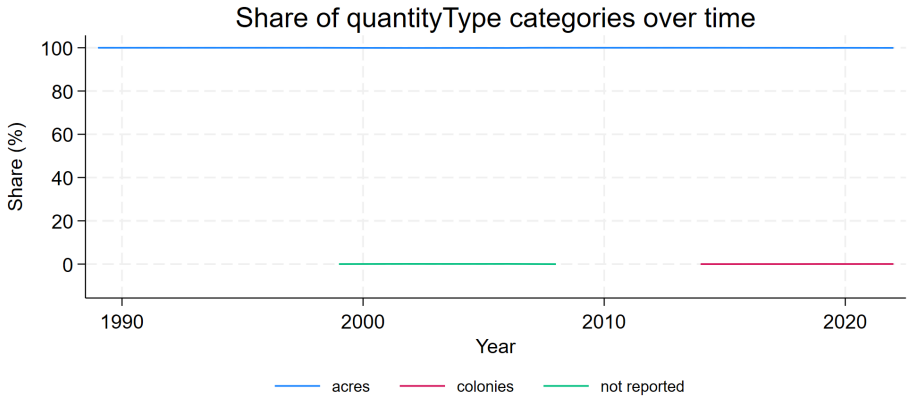
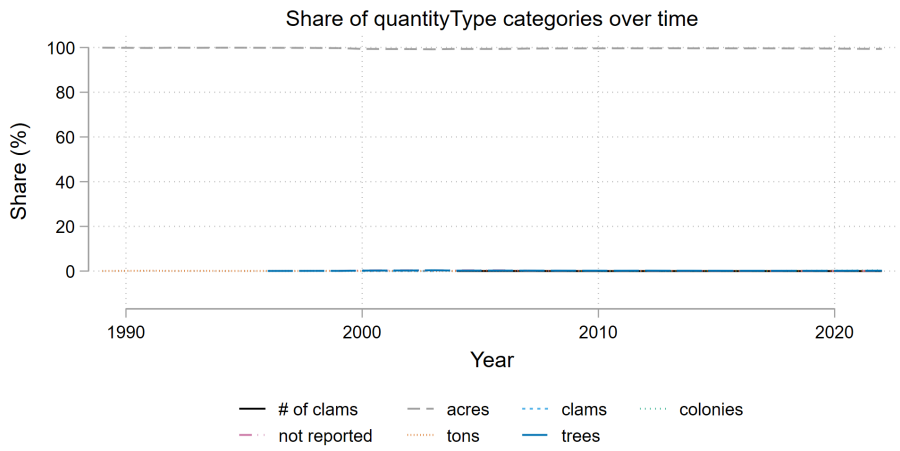


Figure 7: Premium subsidies in yield based on revenue based insurance in all States for corn, soybeans, wheat and cotton

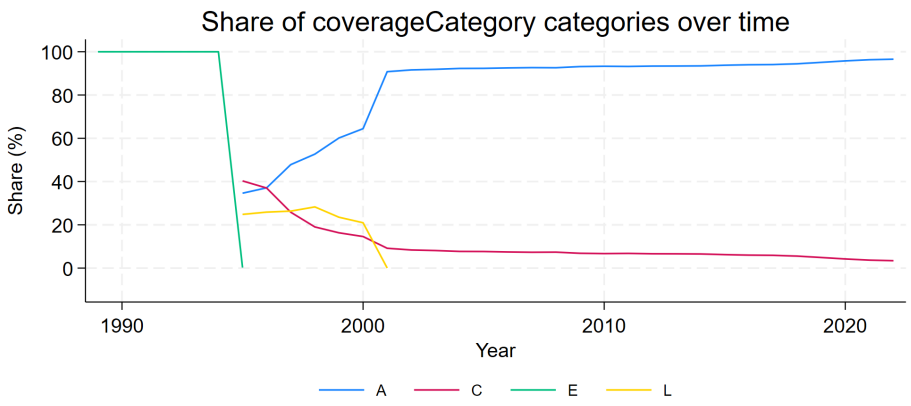


(a) Midwest

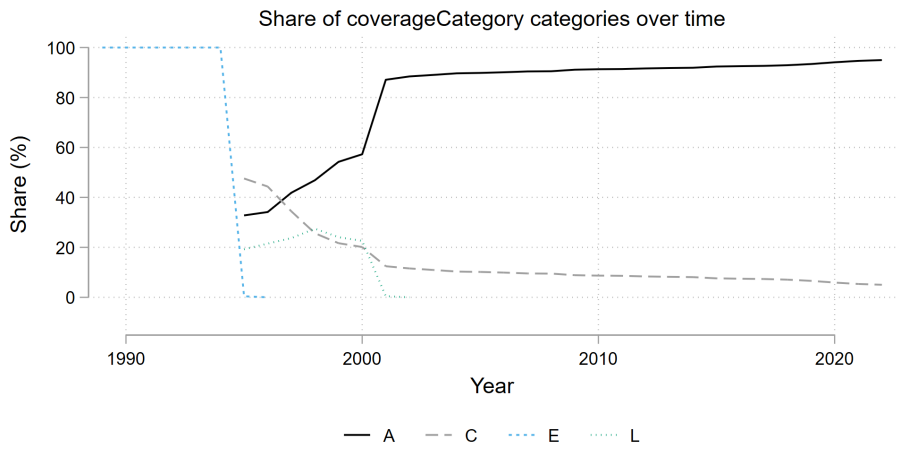


(b) Contingent United States

Figure 8: Share of quantity type categories over time. Data is from U.S. Department of Agriculture, Risk Management Agency (RMA) (2025).

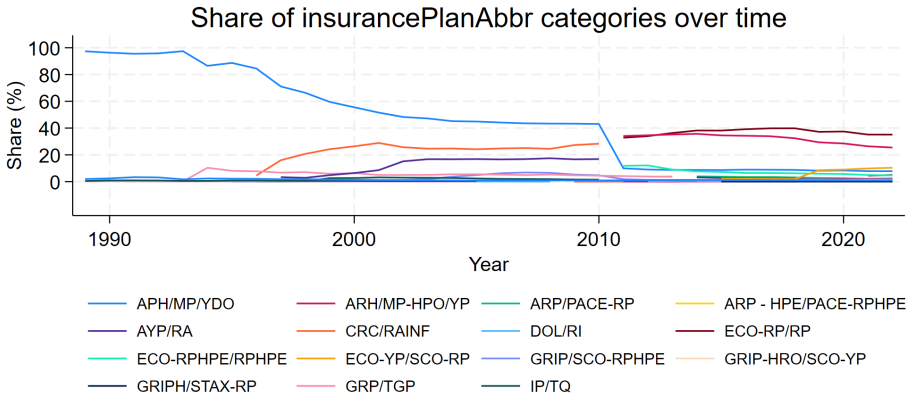


(a) Midwest

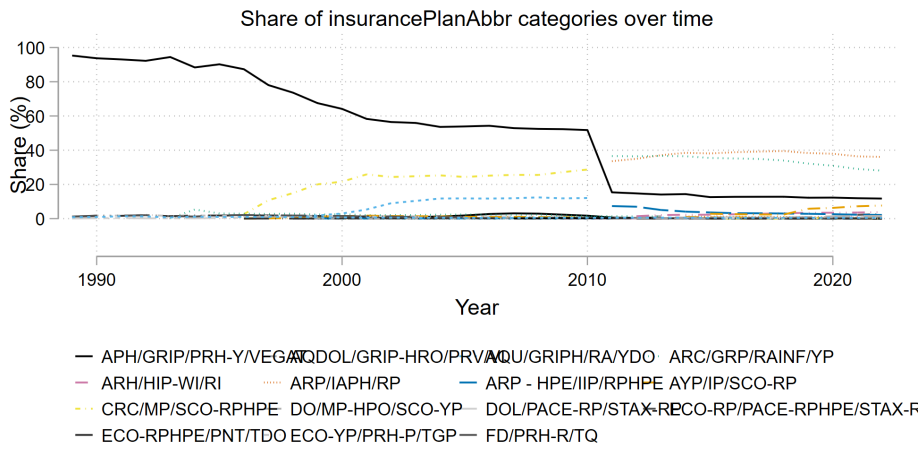


(b) Contingent United States

Figure 9: Share of quantity type categories over time. Data is from U.S. Department of Agriculture, Risk Management Agency (RMA) (2025).

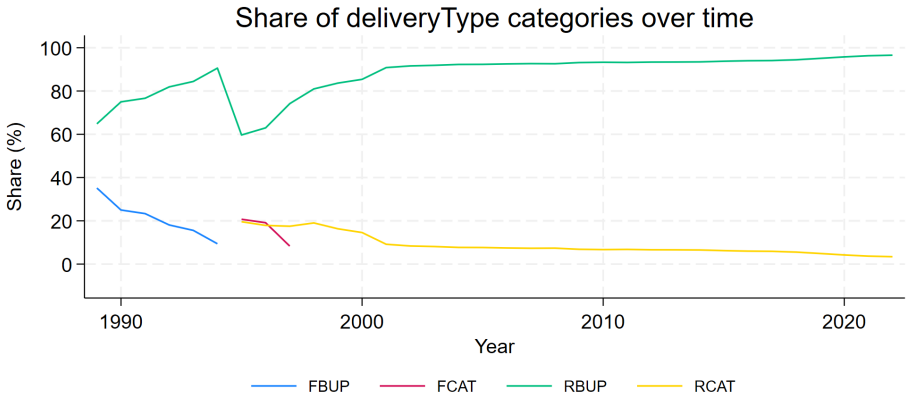


(a) Midwest

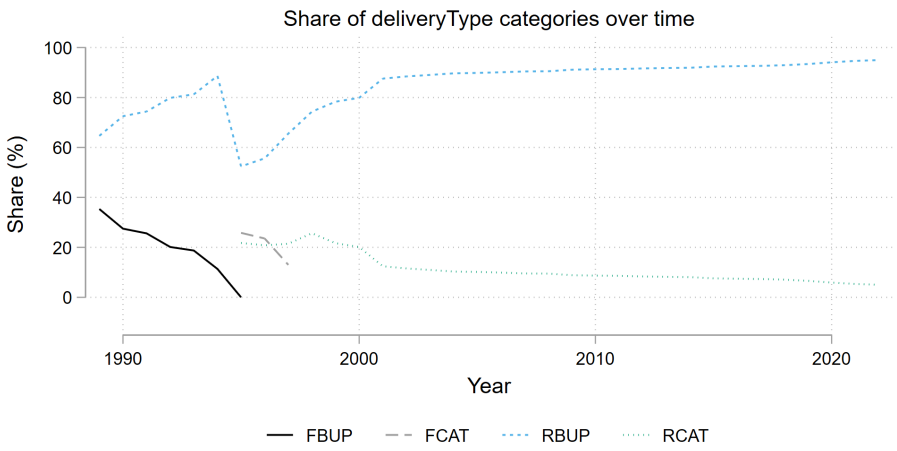


(b) Contingent United States

Figure 10: Share of insurancePlanAbbr categories over time. Data is from U.S. Department of Agriculture, Risk Management Agency (RMA) (2025).

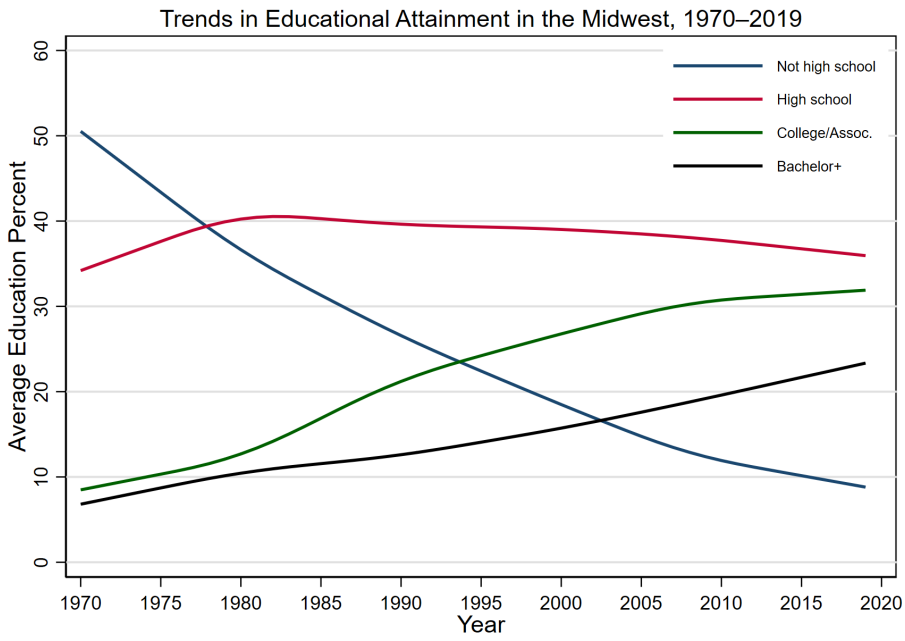


(a) Midwest

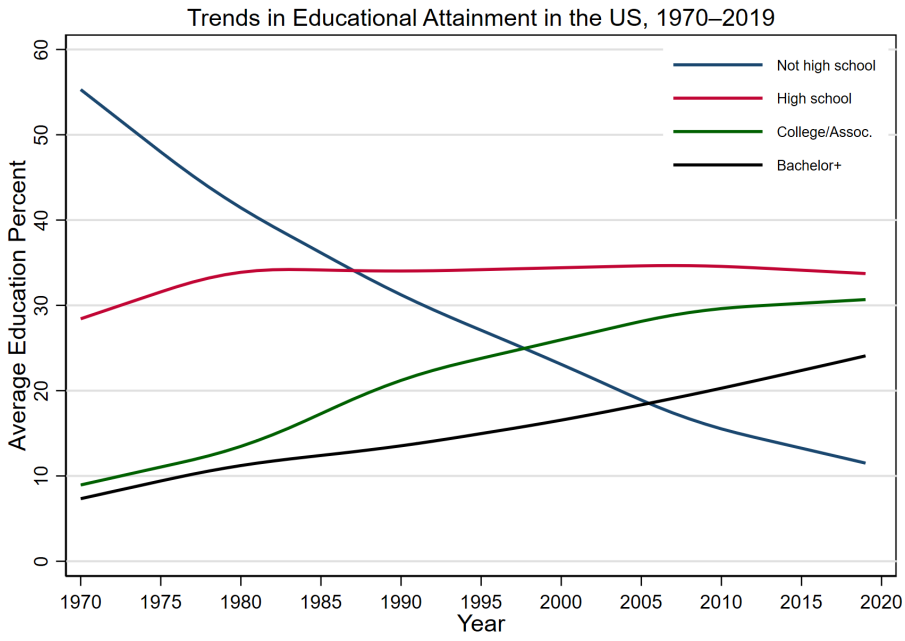


(b) Contingent United States

Figure 11: Share of deliveryType categories over time. Data is from U.S. Department of Agriculture, Risk Management Agency (RMA) (2025).



(a) Midwest



(b) Contingent United States

Figure 12: Percent of adults age 25 and older at different levels of education in the Midwest. Data is from USDA Economic Research Service (2025).

3 Data

Demand for Insurance I follow Barry K Goodwin (1993, p. 427) read on Jan 21, 2026 and use two measures for capturing the demand for crop insurance: ratio of insured to planted acres, and liability per planted acre. Liability is the total value of indemnities that the insurer would have to pay in the event of a complete loss. It equals the product of insured acres, price election, yield guarantee and the FCIC's estimated insurable yield (Barry K Goodwin 1993, p. 427). I call the first measure the external margin, and the second measure, the internal margin of insurance enrollment.

I used databases developed by Annan and Schlenker (2015), Jisang Yu, A. Smith, and Sumner (2018), Tsiboe and Turner (2023a) and Tsiboe and Turner (2023b) in this research. I limited the scope of paper to barley, corn, cotton, oats, rice, sorghum, soybeans, and wheat, in eight Midwest states of Illinois, Indiana, Iowa, Michigan, Minnesota, Missouri, Ohio, and Wisconsin, over the years 1989 to 2013. This choice of geography and crops was due to the fact that these states are major producers of a wide range of food and animal feed for national consumption and international trade. We also focused on corn and soybeans, as these two main commodity crops are grown on 75% of the arable land in the region (Angel et al. 2018).

Annan and Schlenker (2015) combines the publicly available data on county-level insurance coverage (RMA) with data on county-level acres planted and yields for corn and soybeans (USDA National Agricultural Statistical Service). The two data sources derive county aggregates in different ways: the RMA data is the sum of all actual contracts for a given year in the county. The NASS data on the other hand, uses a representative subsample of all farms in the June Agricultural Survey of each year to approximate the planting area in a county and so has sampling noise. As a result, the fraction insured, the ratio of the insured planted area to the total planting area can be larger than one if the NASS estimate in the denominator is an underestimate of the actual planting area. The fraction of land insured and weather variables are from this data set. The weather variables capture moderate heat (degree days between 10-29 degrees Celsius for corn, and 10-30 degrees Celsius for soybeans), extreme heat (degree days above 29 degrees Celsius for corn and above 30 degrees Celsius for soybeans), amount of precipitation in centimeters, and its squared value, during the months of April to September. All parameters are annual and specific to counties.

Jisang Yu, A. Smith, and Sumner (2018) uses the annual county-level information on crop acreage and crop insurance characteristics for major field crops from surveys of the NASS and the RMA's Summary of Business. For price data, the authors have used two prices: 1) futures price data as reported by the Commodity Research Bureau and then deflated each price by the Producer Price Index from the Bureau of Labor Statistics using May 1986 as the base year (U.S. Department of Labor, Bureau of Labor Statistics 2016). 2) Received price, which is the State-level average received price by farms for each crop from NASS. The authors combined the information from both sources and constructed a price variable called "expected price" using both futures prices and average prices received by farms[2]. The county-crop level subsidy per dollar of liability and the expected price of crops are from this data set.

[2]To do so, they regressed the state level annual prices received by farms on state dummies and futures prices for 1989 2014. They used the predicted values from this regression as the expected prices.

Tsiboe and Turner (2023a) has used the summary of business and the cause of loss data sets of the RMA. The authors constructed a county-crop panel dataset from 1948 to 2020 that includes FCIP loss experience information on liability insured by producers, total premium, premium subsidy, and indemnity paid to farmers. The instrumental variable is from this data set.

2 shows the summary statistics of the variables used in this study. There are observations without a specific county, recorded as "other" in the data set. Since they are not tractable through years, we dropped those observations. It made the panel data set unbalanced.

Table 2: Means and Standard Deviations of County–Crop Panels

Variables	**1989–1993**	**2010–2013**	**1989-2013**
Yield	74.35	92.68	86.34
	(41.40)	(54.66)	(50.39)
Planted Acreage	64,801.05	77,085.56	71,705.23
	(54,860.52)	(58,295.91)	(56,463.10)
Insured Acreage	21,388.70	64,330.86	46,947.02
	(29,829.91)	(52,459.90)	(47,008.81)
Fraction Insured (%)	26.94	80.15	59.99
	(20.13)	(13.70)	(26.61)
Subsidy per Liability	0.02	0.06	0.05
	(0.01)	(0.03)	(0.03)
Hot Degree Days	17.47	33.75	21.09
	(18.58)	(30.79)	(21.72)
Moderate Degree Days	156.52	170.50	162.02
	(29.47)	(27.28)	(27.81)
Precipitation	0.62	0.60	0.60
	(0.17)	(0.16)	(0.14)
Observations	6,568	5,007	32,289

Notes: Means shown; standard deviations in parentheses. Periods correspond to the early and late years in the study period. Variables are county–crop level.

NOAA Weather Radio All Hazards (NWR) I downloaded the list of county-level weather stations from the NOAA Weather Radio All Hazards (National Weather Service 2025), and created a measure of number of stations per county. I am deciding weather to use this measure as an IV for the intensive margin of enrollment in the Federal Crop Insurance program, or as an exogenous variable called "weather information access".

Weather station locations and activity periods I obtained the weather station locations and activity periods from NOAA's Global Historical Climatology Network–Daily (GHCN-Daily) dataset (Menne, Durre, Korzeniewski, et al. 2012; Menne, Durre, Vose, et al. 2012; NOAA National Centers for Environmental Information 2025c). I retrieved the Station coordinates from the GHCN-Daily station metadata file (NOAA National Centers for Environmental Information 2025b), and station-year availability from the GHCN-Daily inventory file (NOAA National Centers for Environmental Information 2025a). I then defined the density of weather observation stations as the "number of stations per county" in baseline year of 1989 based on this collected data sets.

NOAA National Centers for Environmental Information (2025b) reports 129,658 weather stations, out of which 75,847 (58.50 percent of total in the world) are based in the United States. Out of

these, 14,217 weather stations (10.97 percent of total in the US) are located in 8 midwest States of this current research. Limiting the time scope to the 1989 afterwards, I find 12,623 of the weather stations (88.79 percent of midwest weather stations) have been active sometime in years 1989 onwards. Figure ?? shows the change in the number of active weather stations over the 1980-2025 period in the 8 midwest states.

Actuarial Data I downloaded the actuarial data of the USDA insurance program from the "Insurance Offer Rate Reports" of U.S. Department of Agriculture, Risk Management Agency (2025). This website reports multiple data: the sub-county rate reports, have data on the base premium rate and rate method code (A, E, C, R and T for actuarial experience, experience blended, county rate, regional rate and transitional rate) by state of irrigation and organic production.

Census Block level boundary and population data To do: to change the numbers to 2000 and refer to the correct citations. To construct population-weighted county centroids for the 1990 period, I use block-level population and census boundary data from the 1990 U.S. Census, obtained through the IPUMS National Historical Geographic Information System (NHGIS) (Schroeder et al. 2025b). This dataset provides the Census block along with total population counts from the 1990 Summary Tape File 1 (STF1). The NHGIS delivery includes both the GIS boundary file containing the full set of 1990 block polygons and a corresponding attribute file with block-level population (P001: Total Population), linked through the unique GISJOIN identifier. These data allow for precise spatial weighting of population within each county, greatly improving the accuracy of location-based measures such as travel distance to USDA offices.

$$\text{Population-weighted centroid} = (\bar{x}, \bar{y}) = \left(\frac{\sum_{i=1}^{n} x_i p_i}{\sum_{i=1}^{n} p_i}, \frac{\sum_{i=1}^{n} y_i p_i}{\sum_{i=1}^{n} p_i} \right)$$

where:

- (x_i, y_i) denotes the geographic centroid of census block i,
- p_i is the population of census block i,
- n is the total number of census blocks in the county.

Bulk Agricultural Census and Survey Datasets (Ag census) The USDA National Agricultural Statistics Service (NASS) provides County-, State- and National-level data on both agricultural surveys and quinquennial agricultural censuses for years 1866 onward(United States Department of Agriculture, National Agricultural Statistics Service 2025). This data reports statistics at annual, monthly, seasonal and occasional levels. I downloaded five data sets on crops, demographics, environmental, economics and animals, retained the county-level annual data for years 1989 onward, and appended these data files for further analysis. In census years, this data set reports two lines for multiple measures (i.e. area planted and area harvested), one for census and one for survey. To avoid multiple counting, I use the census reported values by setting the corresponding values of survey data equal to zero (source description = *survey*). I started off with analysis for corn and soybeans, and due to data availability, I expanded the work to wheat, cotton, forage production, corn, sorghum, peanuts, soybeans, and barley (since Jisang Yu, A. Smith, and Sumner (2018) reported values for insurance liability and crop prices for these crops). I adopted variables on diversification, area planted, area harvested, irrigation, income, etc from this data set to be completed.

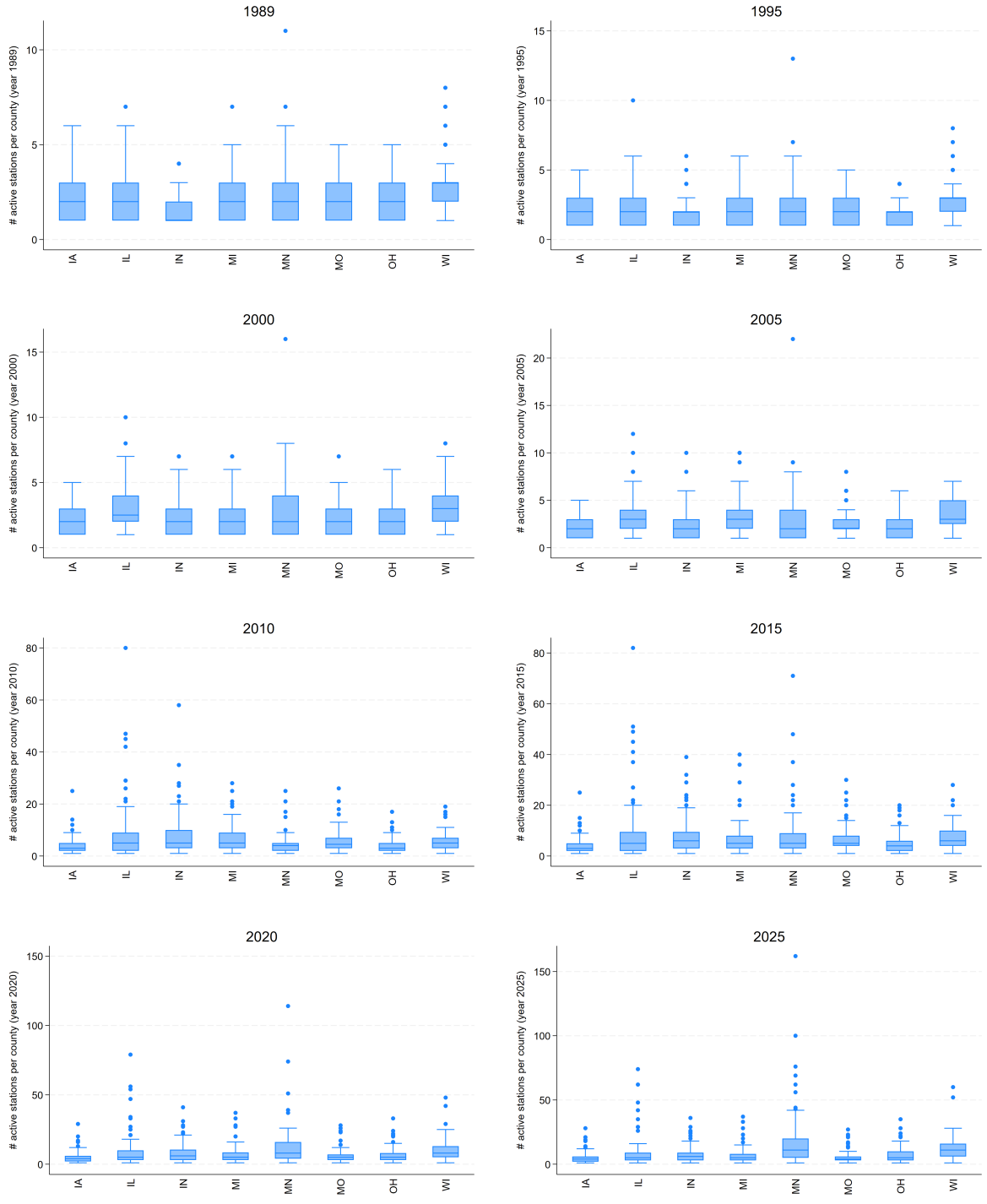


Figure 13: Distribution of the number of weather stations per county by state (1989–2025).

I controlled for diversification and savings, as Wright and Hewitt (1994) suggest that the perceived demand for agricultural insurance may be overstated, because farmers can use to cushion the impact of production shortfalls on consumption. to bring; data on saving of farmers I used the following variables from the census in this analysis 4 and 5.

Operators' Work Experience from Ag census: I construct county-year measures of operator experience by computing the share of operators with at least ten years of experience, separately for experience on any operation and on the present operation. Since operators' experience applies to all crop they grow, I defined this measure ay the county-year level, and not at he county-crop-year level.

Income Measures from Ag census. I construct two county–year measures of farm income. The primary measure is net cash farm income, defined as cash receipts minus cash expenses, which captures realized cash flow but can take zero or negative values in some county–year observations. To retain these observations and limit the influence of extreme values, I transform net cash farm income using the inverse hyperbolic sine (IHS), which is well defined for zero and negative values and approximates the logarithmic transformation for large positive values (Bellemare and Wichman 2020). As a complementary measure, I use farm-related income, which is strictly positive and reflects gross income from farm activities; this variable is transformed using the natural logarithm. Net cash farm income serves as the baseline income measure, while logged farm-related income is used in robustness analyses.

Access to Marketing Assistance Loans from Ag census. I construct county–year measures of Commodity Credit Corporation (CCC) marketing assistance loan exposure, defining net CCC loan position as loans received minus loans repaid, which proxies for county-level access to short-term agricultural credit. Net CCC loan receipts are transformed using the inverse hyperbolic sine to accommodate zero values and reduce right skewness.

Fertilizer Use from Ag census. I construct a county–year measure of fertilizer use by dividing total fertilizer expenditures by total planted acreage, yielding fertilizer expenditure per acre planted. This normalization accounts for differences in county size and serves as a proxy for input intensity. I apply the inverse hyperbolic sine transformation to reduce skewness and retain zero values.

Female Producers from Ag census. I construct a county-year measure of female representation among farmers. I aggregate the number of female and male producers to the county-year level and define the share of female producers as the number of female producers divided by the total number of producers in the county-year. The Census reports data on number of operators and principals as well, that I ignored in my calculations.

Minority Producers from Ag census. I construct a county–year measure of racial and ethnic composition among farmers using producer counts from the Census of Agriculture. I aggregate the number of producers identifying as Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, multi-race, or Hispanic and define the share of minority producers as this total divided by the total number of producers in the county-year. This variable captures persistent cross-county differences in farmer demographics and is used in levels.

Cattle Intensity from Ag census. I control for livestock intensity using county–year cattle inventory per planted acre, which captures the prevalence of mixed crop–livestock farming systems that may influence both insurance participation and crop outcomes.

Land Tenuring from Ag census. I construct a county-year measure of land tenure using data on acres operated by ownership status. To do so, I compute the share of operated farmland that is owned by operators as the ratio of acres owned to the total acres operated (owned plus rented from others). This variable captures persistent differences in land tenure and investment incentives across

counties and is merged into the county–crop–year panel as a time-varying county-level control.

Table 3: County-Level Controls Available Only in Census Years

Variable	Source	Level	Years
Area planted	Ag census/survey	County-crop-year	1989-2024
Area harvested	Ag census/survey	County-crop-year	1989-2024
Planted area irrigated	Ag census/survey	County-crop-year	1989-2024
Planted area non-irrigated	Ag census/survey	County-crop-year	1989-2024
Harvested area irrigated	Ag census/survey	County-crop-year	1989-2024
Harvested area non-irrigated	Ag census/survey	County-crop-year	1989-2024
Share with $\geq$ 10 years on any operation	Ag census	County-year	2017, 2022
Share with $\geq$ 10 years on present operation	Ag census	County-year	2017, 2022
Net cash farm income (IHS)	Ag census	County–year	1997, 2002, 2007, 2012, 2017, 202
Farm-related income (log)	Ag census	County–year	1997, 2002, 2007, 2012, 2017, 202
CCC net payments	Ag census	County–year	1997, 2002, 2007, 2012, 2017, 202
CCC net payments (IHS)	Ag census	County–year	1997, 2002, 2007, 2012, 2017, 202
Share of female producers	Ag census	County–year	2017, 2022
Fertilizer expenditure per acre	Ag census	County–year	1997, 2002, 2007, 2012, 2017, 202
Fertilizer expenditure per acre (IHS)	Ag census	County–year	1997, 2002, 2007, 2012, 2017, 202
Share of minority producers	Ag census	County–year	2017, 2022
Cattle per planted acre	Ag survey	County-year	1989-2022
Cattle per planted acre (IHS)	Ag survey	County-year	1989-2022
Share of acres owned	Ag census	County–year	1997, 2002, 2007, 2012, 2017, 202

Notes: Variables listed in this table are observed only in quinquennial Census years. When both Census and survey values are reported fo measure, I retain the Census-reported value to avoid double counting. Area planted and harvested (including (not)irrigated) are for the se selected crops ("barley", "corn", "cotton", "peanuts", "sorghum", "soybeans", "wheat"). All other variables are for all crops. Filling miss values: To construct a balanced county–year panel, I first reorganize the data so that each county is observed in every year of the analysis peri explicitly creating county–year observations where data are missing. For variables that evolve slowly over time, such as demographic or owners shares, missing observations are filled by carrying forward the most recent available value within a county and, where necessary, carrying val backward to earlier years. For continuous variables, such as income, expenditures, or production intensity, missing values are filled by linea interpolating between observed years within the same county, with extrapolation applied only at the boundaries of the sample. After complet the imputation of the underlying variables, all transformed measures (e.g., inverse hyperbolic sine or logarithmic transformations) are recompu to ensure consistency.

Number of Heating Degree Days, Moderate Degree Days, Cooling Degree Days, and Rainfall I built up on the method of Annan and Schlenker (2015) and calculated the HDD and MDD for the period of study for the contiguous United States. Unlike the past efforts for capturing the impact of climate variables on agriculture output that only included the *cropping-season* values, I followed Farris et al. (2023) and studied the impact of these climate variables in over the year, and over the cropping season.

Summary of Business Data Many farmers insured under the Federal program, particularly in the mid 1990s, held only catastrophic coverage (CAT) for their crops. CAT coverage is fully subsidized by the government, requiring only a nominal administrative fee (typically $50 or $100, depending on the year) for each crop a farmer chooses to enroll in a given county, regardless of acreage total. Coverage provided by CAT, however, is very low, typically covering losses in excess of 50 percent of expected yield at 55 percent of the price. To focus on insurance more likely to influence behavior, we define the *insured* as those having coverage in excess of CAT coverage, generically referred to as "buy up" coverage in much of the literature. I followed the approach used by Roberts, E. J. O'Donoghue, and Key (2007) and considered all insurance coverage above the CAT level "insured" and considered CAT or no insurance as "uninsured." This paragraph above is also from Roberts, E. J. O'Donoghue, and Key (2007) and needs rewriting.

Table 4: List of variables included in the regressions

Variable	**Description**
Crop insurance	
Crop insurance	Self-reported adoption status (= 1 if adopted; = 0 otherwise)
Socioeconomic characteristics	
Age	Farmer's age (number of years)
Male	Farmer's sex (= 1 if male; = 0 otherwise)
Reside in plot	Place of residence (= 1 if agricultural plot; = 0 otherwise)
Education	Farmer's level of education (number of years)
Dependence	% total income obtained from agriculture (1 = less than 25%, ..., 4 = 75% or more)
Institutional characteristics	
Credit	Access to credit instruments by public or private institutions during the last 2 years (=
Secure tenure	% family-owned and rental land hectares over total hectares of the farm (% hectares)
Extension services	Benefited from extension services during the last 2 years (= 1 if yes; = 0 otherwise)
Participation in organisations	Membership in agricultural organisations (= 1 if yes; = 0 otherwise)
Number insurance adopter	Insurance adopters per locality, regardless of crops (number of adopters)
Farm size	
Yield	Agricultural yield (kg/hectare)
Total surface	Size of the agricultural land/farm (number of hectares)
Capital	Agricultural machinery and tools available in the farm (index)
Technology adoption	
Improved seed	Self-reported adoption status (= 1 if adopted; = 0 otherwise)
Modern irrigation	Self-reported adoption status (= 1 if adopted; = 0 otherwise)
Biological control	Self-reported adoption status (= 1 if adopted; = 0 otherwise)
Environmental and location variables	
Rainfall	Cumulative precipitation (millimetres)
Soil quality	% of agricultural land with non-eroded and slightly eroded soil (% hectares)
Location variables	
North zone	Agricultural land/farm located in the North zone (= 1 if yes; = 0 otherwise)
Central zone	Agricultural land/farm located in the Central zone (= 1 if yes; = 0 otherwise)
South zone	Agricultural land/farm located in the South zone (= 1 if yes; = 0 otherwise)

Table 5: Descriptive statistics of major explanatory variables (type of farmer)

	Total		**Family**		**Non-family**	
Variable	Mean	SD	Mean	SD	Mean	SD
Crop insurance	0.022	0.148	0.015	0.122	0.120	0.325
Socioeconomic variables						
Age	57.890	14.462	57.816	14.458	59.017	14.491
Male	0.753	0.431	0.751	0.433	0.796	0.403
Reside in plot	0.748	0.434	0.772	0.420	0.441	0.497
Education	1.909	1.761	1.784	1.609	3.806	2.655
Dependence	1.411	1.219	1.371	1.214	1.940	1.176
Institutional variables						
Credit	0.084	0.277	0.060	0.237	0.394	0.489
Secure tenure (% hectares)	0.851	0.329	0.843	0.337	0.948	0.196
Extension services	0.220	0.414	0.226	0.419	0.141	0.348
Participation in organisations	0.204	0.403	0.184	0.388	0.457	0.498
Number insurance adopter	25.374	35.352	24.641	34.875	34.781	39.801
Farm size						
Yield (kg per hectare)	2,814.4	3,315.3	2,682.4	3,349.4	4,505.5	2,234.5
Total surface (hectares)	38.063	194.710	17.199	22.458	305.503	663.655
Capital	0.083	1.208	−0.092	0.585	2.317	3.226
Technology adoption						
Certified seed	0.213	0.409	0.180	0.384	0.628	0.483
Biological control	0.038	0.191	0.030	0.170	0.142	0.350
Modern irrigation	0.050	0.218	0.041	0.198	0.171	0.377
Environmental and location variables						
Rainfall (mm)	911.195	917.788	323.163	826.687	366.860	—
Soil quality (ratio)	0.283	0.189	0.283	0.186	0.289	0.223
Location variables						
North zone	0.007	0.081	0.005	0.073	0.024	0.153
Central zone	0.159	0.366	0.145	0.352	0.352	0.478
South zone	0.833	0.373	0.850	0.357	0.624	0.485
Observations	42,531		39,874		2,642	

Source: Authors' estimates based on 2007 Census data.

Effective Subsidy Rate The USDA/RMA do not report a time series of subsidy rates on crop insurance. Rather, the summary of business data set of the RMA reports the total monetary values of premium payments of farmers, and subsidies received by farmers at the county-crop-year. The subsidy schedules are a function of coverage levels that farmer adopt (such as 50%, 55% etc.), unit structure of their farms (optional, basic, enterprise, or whole farm), and insurance plan (APH vs Revenue insurance, PRF, WFRP, etc). As the mix moves toward higher coverage/revenue plans after ARPA-2000, the national share rises even if schedules themselves don't change.

I used the data columns of liability, total premium, subsidy, producer premium, plan and coverage level to calculate the subsidy rates presented to farmers over the study period and the subsidy per \$100 liability for crop c in year t as follows[3]

$$\text{Effective Subsidy Rate}_{t,c} = \frac{\sum_c \text{Subsidy}_t}{\sum_c \text{Total Premium}_t}, \qquad \text{Subsidy per \$100 Liab}_{t,c} = 100 \cdot \frac{\sum_c \text{Subsidy}_t}{\sum_c \text{Liability}_t}.$$

4 Methodology

I use six different means to measure participation in the crop insurance program: percent of planted acres enrolled, level of total liability per acre, total acres enrolled, number of acres enrolled in in buy-up coverage[4], level of total premiums, and level of total premiums per acre [5]. The level of total premiums is the first once actually included? it was on page 14 of this paper, level of total premiums per acre, level of total liability per acre and percent of planted area that is insured, capture the intensity of use of crop insurance. Large movements in crop prices can cause these variables to change from year to year while yields tend to trend upwards over time with technology advancement, making it difficult to compare a policy purchased in one year with a policy purchased in a subsequent year. For this reason, I normalized these measures throughout the the analysis to 2002 prices and yields. With this, if the value of total premiums or liabilities changed over time, it would be due to the underlying changes in the quantity of crop insurance demanded and would not be attributed to price or yield changes. The premiums, premiums per acre, and liabilities per acre were all multiplied by the ratio of 2002 expected revenues to 1997 expected revenues (expected yields are derived from a linear regression of normalized yields from 1966 through 2001 on a time dummy for 2002 and expected prices come from Chicago Board of Trade futures prices for the relevant crop's harvest price at the time of planting for the 2002 crop; similar calculations were used for the 1997 prices and yields) (E. O'Donoghue 2014)..

I defined a transfer payment multiplier for crop subsidies that government offers to the farmers. The transfer payment multiplier in the context of crop subsidies refers to the ratio of the total income generated in the recipient economy to the amount of subsidy transferred. In other words, it measures how much additional income is created for farmers (and sometimes the broader rural economy) for every unit of government subsidy provided. Multiplier at horizon h is

$$\mathbf{M}_h \equiv \Delta Y_{t+h} / \Delta T_t \tag{1}$$

where T_t is government transfers to farmers and Y is output (GDP, local earnings, etc.).

[3] I ignored the coverage category of catastrophic insurance (CAT) for which farmers received 100% insurance subsidy in this calculation. I also focused on the yield-based and revenue-based insurance plans of yield protection (YP), average production history (APH) and revenue protection (RP) in the calculation.

[4] those policies above the fully subsidized catastrophic coverage available to farmers

[5] In the earlier versions of this paper, I measured participation in crop insurance program only using the first two parameters. I expanded the list following the method of E. O'Donoghue (2014).

Choose of Variables I used Lasso method (Tibshirani 1996) for choosing the relevant variables out of the pool of covariates I had.

Should I have the insurance subsidy percentage as the shock, or the percentage changed compared to the baseline year, say, 1989? Many policy changes have taked place over the time. How can I single out the effect of 1 policy out of all others? Define policy-episode shocks tied to legal changes (e.g., 1994, 2000, 2014 reforms), and measure each unit's exposure to that episode using formula-driven, pre-policy characteristics. Then run event-study or continuous-treatment DiD around those episodes.

The PROCAMPO program in Mexico is a well-documented example for transfer payment multiplier. Studies found that the income multiplier for recipients ranged from 1.5 to 2.6, meaning each peso transferred generated 1.5 to 2.6 pesos in total income for farm households, especially where liquidity constraints were significant 216. In China, the grain subsidy program was also analyzed for its multiplier effects, with findings indicating a high income multiplier and significant income promotion effects for farmers 1. These multipliers are typically higher for larger or better-endowed farms and in regions with more economic opportunities, reflecting the role of liquidity constraints and asset endowments 216. See F. Yi, W. Lu, and Zhou (2016) and Sadoulet, De Janvry, and Davis (2001)

Dependent Variables The outcome of interest is the change in crop insurance demand measured one of six ways: percent of planted acres enrolled, level of total liability per acre, total acres enrolled, number of acres enrolled in in buy-up coverage, level of total premiums, and level of total premiums per acre.

Control Variables The control variables in this study that vary over time are

- The lagged change in the number of acres of a particular crop in the county,
- The lagged change in the number of acres run by female operators. Since it has been shown that the gender of the individual matters when it comes to dealing with risk (Halek and Eisenhauer 2001; Pawlowski, Atwal, and Dunbar 2008; Byrnes, D. C. Miller, and Schafer 1999).
- The 1-year lagged change in the ratio of yield policies to revenue policies. what does it mean? It is from page 18 of E. O'Donoghue (2014).
- The mean and variance of yield distribution: to explore any changes to the yield distribution, I use county-level yield data and detrends it and normalizes it to a base year's yield (1989) for each crop. This allows the yields to be compared over time and viewed as a distribution of yields for each year examined. I then constructed the mean and variance of these yield distributions and differenced to control for any changes in the yield distribution that may have induced changes in crop insurance participation. I adopted this treatment of the crop yield before using in the analysis from E. O'Donoghue (2014, pp.18-20).
- The change in a 3-year measure of lagged returns to crop insurance measured as total indemnities divided by total premiums paid by the farmer
- The difference in a 1-year lagged, actual-versus-expected revenue, differenced over time.

I used the last two sets of variables to capture the general state of affairs in the years (or year) leading up to the period examined, as recommended by E. O'Donoghue (2014, page 18). In terms of time-invariant control variables, because the regression is run at the county level, time-invariant

county-level controls are controlled for through the differencing. Essentially, differencing acts as a county-level fixed effect, which eliminates the need to include variables that do not change over time, such as farmers' age, average farm size, average county size, and land quality. To do: variables to add when dependent variable that is not a measure of change. The regression also includes region-by-year fixed effects This variable is from E. O'Donoghue (2014). I think region is from the Quickstats dataset that generate comparisons among counties within regions that were created based on soil and climatic attributes (crop reporting districts). Note that this is a fixed effect that captures trends that can vary by region. Any changes that differ across regions will be captured by these variables, such as weather shocks and price movements not picked up by other variables. Finally, the error, captures other unobserved factors affecting crop insurance demand, such as within-region weather variations.

A third group of control variables address the alternative methods for farmers to deal with risk, which might affect producers' willingness to buy crop insurance, including the PLC program and the ad-hoc disaster assistance. Inclusion of these variables sets this paper apart of papers like E. O'Donoghue (2014) that narrow the time window of study and limit the need for considering other concurrent policies.

Variables construction These paragraphs are copied from pages 20-21 of E. O'Donoghue (2014)The mean and variance of the yields are calculated using NASS yields collected from 1975 through 2002. For each county, yields are first detrended using a simple linear model, regressing the 27 years of data on a year variable. Following the study of Barry K Goodwin and Alan P Ker (1998), who found that the standard deviations of the yield tend to be proportional to the level of the average yield, normalized yields were created using the intercept, slope, and residuals from the regression in the following manner:

$$\bar{y}_t = y_{1989} \times (1 + \frac{e_t}{y_t}) \tag{2}$$

where $\bar{y}_t$ denotes the normalized yield for time t, e_t represents the residual from the regressionI think it means the detrending regression, and y_t is the predicted yield from the linear regression. With 27 years of data, equation 2 generates normalized yield observations for each county, allowing one to calculate a separate mean and variance for crop yields for each county. Distributions are constructed for both 1996 and 2001, and means and variances are constructed for each of these lagged years. They are then differenced to provide a change in the mean and variance of the detrended, normalized county-level yields.

The 3-year returns to crop insurance variable was constructed by dividing the indemnities by the premium paid by the producer for each insurance plan for the relevant crop and summing them together, weighted by their share of total acres enrolled in each plan. This is done for each of the 3 years preceding 1997 and 2002 (e.g., for 1997, the years 1996, 1995, and 1994 were used). The returns for the 3 years were then averaged to obtain two, 3-year average returns to crop insurance for each crop in each State (one for the years leading up to 1997; a second for the years leading up to 2002). These were then differenced and used in the regression.

If a producer experienced a loss in the year prior to that examined, he or she may be more inclined to enroll in crop insurance in the following year. To observe this over time, if a farmer fared worse in 2001 relative to 1996, he or she might find crop insurance more attractive in 2001 than in 1997. Therefore, to construct such a variable, the actual and expected revenues were calculated for 1996 and 2001. Actual revenues were generated using NASS price and yield data. Expected revenues were generated using the predicted yields from the detrending linear regression process discussed earlier

and national level futures commodity prices (assuming away basis differences between counties). After the actual and expected revenues for 1996 and 2001 were constructed, they were differenced. The difference is the gain/loss for the year. The resulting gain/loss for 1996 was then subtracted from that of 2001 to obtain a measure of relative gain/loss over time.

The crop insurance ratio used in the analysis was constructed using RMA data from the Summary of Business. The total premiums of all yield policies were summed up and divided by the total level of the sum of all yield policy and revenue policy premiums. Finally, when running the regressions, if a county had a change in total premiums per acre, change in total liabilities per acre, or a change in total insured acres that fell outside of four standard deviations from the mean of those variables, the observation (county) was omitted from the analysis. Results were also run (but not reported—available upon request) without omitting any variables, and results remained very similar.

Endogeneity I use the introduction of the premium subsidies as the source of price variation. This necessitates the careful examination of the relationship between the demand for crop insurance and the level of subsidies to derive causality. I am trying to identify a change in subsidy levels due to the policy change from the result of a change in demand and as a result, unrelated to the new policy. The second scenario is also probable, since subsidy levels are a function of the level of coverage demanded by the producer.

4.1 Instrumental Variable (IV)

I use an instrumental variables framework to decompose the change in crop yield into two channels: extensive margin of enrollment in crop insurance (by fraction of planted land insured), intensive margin of enrollment in crop insurance (amount of insurance liabilities of farmers). I control for climate variables and other demography variables in the regression as well. I use instruments that provide plausibly exogenous variation in intensive and extensive margins, and I allow the effects to vary for areas by the amount of exposure to extreme weather conditions. The instrument for extensive margin is the driving distance between USDA county offices and population-weighted county centroid. The instrument for intensive margin is the number of weather stations in each county.

I used this distance as instrumental variable for the identification in the model. In case of having two instruments for one endogenous variable, do 2sls method. Do the Durbin-Wu-Hausman Test for endogeneity.

1- Past years as an instrument. I used the change in subsidy rate between the previous two rounds of Farm Bills as an instrument for the immediate following Farm Bill. Take the example of 1994 and 1996 Farm Bills. In order to study the effect of changes in subsidy rates between 1995 and 1996, I instrument for subsidy change with the change in subsidy between the 1985 and 1994 Farm Bills, and use the difference between subsidy rates in 1994 and 1989 (first year in our data set). This is a valid instrument, because by the time farmers decide about crop insurance in crop year 1995, the decision about crop insurance in 1994 is already made. Therefore, the decision for 1994 crop year and its change from 1989 can be considered exogenous to the decision about quantity of crop insurance demanded in 1995.

2- Actuarial rating parameters of the USDA as an instrument. I used the actuarial rating parameters of the USDA RMA as instrument. These govern premium rates in a way that is plausibly

exogenous to any individual farmer's crop insurance decisions (Woodard and J. Yi 2020; Tsiboe and Turner 2023b).

explain the instrument more. Is there a graph you make to illustrate the first stage? A standard test of an IV is a covariance balance test. Include it. Show you F-stats for your first-stage regressions. The decision of farmers over the acreage of land to insure is a function of the insurance price (= premium rate) they face. This price itself is set as a function of risk for a given insurance pool[6], as well as the producer's production choices about the level of insurance coverage. This is because an insurance pool with a higher risk of initiating an indemnity payment has a higher premium rate for a given crop insurance choice. This creates a case for endogeneity and simultaneity problems. Farmers' decisions about production, insurance purchase, and the price they face for insurance are intertwined, through a link of unmeasurable "behavioral" characteristics of farmers.

To address this endogeneity problem, we adopt the identification strategy developed by Tsiboe and Turner (2023a) and Tsiboe and Turner (2023b). The authors rely on the rating framework used by RMA (i.e., continuous rating) and develop an instrumental variable (τ). The main idea is that FCIP premium rates are determined using exogenously set rating parameters. This lets us use the parameters to generate exogenous variation in premium rates that producers face when making crop insurance decisions. Tsiboe and Turner (2023a) and Tsiboe and Turner (2023b) estimate the parameters used by the RMA and then use these values as instrumental variables to resolve the endogeneity problem of farmer's decision about farming and purchasing crop insurance.

RMA updates rating parameters on a 3-year cycle using historic loss experience data from the previous 20 years starting from two crop years before the current crop year (Tsiboe and Turner 2023b). The previous yields reported by producers contribute to the producer's historical loss experience but are fixed and not under the control of producers once reported. Consequently, FCIP premium rates can be thought of as sequentially exogenous to producer insurance decisions. This assumption is plausible because the rating parameters are finalized in advance. This exogenous portion of PCIP premium rates is driven by the continuous rating parameter and the endogenous portion by the insured's choice of insurance unit, coverage level, and their historic production experience.

3- Number of livestock that farmers have as an instrument It is from the US agricultural census data.

4- Density of weather observation stations. I exploit spatial variation in the historical density of weather observation stations. The underlying economic mechanism is that a denser local observation network improves the accuracy of area, as level weather measures used in insurance rating and indemnity determination, thereby reducing basis risk and measurement error in the insurance contract. Lower basis risk increases the expected accuracy of indemnity payments from the farmer's perspective, making crop insurance more attractive and increasing insurance take up. Importantly, this channel operates through the quality of the insurance contract and the information used in risk assessment, rather than through any direct effect on agronomic decisions or crop yields themselves. Figure 14 shows the number of weather stations in the study area in the base year of 1989.

[6] RMA defines an insurance pool as a group of producers formed by the unique combination of county, crop, crop type, and production practices (Tsiboe and Turner 2023b).

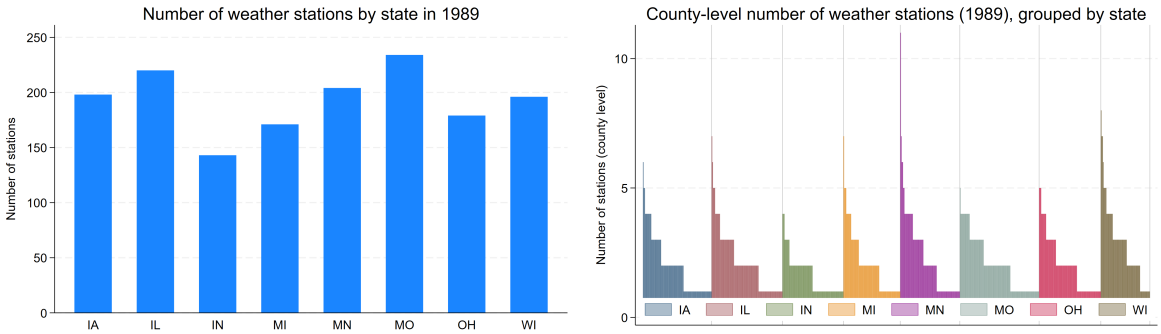


Figure 14: Number of weather stations by state in 1989 (authors' calculation using NOAA National Centers for Environmental Information (2025b) and NOAA National Centers for Environmental Information (2025a)).

5- Sales closing dates of crop insurance Under the Federal Crop Insurance Program, insurance coverage decisions must be made by a crop- and region-specific sales closing date set administratively by the Risk Management Agency. These dates are uniform across counties within a given crop–region and occur prior to planting and before the realization of growing-season weather shocks. For major row crops such as corn and soybeans, sales closing dates have remained largely unchanged since the 1990s, with only infrequent, policy-driven adjustments. This institutional feature ensures that insurance take-up is predetermined with respect to contemporaneous weather realizations and limits strategic timing by producers.

6- Shortest/mean/median driving distance between the County centroids and USDA offices as an instrument. Given the historical nature of the period of this study, and the requirement for in-person processing of the insurance contract by farmers., access to the USDA local offices could be affected by the proximity or distance of farmers from those offices. This is a classic "geographic access to public service" problem. To my knowledge, the only prior work that has studied the impact of distance from USDA offices on farmers' choices, is Diekmann et al. (2012). Diekmann et al. (2012) used survey data on Ohio farmers to study the trade offs that farmers are willing to make when choosing between local agriculture extension offices based on distance to the local office, differences in educational programs offered, expertise level of staffing educators, and the level of local taxation required to support an extension office.

The empirical strategy exploits this shift by instrumenting insurance enrollment using measures of exogenous variation in access to FCIP administration, such as distance to the nearest USDA or Risk Management Agency (RMA) service center. Following the MPCI expansion, proximity to program offices plausibly affected enrollment through informational frictions, transaction costs, and assistance with policy selection, while being orthogonal to short-run yield realizations and idiosyncratic weather shocks (Glauber 2013). Importantly, this exclusion restriction is less credible in the earlier named-peril regime, when insurance availability and contract design were explicitly tailored to local risk conditions.

To quantify geographic access to USDA service centers, I construct a measure of the road travel distance between each county's geographical center and the nearest USDA office using the contemporary roads network[7]. First, I use 1990 county boundary shapefiles from the U.S. Census and compute the geographic centroid of each county. I pair these with the geocoded locations of USDA offices in the six

[7] I am in the process of getting the historical road data, and will update the analysis with that once I receive it.

Midwestern states of Illinois, Indiana, Iowa, Michigan, Minnesota, Missouri, Ohio and Wisconsin. For each state, I extract the full road network from OpenStreetMap using OSMnx (in Python) and build a state-specific drivable graph. I then project both county centroids and USDA offices onto the nearest nodes of that road network and compute the shortest road distance between each county and every USDA office using Dijkstra's algorithm (Dijkstra 1959) with distance as the weight. For computational stability, states are processed independently, and the resulting distances are saved separately before being merged into a unified dataset. The final variable, *distance to nearest USDA office*, captures minimum drivable kilometers and serves as a proxy for a county's access to USDA administrative services. How about calculating the minimum, mean and median driving distance and experiment with them? Because the IV with minimum driving distance turned out to be weak.

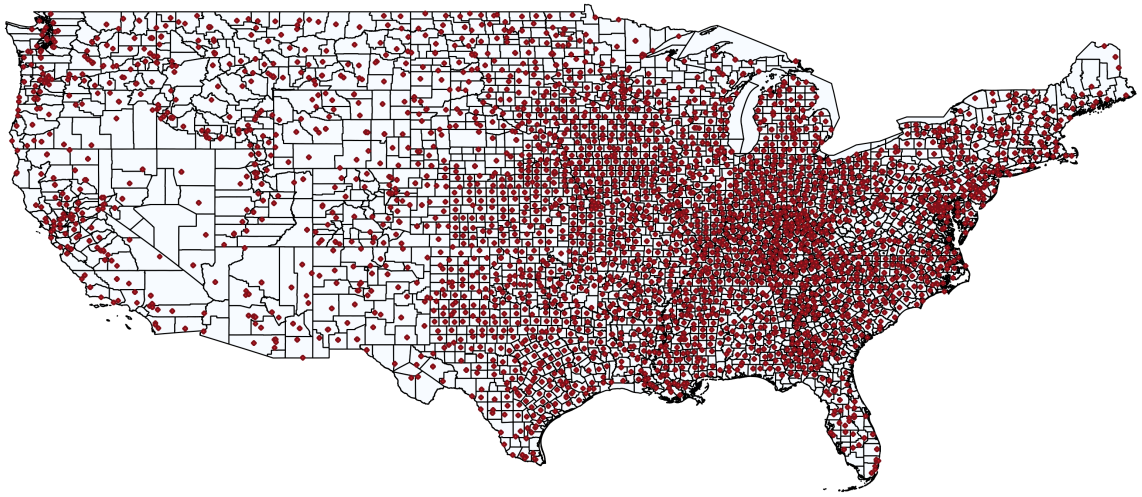
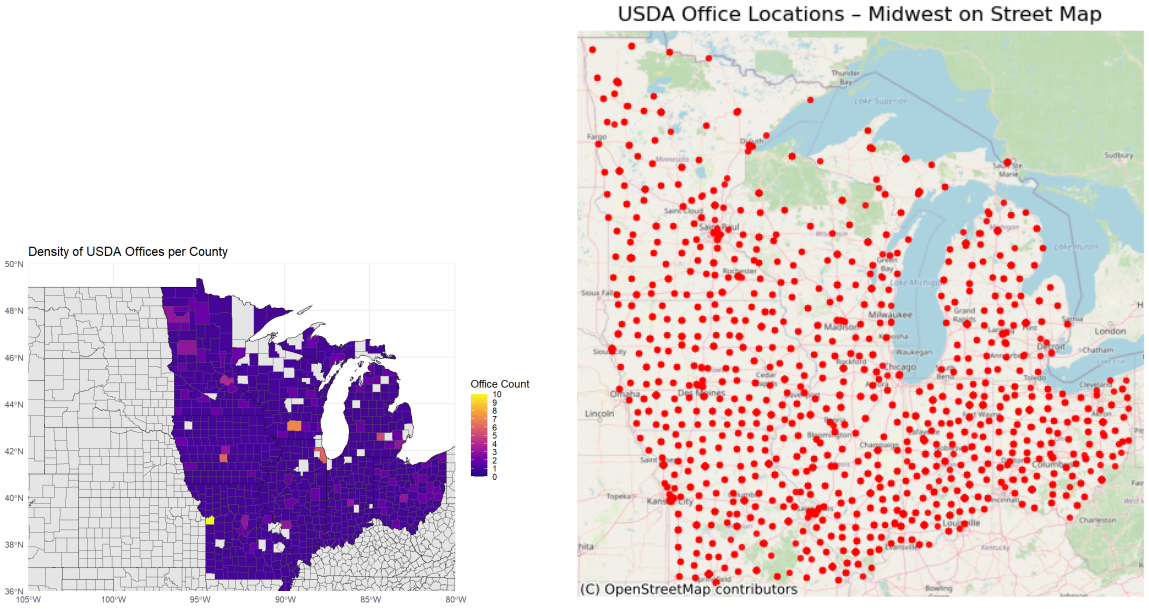


Figure 15: Locations of USDA offices in 2025 (based on 1990s county boundaries and 2025 USDA office locations)

Following standard IV diagnostics, I perform a *covariate balance test* test to verify whether the instrument (distance to USDA office) predicts predetermined covariates. I regress each covariate on the instrument with county and year fixed effects. Because the instrument is time-invariant and the specification includes county-crop fixed effects, all identifying variation is absorbed. Table 6 reports the results for this test. As expected, the instrument is omitted from these regressions, indicating that it has no residual correlation with any of the predetermined covariates. This supports the orthogonality condition required for instrument validity.

Figure 17 plots county-level averages of predetermined covariates against distance to the nearest USDA office. As expected, distance is correlated with cross-sectional features such as precipitation and historical crop prices: USDA offices tend to be located closer to high-production agricultural counties. However, these cross-sectional correlations are fully absorbed by county–crop fixed effects in the empirical specification. Indeed, the formal covariate-balance test shows that the instrument is omitted once fixed effects are included, indicating no remaining correlation with time-varying predetermined covariates. Therefore, identification relies on cross-sectional variation but is not threatened by omitted time-varying confounders.

The first-stage regression shows that distance to the nearest USDA office is statistically associated with insurance take-up ($\hat{\beta} = -0.091$, $p < 0.01$). However, the first-stage F-statistic is 7.60, below the conventional threshold of 10 for strong instruments (Staiger and Stock 1997). This indicates that although the instrument is relevant, it is relatively weak once state and year fixed effects and



(a) Density of USDA offices

(b) USDA offices and their locations relative to roads

Figure 16: The condition of access to USDA offices in the Midwest. Base map is from OpenStreetMap contributors (2025), county boundaries based on the 1990 census (U.S. Geological Survey 1990) and USDA office locations if from USDA Farm Service Agency (2025).

Table 6: Covariate Balance: Distance to USDA Office and Predetermined Covariates

	Outcome: Predetermined Covariate X		
Covariate X	Coefficient on distance (km)	Std. error	Observations
$ddayMod_100$	0 (omitted)	–	29,990
$ddayHot_100$	0 (omitted)	–	29,990
$prec_cm$	0 (omitted)	–	29,990
$prec2_cm$	0 (omitted)	–	29,990
$price_received$	0 (omitted)	–	29,990
County–crop FE	Yes		
Year FE	Yes		
Cluster level	County–crop		

Notes: Each row reports results from a separate regression of the listed predetermined covariate on the instrument, distance (in km) to the nearest USDA office, controlling for county–crop and year fixed effects. Standard errors are clustered at the county–crop level. Because the instrument is time-invariant and perfectly collinear with the county–crop fixed effects, it is absorbed in all specifications and the coefficient on distance is omitted. This implies that the instrument has no residual explanatory power for the predetermined covariates after conditioning on fixed effects, consistent with covariate balance.

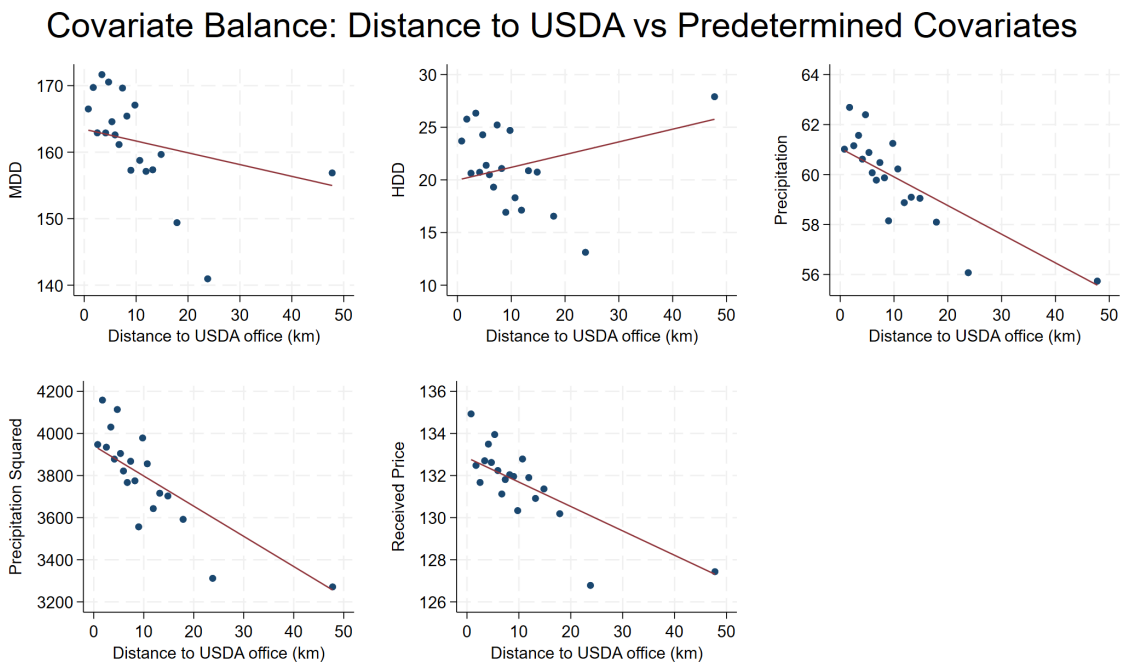


Figure 17: Covariate Balance: Distance to USDA vs Predetermined Covariates using the OpenStreetMap and simple county centroids

weather controls are included. Consequently, IV estimates using this instrument may suffer from weak-instrument concerns, and inference should be interpreted with caution.

Instrumental variable strength and validity: We assess the strength of our instrumental variables by estimating the first-stage regressions for both endogenous variables: subsidy per dollar of liability and the percentage of land insured.

$$ln(\Gamma_{ijt}) = \alpha_0 + \alpha_1 \tau_{ijt} + \alpha_2 EP_{ijt} + \alpha_3 Y_{ijt-1} + \sum_j \alpha_{j4} Time_{jt} + + \alpha_{j5} W_{ijt} + \alpha_{jw6} W_{ijt} \Gamma_{ijt} + \nu_{ij} + \epsilon_{ijt} \quad (3)$$

Fixed effects are at the county-crop level, and robust standard errors are clustered at the county-crop.

Report F-stats

Table 8 shows that across all four specifications, the RMA rating parameter (τ) is both statistically strong and economically meaningful. ??? Give me the F-stats, that is how people assess instrument strength. When the endogenous variable is sublia ??, the coefficient on τ lies between 0.062 and 0.068, highly significant at the 1% level. Similarly, when the dependent variable is fracInsured, the coefficient on τ ranges between 50 and 55, again precisely estimated and economically large. These results indicate that exogenous variation in actuarial rating rules translates directly into variation in subsidy incidence and acreage insured. This satisfies the relevance condition for instrumental variables and mitigates concerns about weak instruments. Moreover, the inclusion of weather interactions, lagged yields, and crop-specific time trends ensures that identification is not confounded by omitted risk factors. Together, these results provide a credible empirical foundation for using τ as an instrument to estimate the causal

Table 7: First-stage Regression: Insurance Take-up on Distance

	(1) fracInsured
drive_km_to_usda	-0.0909**
	(0.0330)
ddayMod_100	-0.0327
	(0.0223)
ddayHot_100	-0.0654***
	(0.0166)
prec_cm	-0.0895
	(0.0471)
prec2_cm	0.000486
	(0.000323)
price_received	-0.0243***
	(0.00547)
Constant	76.71***
	(4.186)
Observations	30003

Standard errors in parentheses
State and year fixed effects included.
,
Standard errors clustered at the county level.
,
Instrument F-statistic (drive_km_to_usda): 7.60
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8: First stage results of the instrument variable for the baseline (no weather variables), and the complete model.

VARIABLES	(1) sublia	(2) sublia	(3) fracInsured	(4) fracInsured
τ	0.0679***	0.0625***	55.06***	50.27***
	(0.00686)	(0.00688)	(5.060)	(5.026)
Last period yield	5.61e-05***	7.67e-05***	-0.00995**	-0.00514
	(4.73e-06)	(5.08e-06)	(0.00434)	(0.00472)
Received price	-3.87e-05***	-4.86e-05***	-0.114***	-0.111***
	(3.80e-06)	(4.01e-06)	(0.00312)	(0.00322)
MDD		-8.51e-05***		-0.0101*
		(6.56e-06)		(0.00570)
HDD		-6.21e-05***		0.0785***
		(7.03e-06)		(0.00538)
Precipitation		0.000186***		0.803***
		(3.51e-05)		(0.0332)
Squared precipitation		-1.37e-06***		-0.00573***
		(2.49e-07)		(0.000238)
Corntime	0.00202***	0.00208***	2.442***	2.406***
	(4.03e-05)	(4.10e-05)	(0.0293)	(0.0300)
Soybeantime	0.00209***	0.00217***	2.488***	2.475***
	(3.33e-05)	(3.38e-05)	(0.0262)	(0.0266)
Constant	0.0117***	0.0199***	40.76***	14.42***
	(0.00105)	(0.00169)	(0.778)	(1.503)
Observations	28,189	28,189	28,189	28,189
R-squared	0.570	0.579	0.611	0.620
Number of panelID	1,291	1,291	1,291	1,291

Notes: Robust standard errors in parentheses (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.).
Model 1-2 have subsidy per dollar of liability (sublia) as endogenous variables, and models 3-4 have fraction of land insured (fracInsured) as endogenous variable.

effect of crop insurance subsidies on yield responses to weather extremes. You'll need to do a much better job of explaining what the IV does and why it satisfies the exclusion restriction. I am not convinced, and IV papers live (and die) by how good a job authors do of convincing readers that the IV is valid. I stopped reading here..

7- Shortest driving distance between the population-weighted County centroids and USDA offices as an instrument. To measure geographic access to agricultural service infrastructure, I construct a county-level measure of driving distance to the nearest USDA offices. First, I compute population-weighted county centroids using Census 2000 block-level population counts from IPUMS NHGIS Schroeder et al. (2025a). For each census block, the geometric centroid is calculated, and county-level centroids are obtained as the population-weighted average of block centroids within each county, ensuring that the resulting location reflects the spatial distribution of residents rather than land area.

To approximate travel distance along the road network, I use the U.S. Census Bureau's TIGER/Line road network files from the year 2000 (Bureau 2000). Road segments are converted into a graph representation in which nodes correspond to segment endpoints and edges represent road links weighted by their physical length. For each population-weighted county centroid and each USDA office location, the nearest road network node is identified. Driving distance is then computed as the shortest-path distance along the road network between the centroid's nearest node and the nearest node associated

with a USDA office.

All spatial layers are projected to a common projected coordinate reference system (NAD83 / USA Contiguous Albers Equal Area) prior to distance calculations to ensure consistency and metric accuracy. Distances are computed separately for each state to reduce computational complexity, and for each county the minimum driving distance across all USDA offices within the state is retained. The resulting measure captures effective geographic accessibility to USDA offices while accounting for both population location and the structure of the historical road network.

Covariance Balance Test for this IV. To assess the plausibility of the exclusion restriction, I examine whether counties located farther from USDA offices differ systematically in predetermined characteristics (Table 9). Regressions of historical climate variables and crop prices on distance to the nearest USDA office, controlling for year fixed effects and clustering at the county level, reveal no statistically significant relationships for temperature exposure or output prices. While precipitation levels are modestly correlated with distance, this reflects well-known spatial climate gradients and is directly controlled for in all specifications. While precipitation is modestly correlated with distance, this pattern reflects smooth spatial climate gradients rather than institutional sorting, and precipitation is flexibly controlled for in all specifications. Consistent with these results, binscatter plots using long-run county–crop averages in Figure 19 reveal no economically meaningful nonlinearities or discontinuities. Together, these findings support the plausibility of the exclusion restriction. Overall, these results suggest that distance to USDA offices is not strongly associated with observable determinants of insurance demand, supporting its use as an instrument. Create an appendix for tables and move the results of balanced tests there.

Table 9: Covariate Balance Test for IV of Driving Distance Between Population-Weighted County Centroids and USDA Offices

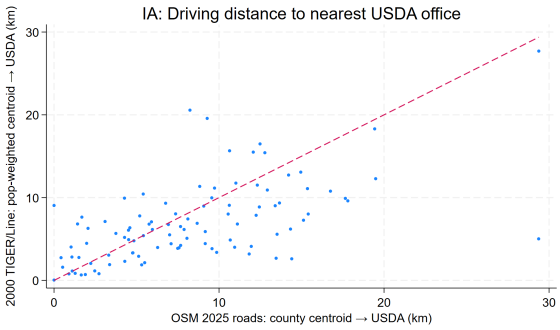
	MDD	HDD	Precipitation	Precipitation squared	Received price
drive_km_to_usda	-0.056	0.137	-0.079***	-10.307***	-0.069
	(0.207)	(0.124)	(0.023)	(2.930)	(0.047)
Observations	29616	29616	29616	29616	29616
R-squared	0.161	0.278	0.267	0.276	0.172

Standard errors in parentheses

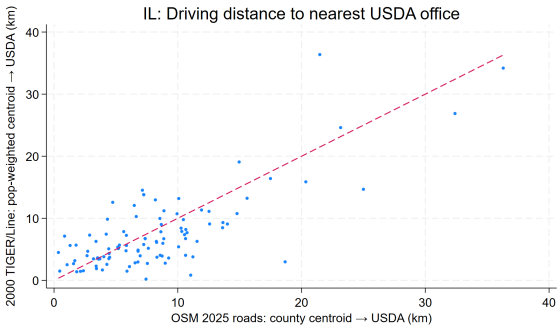
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Instrumental variable strength and validity: I assess the strength of this instrumental variable by estimating the first-stage regressions for both endogenous variables: subsidy per dollar of liability and the percentage of land insured.

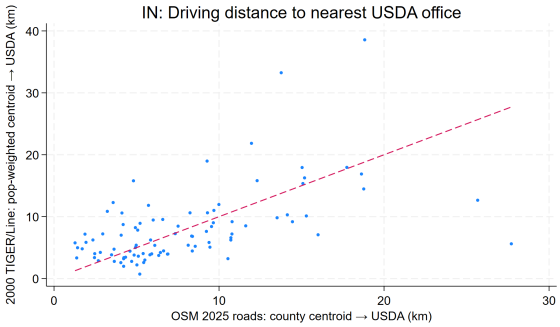
$$Old: ln(\Gamma_{ijt}) = \alpha_0 + \alpha_1 \tau_{ijt} + \alpha_2 EP_{ijt} + \alpha_3 Y_{ijt-1} + \sum_j \alpha_{j4} Time_{jt} + +\alpha_{j5} W_{ijt} + \alpha_{jw6} W_{ijt} \Gamma_{ijt} + \nu_{ij} + \epsilon_{ijt}$$



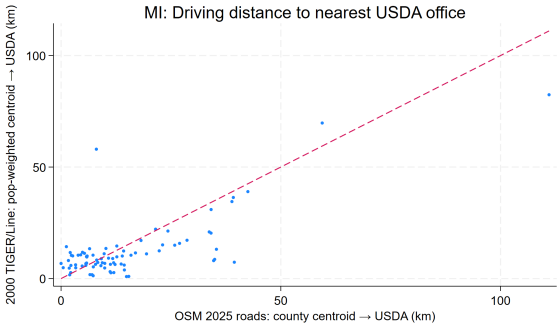
(a) Iowa



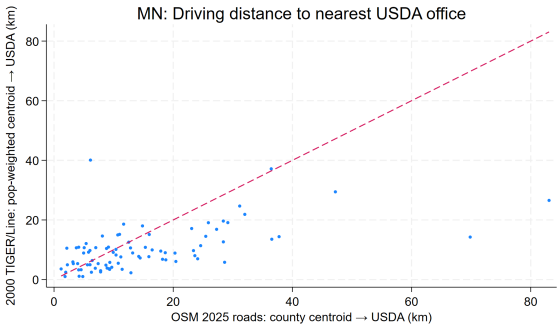
(b) Illinois



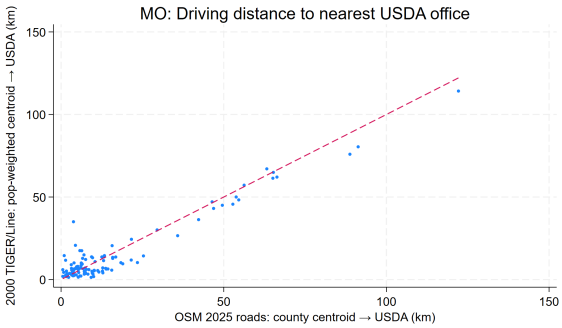
(c) Indiana



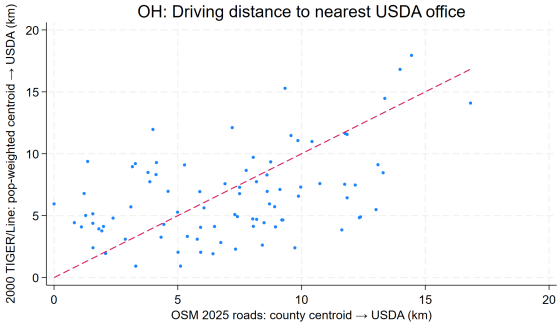
(d) Michigan



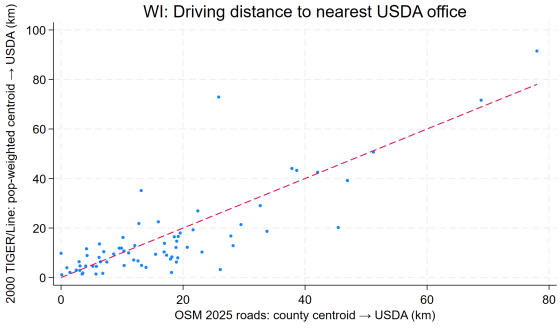
(e) Minnesota



(f) Missouri



(g) Ohio



(h) Wisconsin

Figure 18: Correlation between driving distance from county centroids to the nearest USDA office using the 2025 OpenStreetMap road network and from population-weighted county centroids using the 2000 Census TIGER/Line road network, for eight Midwest states.

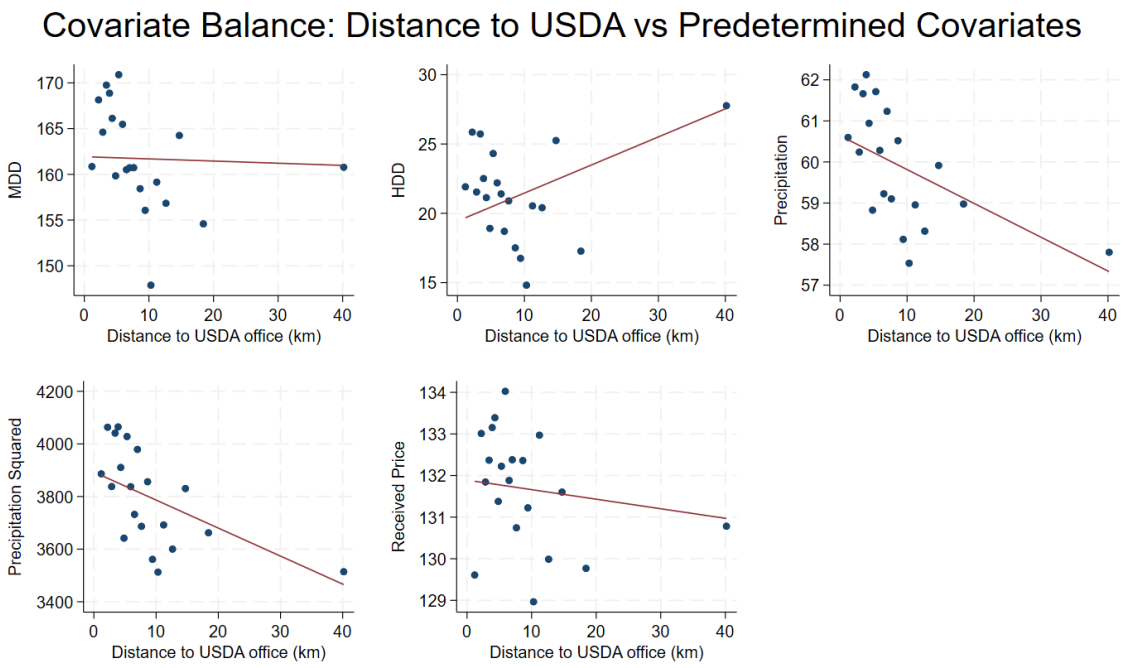


Figure 19: Long-run county–crop averages

$$\text{fracInsured}_{ict} = \pi_1 \,\text{drive_km_to_usda}_i + \mathbf{X}'_{ict}\boldsymbol{\beta}_1 + \gamma_{s(i)} + \delta_t + \varepsilon^{(1)}_{ict}, \tag{4}$$

$$\ln(\text{fracInsured}_{ict}) = \pi_2 \,\text{drive_km_to_usda}_i + \mathbf{X}'_{ict}\boldsymbol{\beta}_2 + \gamma_{s(i)} + \delta_t + \varepsilon^{(2)}_{ict}, \tag{5}$$

$$\text{sublia}_{ict} = \pi_3 \,\text{drive_km_to_usda}_i + \mathbf{X}'_{ict}\boldsymbol{\beta}_3 + \gamma_{s(i)} + \delta_t + \varepsilon^{(3)}_{ict}, \tag{6}$$

$$\ln(\text{sublia}_{ict}) = \pi_4 \,\text{drive_km_to_usda}_i + \mathbf{X}'_{ict}\boldsymbol{\beta}_4 + \gamma_{s(i)} + \delta_t + \varepsilon^{(4)}_{ict}. \tag{7}$$

$$\mathbf{X}_{ict} = (\text{ddayMod_100}_{ict}, \text{ddayHot_100}_{ict}, \text{prec_cm}_{ict}, \text{prec2_cm}_{ict}, \text{price_received}_{ict})\,.$$

Fixed effects are at the state and year level, in order to control for state-level time-invariant factors like institutions and year shocks like macro conditions, and state of technology. With this choice, my identification focuses on the within-state, cross-county variation over time. Standard errors are clustered at the county (FIPS) level, allowing for arbitrary correlation in regression residuals across crops and years within a county.

Table 10 shows that the distance to USDA offices strongly predicts insurance liability but weakly predicts participation rates, indicating that the instrument primarily shifts the intensive margin of insurance coverage rather than the extensive margin. Accordingly, I focus on insurance liability as the endogenous regressor in the baseline IV specification.

Partialled-out first-stage analysis I next studied the relationship between the instrument and the set of endogenous variables after netting out all controls and fixed effects included in the first-stage regression using the partialled-out first-stage plots. Specifically, both the instrument (distance to the

Table 10: First-stage Regression: Insurance Take-up on Distance

	(1) fracInsured	(2) ln(fracInsured)	(3) sublia	(4) ln(sublia)
drive_km_to_usda	-0.0630	-0.00245*	0.000321***	0.00365*
	(0.0488)	(0.00118)	(0.0000861)	(0.00147)
ddayMod_100	-0.0430	-0.000648	-0.000123	0.00186
	(0.0283)	(0.000716)	(0.0000714)	(0.00121)
ddayHot_100	-0.0689***	-0.00246***	0.000271***	0.00386***
	(0.0175)	(0.000433)	(0.0000275)	(0.000507)
prec_cm	-0.105	-0.00402*	-0.000592***	-0.0103***
	(0.0614)	(0.00171)	(0.000127)	(0.00196)
prec2_cm	0.000578	0.0000281*	0.00000479***	0.0000819***
	(0.000418)	(0.0000124)	(0.000000847)	(0.0000130)
price_received	-0.0241***	-0.000690***	-0.0000186***	-0.000715***
	(0.00258)	(0.0000635)	(0.00000355)	(0.0000683)
Observations	29616	29616	29616	29616
R-squared	0.694	0.628	0.519	0.710
First-stage F	1.66	4.34	13.90	6.17

Standard errors in parentheses
State and year fixed effects included.
,
Standard errors clustered at the county (FIPS) level.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

USDA office) and the endogenous variable are residualized with respect to state fixed effects, year fixed effects, and observable covariates, and the plot then relates these residuals to each other. As a result, the slope of the fitted line represents the conditional first-stage effect that identifies the IV, using only within-state, within-year variation. In Figure 20, the slopes are positive and steep for insurance liability (in both levels and logs), indicating a strong conditional association between distance and the intensive margin of coverage. In contrast, the slopes are negative but relatively flat and noisy for insurance participation (in both levels and logs), reflecting a much weaker conditional relationship once fixed effects and controls are removed.

Taken together, the partialled-out first-stage plots show that distance to USDA offices primarily affects the intensive margin of crop insurance coverage rather than the extensive margin. The instrument has substantial explanatory power for insurance liability, as evidenced by both the clear positive slopes and the corresponding strong first-stage F-statistics, while its relationship with participation rates is weak and imprecise. This visual and statistical evidence indicates that the instrument shifts how much insurance farmers carry more than whether they participate at all, supporting the choice of insurance liability as the endogenous regressor in the baseline IV specification and clarifying the economic channel through which distance operates.

Instrumental Variables Diagnosis for Two IVs of driving distance from population-weights county centroids and the number of weather stations Table 11 reports covariate balance tests in which predetermined weather and price variables are regressed on the full set of instruments. While some coefficients are statistically significant, the pattern is not systematic across covariates, and

magnitudes are modest. Given the spatial and historical nature of the instruments, some correlation with long-run agro-climatic conditions is expected. Our identifying assumption is conditional on fixed effects and realized weather controls, which absorb these channels in the main IV specification.

Table 11: Covariance Balance Test for the Instrumental Variables of Driving Distance Between Population-Weighted County Centroids and USDA Offices, and the Number of Weather Stations per County in 1989.

	MDD	HDD	Precipitation	Precipitation squared	Received price
drive_km_to_usda	-0.012	0.157	-0.075***	-9.746***	-0.063
	(0.206)	(0.124)	(0.023)	(2.897)	(0.047)
stations_count1989	-2.095**	-0.750	-0.234	-31.003	-0.438**
	(0.856)	(0.530)	(0.188)	(23.979)	(0.188)
Observations	29002	29002	29002	29002	29002
R-squared	0.171	0.281	0.270	0.279	0.172

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 21 provides a visual assessment of covariate balance by plotting predetermined weather and price variables against the instruments. Both instruments exhibit smooth relationships with long-run climatic conditions, reflecting their spatial and historical nature. Importantly, there is no evidence of sharp discontinuities or nonlinear patterns indicative of endogenous sorting. Combined with fixed effects and controls for realized weather shocks in the main specification, these results support the plausibility of the exclusion restriction.

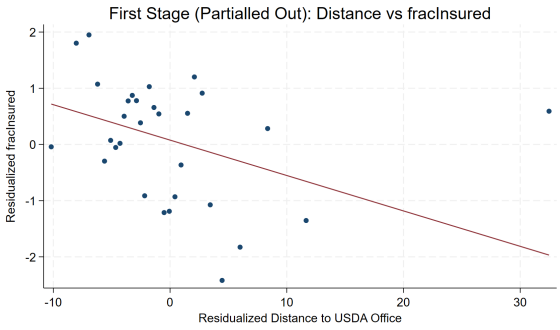
Instrumental variable strength and validity (First-stage of IV regression). I assess the strength of the instrumental variables by estimating first-stage regressions for both endogenous variables: (i) insurance participation, measured by the fraction of planted land that is insured, and (ii) insurance intensity, measured by subsidy per dollar of insured liability. Specifically, I estimate the following first-stage specification:

$$D_{ict}^{(k)} = \boldsymbol{\pi}_k' \mathbf{Z}_i + \mathbf{X}_{ict}' \boldsymbol{\beta}_k + \gamma_{s(i)} + \delta_t + \varepsilon_{ict}^{(k)}, \tag{8}$$

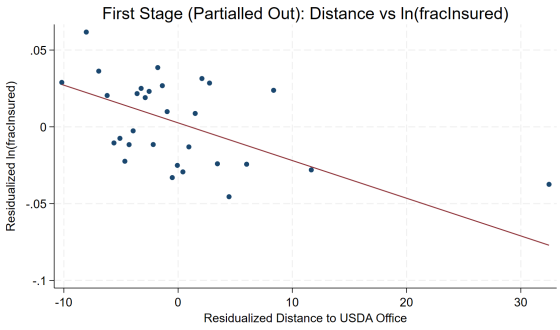
where $D_{ict}^{(k)} \in \{\text{fracInsured}_{ict},\ \ln(\text{fracInsured}_{ict}),\ \text{sublia}_{ict},\ \ln(\text{sublia}_{ict})\}$ denotes the endogenous insurance variable, and $\mathbf{Z}_i = (\text{drive_km_to_usda}_i,\ \text{stations_count1989}_i)$ is the vector of instrumental variables. The vector of controls $\mathbf{X}_{ict}$ includes predetermined weather and price variables:

$$\mathbf{X}_{ict} = (\text{ddayMod_100}_{ict}, \text{ddayHot_100}_{ict}, \text{prec_cm}_{ict}, \text{prec2_cm}_{ict}, \text{price_received}_{ict})\,.$$

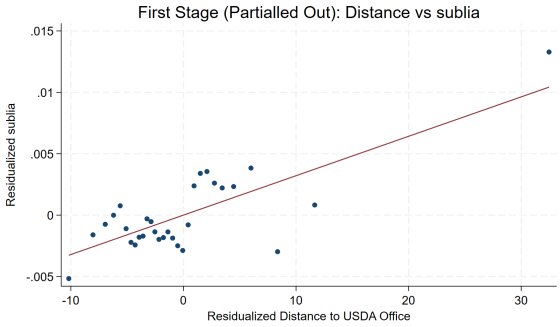
State fixed effects $\gamma_{s(i)}$ absorb time-invariant institutional and geographic differences across states, while year fixed effects δ_t capture common macroeconomic and technological shocks. Identification therefore relies on within-state, cross-county variation over time. Standard errors are clustered at the county (FIPS) level, allowing for arbitrary correlation in regression residuals across crops and years within a county. Instrument relevance is assessed using joint first-stage tests of the null hypothesis that the coefficients on both instruments are equal to zero.



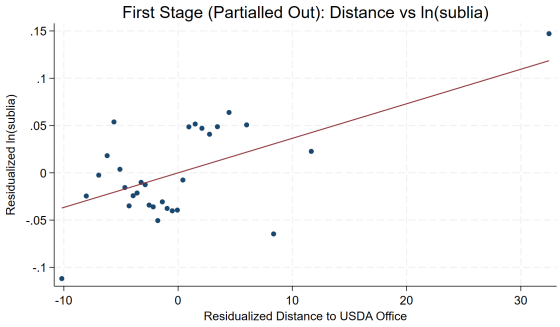
(a) Insurance participation (level)



(b) Insurance participation (log)



(c) Liability (level)



(d) Liability (log)

Figure 20: Partialled-Out First Stage: Distance to USDA Offices Predicts Insurance Liability More Strongly Than Participation

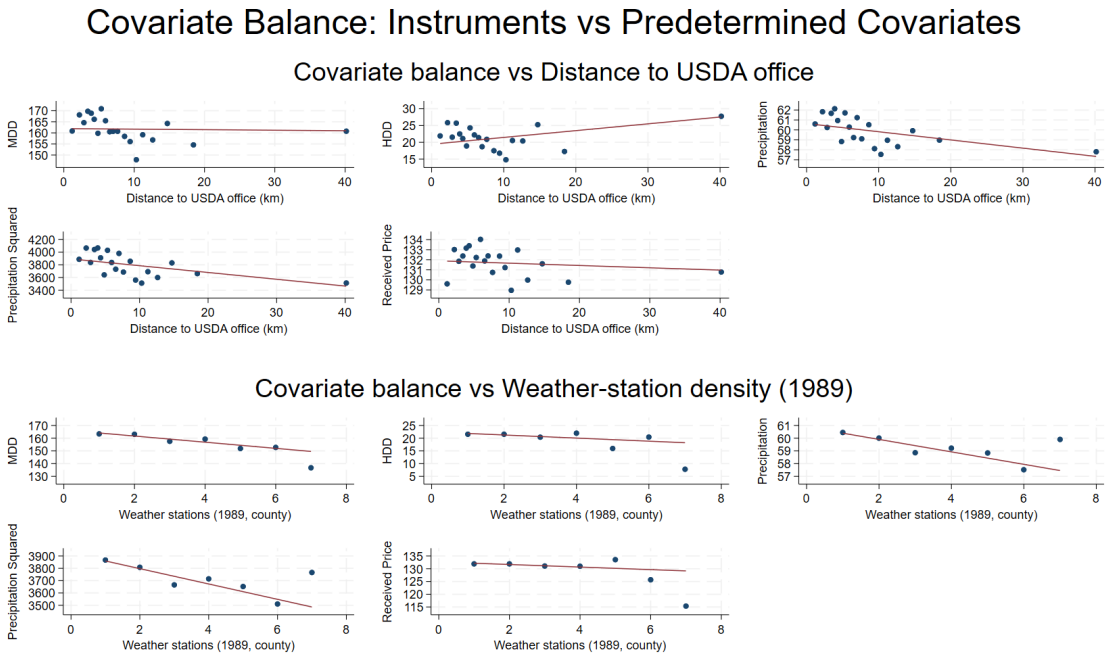


Figure 21: Results of Covariance Balance Tests: Instruments vs Predetermined Covariates

Weak-instrument robustness. Given the moderate first-stage strength for some participation outcomes, all inference is interpreted cautiously, and results are shown to be robust across alternative specifications and insurance intensity measures for which the instruments exhibit stronger relevance.

Table 12 reports the first stage regressions for insurance take-up and liability measures. The instruments exhibit stronger relevance for insurance liability than for fraction of planted land insured, with joint first-stage F-statistics ranging from 3.2 to 7.4 for liability outcomes. In contrast, relevance is weaker for participation outcomes, consistent with the interpretation that distance to USDA offices primarily affects coverage intensity rather than the extensive margin of insurance adoption.

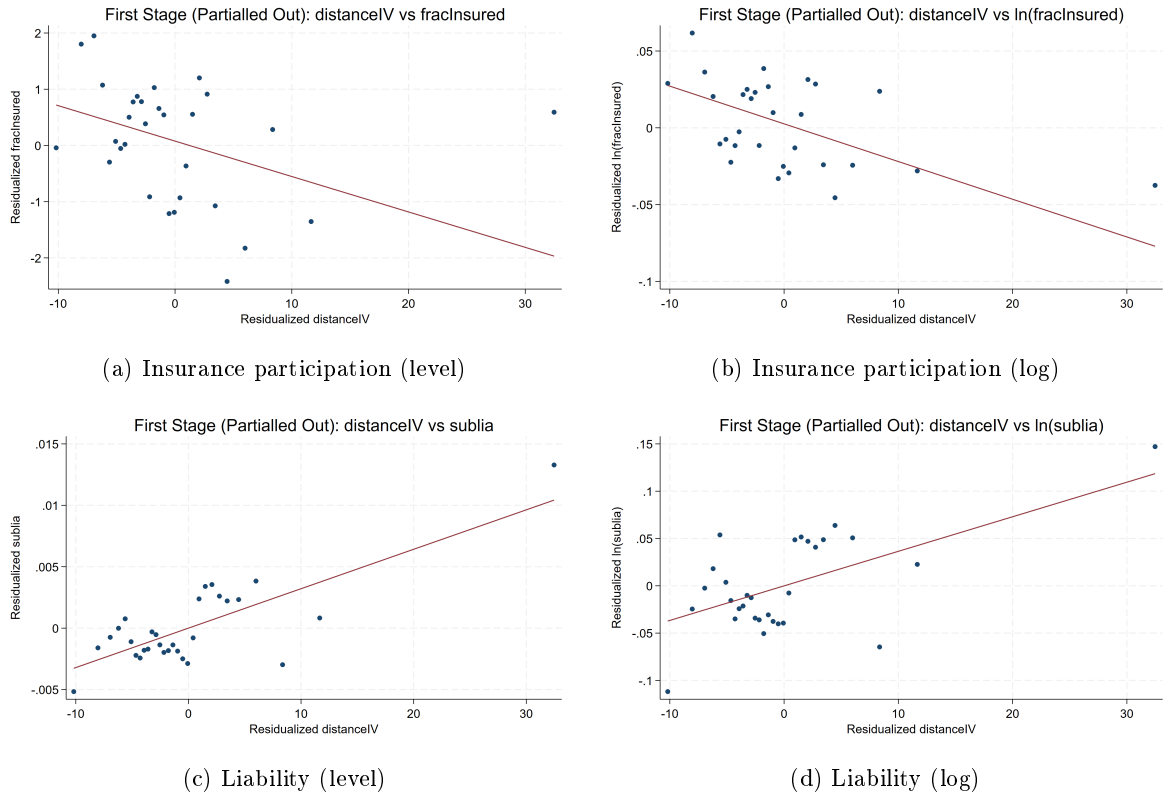
Table 12: First-stage Regressions: Insurance Take-up on Instruments

	(1) fracInsured	(2) ln(fracInsured)	(3) sublia	(4) ln(sublia)
drive_km_to_usda	-0.0565	-0.00237*	0.000313***	0.00353*
	(0.0498)	(0.00119)	(0.0000857)	(0.00145)
stations_count1989	-0.357	-0.00401	0.000549	0.00683
	(0.338)	(0.00746)	(0.000596)	(0.0114)
ddayMod_100	-0.0442	-0.000667	-0.000135	0.00163
	(0.0287)	(0.000728)	(0.0000705)	(0.00121)
ddayHot_100	-0.0702***	-0.00252***	0.000271***	0.00388***
	(0.0177)	(0.000441)	(0.0000279)	(0.000515)
prec_cm	-0.113	-0.00435*	-0.000612***	-0.0105***
	(0.0623)	(0.00174)	(0.000130)	(0.00199)
prec2_cm	0.000625	0.0000307*	0.00000491***	0.0000834***
	(0.000425)	(0.0000126)	(0.000000859)	(0.0000132)
price_received	-0.0244***	-0.000696***	-0.0000185***	-0.000710***
	(0.00261)	(0.0000641)	(0.00000361)	(0.0000693)
Observations	29002	29002	29002	29002
R-squared	0.694	0.628	0.523	0.713
First-stage F (joint IVs)	1.42	2.25	7.36	3.24

Standard errors in parentheses
State and year fixed effects included.
,
Standard errors clustered at the county (FIPS) level.
,
First-stage F is the joint test: drive_km_to_usda = 0 and stations_count1989 = 0.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Partialled-out first-stage analysis. Figures 22 and 23 present partialled-out first-stage relationships between the two instrumental variables—distance to the nearest USDA office and county-level weather-station density in 1989—and insurance outcomes. Across all panels, the fitted lines reflect systematic partial correlations between the instruments and insurance take-up and liability measures after controlling for weather covariates and state and year fixed effects. Distance to USDA offices exhibits a clear negative relationship with insurance participation and a positive relationship with insured liability, consistent with higher administrative or information costs discouraging participation but increasing per-policy liability among remaining participants. Weather-station density displays weaker but directionally consistent relationships, particularly for liability measures, suggesting that greater

availability of local weather information is associated with higher insured exposure. Importantly, these patterns persist after residualizing both the instrument and the endogenous variable, indicating that the observed relationships are not driven by confounding weather conditions or common state-level trends. Together, the figures provide visual evidence supporting the relevance condition for both instruments, while the absence of strong residual correlations with predetermined covariates (shown elsewhere) supports their plausibility as exogenous sources of variation in insurance outcomes.



(a) Insurance participation (level)

(b) Insurance participation (log)

(c) Liability (level)

(d) Liability (log)

Figure 22: First-stage relationships between distance to USDA offices and insurance outcomes. Each panel plots the residualized insurance outcome against the residualized distance to the nearest USDA office after partialling out weather controls and state and year fixed effects. The fitted line depicts the partial first-stage relationship implied by the Frisch–Waugh–Lovell theorem.

Second-stage specification. The causal effect of insurance coverage on outcomes is estimated using two-stage least squares, instrumenting insurance participation and intensity with distance to USDA offices and historical weather-station density.

Let Y_{ict} denote the outcome of interest for county i, crop c, and year t. The second-stage equations are:

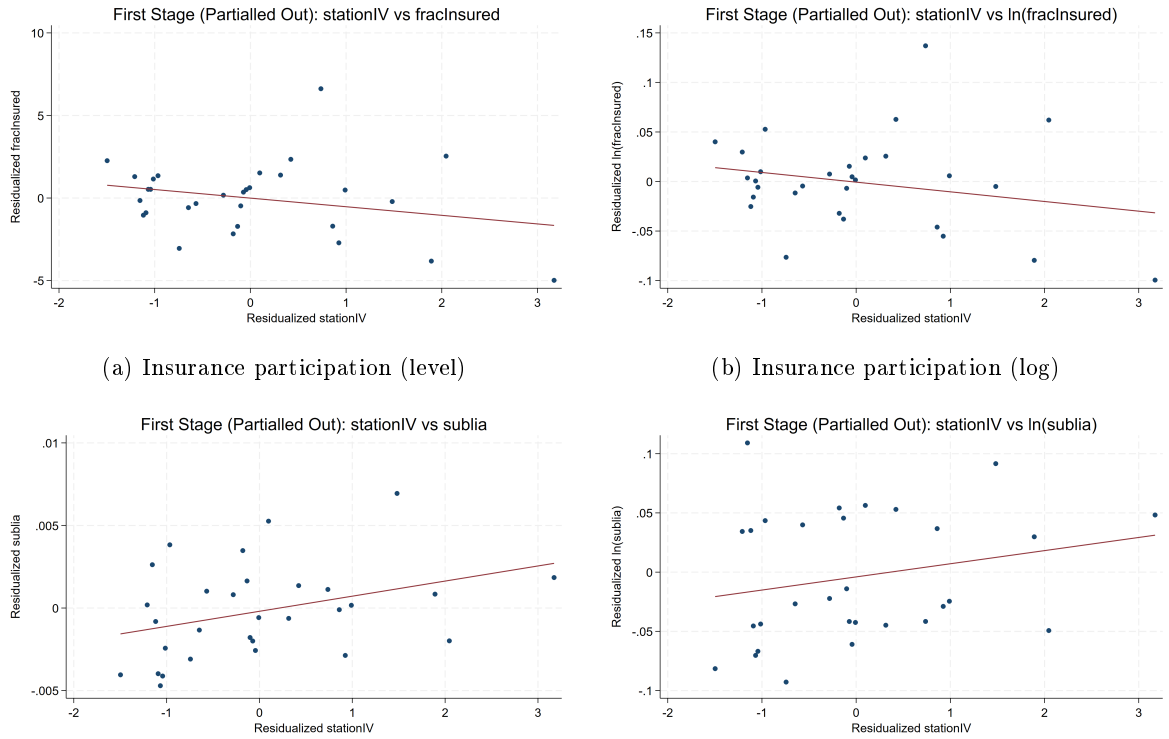
$$Y_{ict} = \beta_1 \widehat{\text{fracInsured}}_{ict} + \mathbf{X}'_{ict}\boldsymbol{\theta}_1 + \gamma_{s(i)} + \delta_t + u^{(1)}_{ict}, \tag{9}$$

$$Y_{ict} = \beta_2 \ln(\widehat{\text{fracInsured}})_{ict} + \mathbf{X}'_{ict}\boldsymbol{\theta}_2 + \gamma_{s(i)} + \delta_t + u^{(2)}_{ict}, \tag{10}$$

$$Y_{ict} = \beta_3 \widehat{\text{sublia}}_{ict} + \mathbf{X}'_{ict}\boldsymbol{\theta}_3 + \gamma_{s(i)} + \delta_t + u^{(3)}_{ict}, \tag{11}$$

$$Y_{ict} = \beta_4 \ln(\widehat{\text{sublia}})_{ict} + \mathbf{X}'_{ict}\boldsymbol{\theta}_4 + \gamma_{s(i)} + \delta_t + u^{(4)}_{ict}. \tag{12}$$

The fitted values $\widehat{\text{fracInsured}}_{ict}$, $\ln(\widehat{\text{fracInsured}})_{ict}$, $\widehat{\text{sublia}}_{ict}$, and $\ln(\widehat{\text{sublia}})_{ict}$ are obtained from the



(a) Insurance participation (level)

(b) Insurance participation (log)

(c) Liability (level)

(d) Liability (log)

Figure 23: First-stage relationships between weather-station density and insurance outcomes. Each panel plots the residualized insurance outcome against the residualized number of weather stations in the county (measured in 1989), after partialling out weather controls and state and year fixed effects. The fitted line represents the partial first-stage relationship implied by the Frisch–Waugh–Lovell theorem.

corresponding first-stage equations (??)–(??), which instrument insurance outcomes using $\text{drive_km_to_usda}_i$ and $\text{stations_count1989}_i$. The control vector $\mathbf{X}_{ict}$ includes predetermined weather and price variables:

$$\mathbf{X}_{ict} = (\text{ddayMod_100}_{ict}, \text{ddayHot_100}_{ict}, \text{prec_cm}_{ict}, \text{prec2_cm}_{ict}, \text{price_received}_{ict}) .$$

State fixed effects $\gamma_{s(i)}$ absorb time-invariant institutional and geographic differences, while year fixed effects δ_t control for common macroeconomic and technological shocks. Identification therefore relies on within-state, cross-county variation over time. Standard errors are clustered at the county (FIPS) level.

8- Population size in the past (Ciccone and hall 1996) as an IV Idea from Rossi-Handberg's class.

9- Immigrant population (card 2001) as an IV Idea from Rossi-Handberg's class.

10- Presence of land grant colleges (Moretti 2004b) as an IV Idea from Rossi-Handberg's class.

4.2 Ex-ante Production Risk

To construct a county–year–crop measure of ex ante production risk while keeping causal identification separate from prediction, I use a two-step "predict-then-estimate" procedure. In the first step, we train a flexible predictive model that maps weather realizations to yield outcomes. In the second step, we use the resulting out-of-sample predicted yield risk as (i) a control for baseline risk, (ii) a source of heterogeneity, or (iii) the basis for a yield-shock index in downstream econometric specifications.

Step 1: Define the prediction target (yield anomaly / loss). Raw yields vary systematically across counties and exhibit long-run trends due to technological change and compositional shifts. To focus the prediction model on short-run production risk, I work with a detrended yield measure. Specifically, for each county-crop pair I estimate a linear time trend in log yields and define the yield anomaly as the residual:

$$\tilde{y}_{itc} \equiv \log(y_{itc}) - \log(\bar{y})_{ic(t)}, \tag{13}$$

where i indexes counties, t years, and c crops. We optionally convert anomalies into a nonnegative "loss" target,

$$\text{loss}_{itc} = \max(0, -\tilde{y}_{itc}) , \tag{14}$$

so that larger values represent worse yield realizations relative to trend.

Step 2: Predicting yield risk using cross-validated machine learning. We predict short-run yield risk from weather using a feed-forward neural network (multi-layer perceptron) implemented in PyTorch. The model maps a parsimonious set of exogenous weather variables—heating degree days (HDD), moderate degree days (MDD), total precipitation, and precipitation squared—to the de-trended yield outcome $\tilde{y}_{itc}$ (or alternatively to the nonnegative loss measure $loss_{itc}$). The objective of this step is not causal inference, but the construction of a flexible risk proxy that captures nonlinear responses of yields to weather realizations.

To avoid mechanically fitting the risk measure to the same observations used in subsequent causal estimation, we generate out-of-fold predictions using grouped cross-validation. Specifically, the data are partitioned by county–crop group, and the prediction model is trained on all but one fold, producing predictions for the held-out fold. Concatenating predictions across folds yields an out-of-sample predicted risk measure $\hat{r}_{itc}$ for every observation in the panel.

Step 3: Using predicted risk in causal estimation. The predicted risk measure $\hat{r}_{itc}$ is then incorporated into econometric models in a separate stage, without altering the underlying identification strategy. First, $\hat{r}_{itc}$ can enter as a baseline risk control to improve estimation precision. Second, it can be used to study heterogeneous effects by interacting insurance enrollment (or coverage intensity) with $\hat{r}_{itc}$, testing whether insurance has larger impacts in higher-risk settings. Third, $\hat{r}_{itc}$ can be used to construct a shock index—defined as realized loss minus predicted loss—for local-projection impulse response analysis. Throughout, the machine-learning model is used solely to summarize ex ante production risk, while identification continues to rely on the chosen econometric design, such as fixed effects, instrumental variables, or policy-induced variation.

4.3 Cost of Insurance Subsidy

In order to assess the full economic costs of these government-subsidized programs, (Hazell 1992) presented his analysis in terms of the ratio of paid indemnities to the nonsubsidized portion of the premium (P) paid by the farmer, termed the producer loss ratio (I/P). This measure is distinct from a conventional or gross loss ratio, which is calculated as the ratio of paid indemnities to total original gross premiums. I present results for both.

5 Baseline causal Effects of Subsidy Shocks

In different models, I gradually add fixed effects for county, crop-specific time trend (or time trend), State-specific time trend, lagged rainfall control (for annual or cropping period).

I measured the impact of farm bills using four measures, all endogenously set, and hence, need instrumental variable for deriving their causal effects. The four factors are: 1) the effective subsidy rates experienced by farmers; 2) the subsidy amount per $100 Liability, both calculated in Section 3; 3) The liability of farmers, from Jisang Yu, A. Smith, and Sumner (2018); and 4) The percent of planted land that has insurance coverage. I used these measures in all sections of the paper.

Econometric Model Specification To study the sensitivity of crop yield to crop insurance subsidies, we specify changes in yield as a dynamic panel model. In this paper, sensitivity refers to the responsiveness of yields (measured in log terms) to a one percentage point change in the measure of intensity of insurance enrollment, conditional on weather interactions and other controls.

I modeled the panel data using two-way clustering as formulated by Cameron, Gelbach, and D. L. Miller (2011). The panel data allows me to correct standard errors for serial correlation and heteroskedasticity by exploiting both the cross-sectional and time series dimensions. Clustering by unit accounts for serial correlation within units over time, while clustering by time accounts for simultaneous correlation across units. These adjustments are nonparametric and they rely on empirical covariances rather than assuming a specific correlation structure.

In modeling, I first followed the basic model of Deschênes and Greenstone (2007). The dependent variable is the yield Y_ijt, for county i, crop j, and year t. We capture the intensity of enrollment Γ_{ijt} with two parameters of subsidy per dollar of liability and percentage of land insured, and have them in the model once at a time. The weather-related variables are the number of heating degree days (HDD), number of moderate degree days (MDD), amount of precipitation in centimeter, and its squared value, all denoted collectively as W_{ijt}. The control variables are yield in the previous period Y_{ijt-1}, expected price EP_{ijt}, crop-specific time trends $Time_{jt}$, and county-crop fixed effects ν_{ij}. Why not year FEs (instead of crop-year FE)? Also, the presence of lagged Y (why not lagged $ln(Y)$?) on the RHS will cause its own set of problems—this was the starting observation for Nerlove developing random effects back in the day.. As Deschênes and Greenstone (2007) put, it is impossible to estimate the effect of the long-run climate averages in a model with county fixed effects, because there is no temporal variation in $\overline{W}_{ic}$. Consequently, I replace the climate variables with annual realizations of weather, W_{ict}.

$$ln(Y_{ijt}) = \beta_0 + \beta_1\Gamma_{ijt} + \beta_2 EP_{ijt} + \beta_3 Y_{ijt-1} + \sum_j \beta_{j4} Time_{jt} + \beta_{j5} W_{ijt} + \beta_{jw6} W_{ijt}\Gamma_{ijt} + \nu_{ij} + u_{ijt} \quad (15)$$

u_{ijt} denotes random errors Just call it an error term with mean zero. "Random" can imply iid, which it is not. Our main coefficients of interest are β_{jw6}, but we have also added Γ_{ijt} and W_{ijt} separately. According to Balli and Sørensen (2013), in a regression with interaction terms, the main terms should always be included Not sure—Pearl and also Angriest and Pischke show that identification by blocking backdoor paths is nonparametrically identified. In other words, if you want E(Y|X,W), you just need to control for X and W—doesn't matter if they're interacted or not, squared, etc. provided both are included somehow.; otherwise, excluding them would cause omitted variable bias.

The county-crop fixed effect captures unobserved heterogeneity across counties for each crop and across crops for each county. The crop-specific time trends capture unobserved heterogeneity across crops for each year. The instrumental variables ?? You haven't discussed them yet and the Γ_{ijt} variable trended upwards ? over time. To prevent detecting a spurious relation Granger and Newbold (1974), we include crop-specific trends in our models to mitigate this potential omitted variable bias.

We also have the expected price of farmers, since we expect farmers' planting decisions (such as input use and effort) to respond to the expected crop prices. The lagged yield controls farm productivity heterogeneity (e.g., soil quality, farmer ability, technology adoption, etc.) that persists over time. It can also proxy the farmer's expectations about future yield. But including the lag of the dependent variable could cause a potential "Nickell bias" (Nickell 1981). Our study covers the span of 25 years and has 1382 unique county-crop panel identifiers. Since the Nickell bias is proportional to the inverse of the number of time periods, we expect the bias to be small, on the order of 4 percent. In addition, our interest is to identify the causal relationship among the crop yield, crop insurance premium subsidy, and a measure of insurance enrollment (subsidy per dollar of liability or percentage of land insured), not the dynamic relationship between the current yield and its lag. The bias in the coefficient of the premium subsidy is less problematic when the covariance of the demeaned premium subsidy variable and the demeaned lagged dependent variable is small.

Endogeneity of Γ A riskier crop or a riskier county has a higher premium rate for a given crop insurance choice (Coble and Barnett 2013). As a result, farmers tend to plant less risky crops and plant less in riskier counties, ceteris paribus. Riskiness is positively correlated with Γ_{ijt} and negatively correlated with Y_{ijt}. Thus, omitting the "riskiness" variable that has a positive correlation with Γ_{ijt}, would cause a downward bias in the OLS estimate for coefficient of Γ_{ijt}. A second source of endogeneity is the dependence of Γ_{ijt} on the choice of coverage level. Although the sign of relationship is not determined. Appendix C has derivations of the sign of both sources of bias. Okay but who cares about the sign of the bias? If it's biased, it's biased.

5.1 Results- Baseline causal Effects of Subsidy Shocks

5.1.1 Subsidy per dollar of liability as the measure of intensity of insurance enrollment

Table 13 presents the results for the case of subsidy per dollar of liability (sublia) as a measure of the intensity of the insurance enrollment. In the OLS specifications (columns 1 and 3), the coefficient on subsidies is positive and significant, suggesting that higher subsidies are associated with higher yields. When weather interactions are introduced (column 3), the results show that the yield effect of subsidies depends on weather conditions. The interaction between subsidies and moderate degree days) is positive and statistically significant, indicating that subsidies amplify yield benefits of moderate heat exposure, which typically reflects favorable growing conditions. By contrast, the interaction with extreme heat is negative, consistent with the idea that subsidies exacerbate farmers' vulnerability to high temperature stress, possibly by discouraging long-term adaptive investments. The interaction with linear precipitation is positive, suggesting that subsidies strengthen the productivity benefits of favorable rainfall. However, the interaction with squared precipitation is negative, pointing to declining marginal returns under excessive rainfall.

The IV estimates (columns 2 and 4), which instrument for subsidies to address endogeneity, tell a more cautious story. The main effect of subsidies turns negative and significant, implying that after accounting for selection bias, subsidies are associated with lower yields on average. This aligns with the concern that subsidies reduce incentives for costly adaptation. Nonetheless, the interaction terms confirm that weather conditions mediate these effects. Subsidies heighten yield benefits under moderate temperatures and sufficient rainfall but increase losses under extreme heat or excessive precipitation variability. This pattern underscores that while subsidies provide short-run insurance against income shocks, they may reduce long-run resilience by altering the incentives for adaptive behavior.

Marginal effects at representative values, MERs (Williams 2012): The post-estimation marginal analysis of model (4) reports the result of one percent point increase in sublia on the yield at different levels of each weather variable, holding other variables as their average values[8]. Figure 24 reports that the effect of sublia on yield is negative across the distribution of hot degree days. While the impact is modest and mostly insignificant at lower quantiles, it becomes larger and statistically significant at higher heat levels, reaching yield losses of about 7–9 percent per additional percentage point of adoption. At the 95th percentile and average of hot degree days, a one percentage point increase in sublia causes 8.6 and 6.8 percent decrease in yield.

[8] Appendix E reports tables of marginal analysis for this and the subsequent measures.

Table 13: Estimation of Equation (1) for subsidy per dollar of liability as endogenous variable.

	(1) *log_yield*	(2) *log_yield*	(3) *log_yield*	(4) *log_yield*
Received price	0.00034***	0.00011	0.00026***	0.00015**
	(0.000035)	(0.000072)	(0.000035)	(0.000077)
Last period yield,	0.00031***	0.00052***	0.00032***	0.00050***
	(0.000056)	(0.000082)	(0.000057)	(0.000094)
sublia	0.25**	-3.14***	5.56***	-6.34*
	(0.11)	(0.91)	(0.73)	(3.50)
MDD	0.0055***	0.0053***	0.00558***	0.0031***
	(0.00015)	(0.00018)	(0.00019)	(0.00059)
HDD	-0.0085***	-0.0088***	-0.0088***	-0.0076***
	(0.00016)	(0.00017)	(0.00022)	(0.00055)
Precipitation	0.013***	0.013***	0.019***	0.020***
	(0.00051)	(0.00058)	(0.00092)	(0.00022)
Squared precipitation	-0.00011***	-0.00011***	-0.00016***	-0.00019***
	(4.1e-06)	(4.6e-06)	(6.9e-06)	(0.000016)
Sublia × HDD			0.0068*	-0.023**
			(0.0036)	(0.0099)
Sublia × MDD			-0.011***	0.052***
			(0.0028)	(0.016)
Sublia × precipitation			-0.14***	-0.17***
			(0.018)	(0.049)
Sublia × sq.precipitation			0.0012***	0.0018***
			(0.00013)	(0.00037)
Corntime	0.014***	0.021***	0.014***	0.018***
	(0.00030)	(0.0018)	(0.00028)	(0.0019)
Soybeantime	0.0079***	0.015***	0.0082***	0.012***
	(0.00035)	(0.0020)	(0.00033)	(0.0022)
Constant	2.97***	3.07***	2.76***	3.27***
	(0.027)	(0.039)	(0.043)	(0.14)
Observations	30,500	28,189	30,500	28,189
R-squared	0.422		0.428	
Number of panelID	1,364	1,291	1,364	1,291

Notes: Robust standard errors in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

County–crop fixed effects included; errors clustered at the county–crop level. Model 1: no IV, no weather interaction; model 2, with IV, no weather interaction; model 3, no IV, with weather interaction; and model 4, with IV, with weather interaction.

Figure 25 reports the marginal effects of sublia on yield at different values of number of moderate degree days. The marginal effect of sublia on yield shifts from negligible at low values of moderate degree days to positive and statistically significant at higher values. At the 90th and 95th percentiles, one percentage point increase in sublia adoption results in a 4.1-4.6 percent increase in yield, and at the 99th percentile the effect rises to about 5.4 percent.

Figure 26 presents the marginal effects of sublia on yields at different levels of precipitation. The results indicate that higher subsidies are consistently associated with lower yields, and the adverse effect intensifies as precipitation increases. At the 1st percentile of precipitation, a one percentage point increase in subsidy reduces yield by about 12.3 percent, whereas at the 99th percentile the reduction

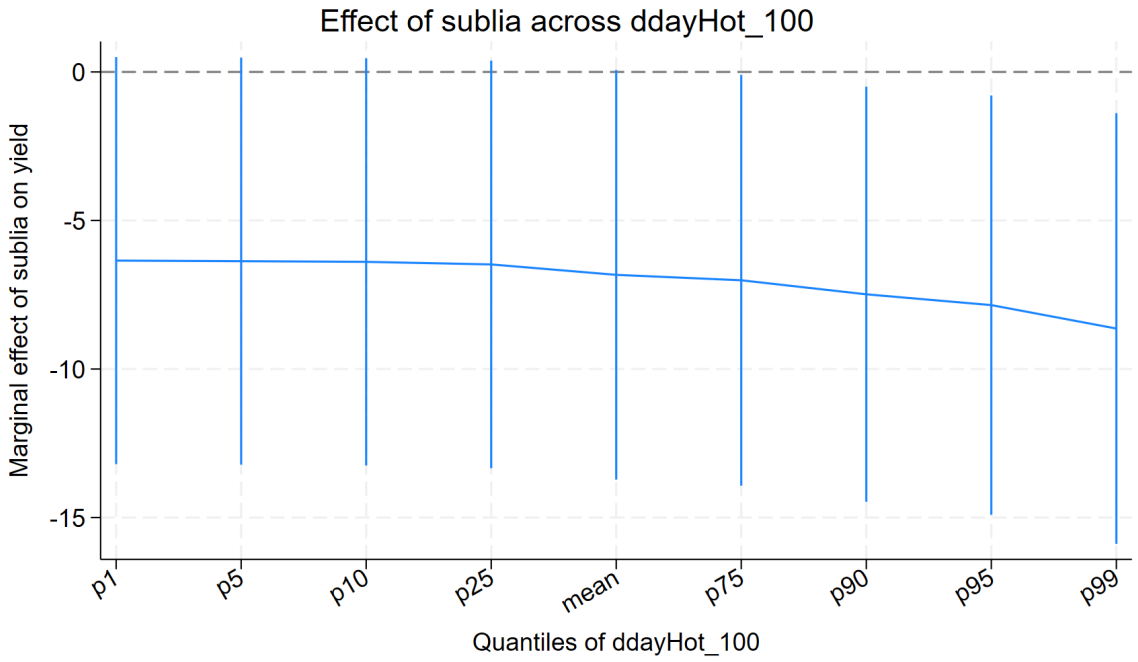


Figure 24: The overall marginal effects of change in sublia across distribution of heating degree days, considering the standalone sublia variable as well as its interaction with weather variable in the regression. All other variables are at their mean values.

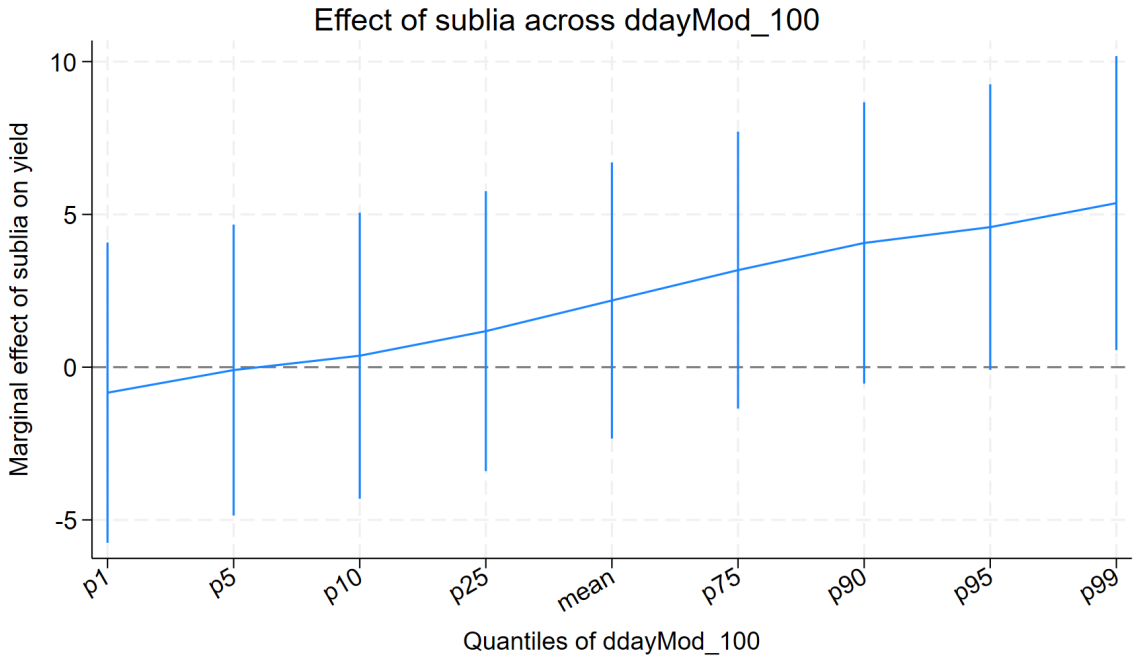


Figure 25: The overall marginal effects of change in sublia across distribution of moderate degree days, considering the standalone sublia variable as well as its interaction with weather variable in the regression. All other variables are at their mean values.

reaches nearly 24.9 percent.

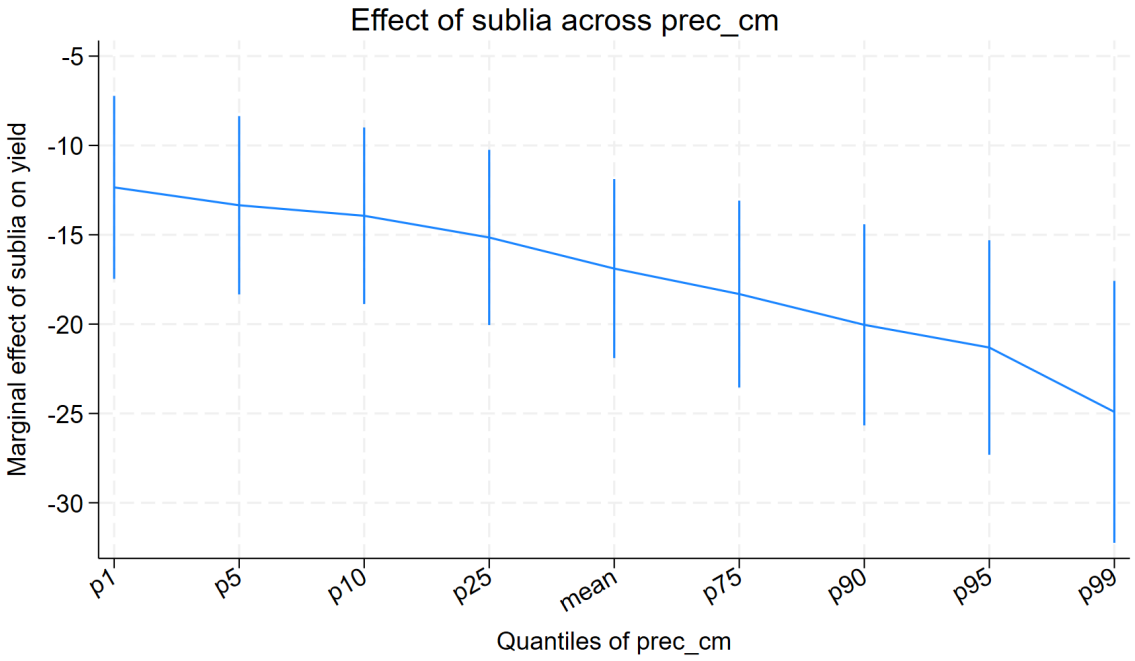


Figure 26: The overall marginal effects of change in sublia across distribution of precipitation, considering the standalone sublia variable as well as its interaction with weather variable in the regression. All other variables are at their mean values.

Figure 27 presents the marginal effects of sublia on crop yield across quantiles of squared precipitation. The results indicate a nonlinear relationship: at lower precipitation squared levels, the marginal effect of sublia on yield is negative but statistically insignificant, suggesting no robust adverse effect in drier or moderately wet conditions. However, as values increase toward the upper tail of the distribution, the marginal effect becomes positive and statistically significant at the 99th percentile, where a one-percentage-point increase in sublia causes a 14.1 percent increase in yield.

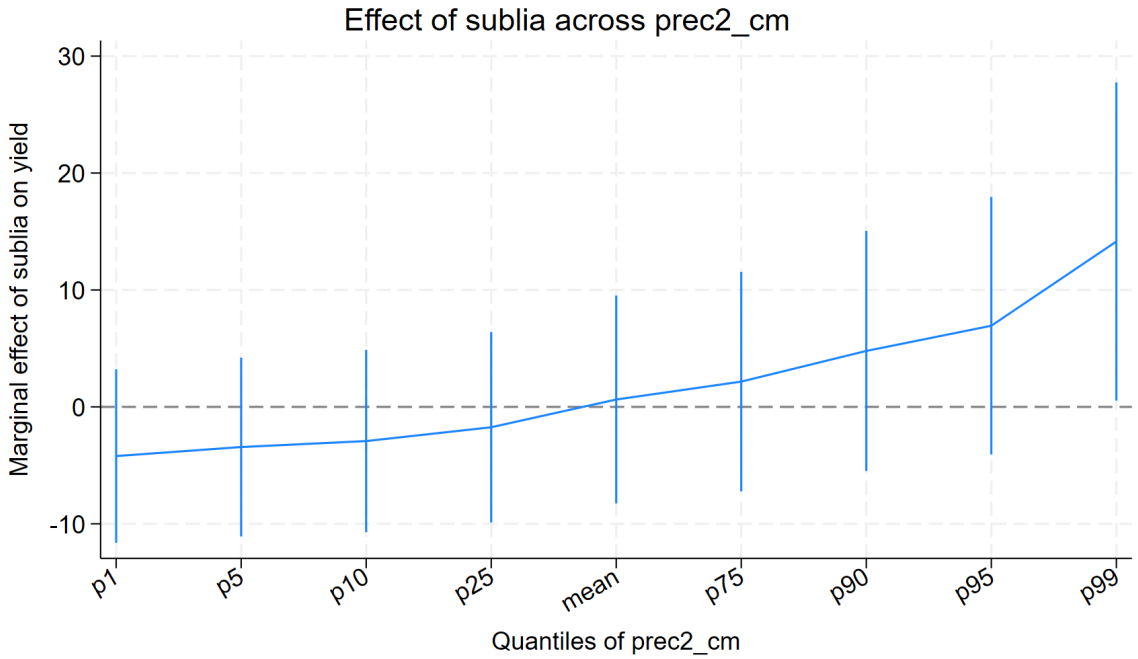


Figure 27: The overall marginal effects of change in sublia across distribution of precipitation squred, considering the standalone sublia variable as well as its interaction with weather variable in the regression. All other variables are at their mean values.

5.1.2 Fraction of planted land insured as the measure of intensity of insurance enrollment

Table 14 shows the results for the case of fracInsured (fraction of land insured) as the measure of the intensity of insurance enrollment. This table presents fixed-effects regressions of log yield on fracInsured, weather variables, and their interactions. Columns (5) and (7) report OLS estimates, while columns (6) and (8) use an IV strategy to address endogeneity of insurance take-up. Columns 1 shows essentially no average association between the insured share of acreage and log yields. When we instrument FracInsured with the rating-based parameter (column 2), the coefficient turns negative and statistically significant, indicating that at average weather conditions a 1 pp increase in insured acreage lowers yields on the order of a few tenths of a percent. This pattern is consistent with selection into more marginal land and/or weakened adaptation incentives when a larger share of acreage is covered. Allowing for heterogeneous effects across the weather distribution (columns 3-4) changes the picture. With interactions, the OLS coefficient becomes positive, but the IV main effect loses precision, while the interaction terms are informative: in the IV specification, the effect of insurance becomes more favorable as hot-degree days increase and less favorable as moderate-temperature days increase; interactions with precipitation are small and not robust. Taken together, the results suggest that a higher insured share is not on average productivity-enhancing, but it reallocates risk across states of nature by penalizing yields in benign conditions while partially offsetting losses in hot years. This heterogeneity reconciles the negative average IV effect with the mixed OLS patterns and is consistent with a portfolio/exposure channel: expanding insured acreage encourages cultivation on riskier margins and dampens baseline adaptation, yet coverage cushions outcomes when heat stress is severe.

Marginal effects at representative values: The post-estimation marginal analysis of model

Table 14: Estimation of Equation (1) for fracInsured as endogenous variable.

	(5) *log_yield*	(6) *log_yield*	(7) *log_yield*	(8) *log_yield*
Last period yield,	0.00032***	0.00026***	0.00033***	-0.00073
	(0.000057)	(0.000057)	(0.00057)	(0.00069)
FracInsured	-0.000075	-0.0039***	0.0061***	0.082
	(0.000071)	(0.0011)	(0.00073)	(0.062)
MDD	0.0054***	0.0055***	0.0058***	0.060**
	(0.00014)	(0.00016)	(0.00025)	(0.029)
HDD	-0.0086***	-0.0083***	-0.0089***	-0.037**
	(0.00016)	(0.00021)	(0.00036)	(0.015)
Precipitation	0.013***	0.016***	0.023***	-0.043
	(0.00052)	(0.0011)	(0.0011)	(0.067)
Squared precipitation	-0.00011***	-0.00013***	-0.00017***	0.00041
	(4.1e-06)	(8.0e-06)	(8.0e-06)	(0.00044)
FracInsured × HDD			2.8e-06	0.00053*
			(4.8e-06)	(0.00029)
FracInsured × MDD			-2.6e-06	-0.00098*
			(3.1e-06)	(0.00054)
FracInsured × precipitation			-0.00016***	0.0015
			(0.000016)	(0.0012)
FracInsured × sq.precipitation			9.8e-07***	-0.000012
			(1.2e-07)	(7.8e-06)
Received price	0.00030***	-0.00017	0.00032***	-0.0040
	(0.000035)	(0.00015)	(0.000036)	(0.0026)
Corntime	0.015***	0.024***	0.014***	0.067**
	(0.00027)	(0.0027)	(0.00027)	(0.031)
Soybeantime	0.0087***	0.018***	0.0086***	0.079**
	(0.00028)	(0.0028)	(0.00027)	(0.039)
Constant	2.98***	3.07***	2.57***	-2.63
	(0.026)	(0.037)	(0.054)	(3.79)
Observations	30,500	28,189	30,500	28,189
R-squared	0.421		0.426	
Number of panelID	1,364	1,291	1,364	1,291

Notes: Robust standard errors in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

County–crop fixed effects included; errors clustered at the county–crop level.

(8) reports the outcome of one percent point in fracInsured on yield at different levels of weather variables, holding other variables as their average values. Figures 28 and 29 present the marginal effects of crop insurance coverage, measured as the share of planted land insured (fracInsured), on yields across the distribution of hot degree days and moderate degree days, respectively. The results suggest that the marginal impact of insurance coverage on yield is positive across all quantiles, but statistically insignificant at conventional levels. Figure 28 shows a relatively flat profile, with modest increases in the effect of insurance as heat exposure intensifies, particularly at the upper tail (95th–99th percentiles).

Figure 30 report the marginal effect of fracInsured on yield across the precipitation distribution. The results show that insurance consistently has a positive effect on yield, with the magnitude of the effect increasing as precipitation increases. At the 1st percentile of precipitation, a one percentage

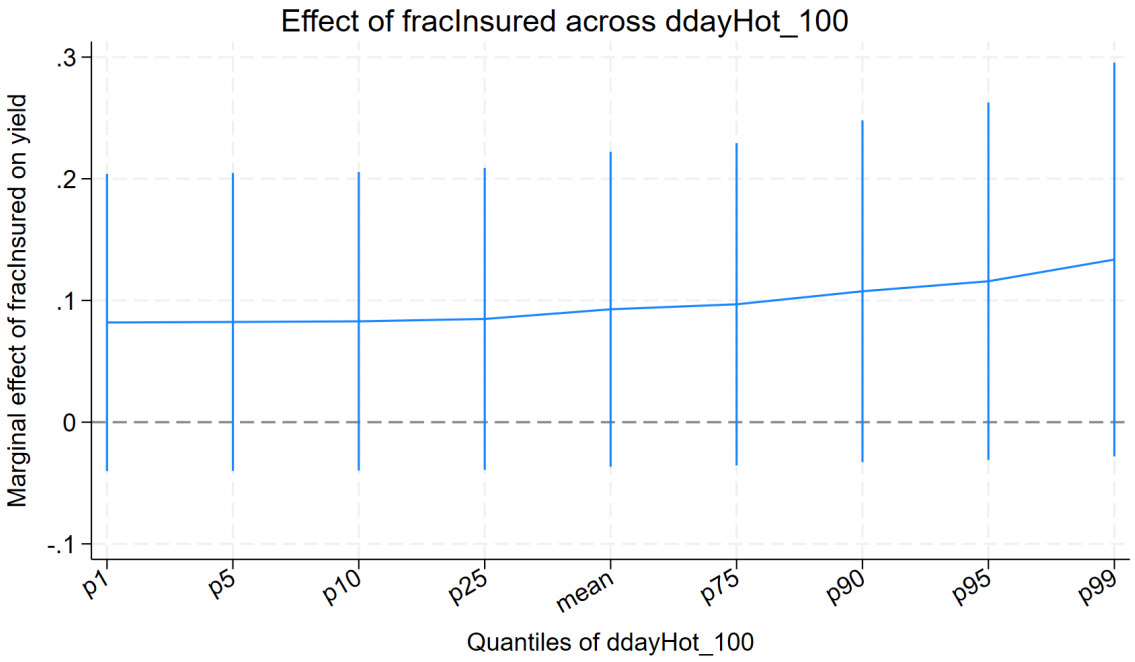


Figure 28: The overall marginal effects of change in fracInsured across distribution of number of heating degree days, considering the standalone fracInsured variable as well as its interaction with weather variable in the regression. All other variables are at their mean values.

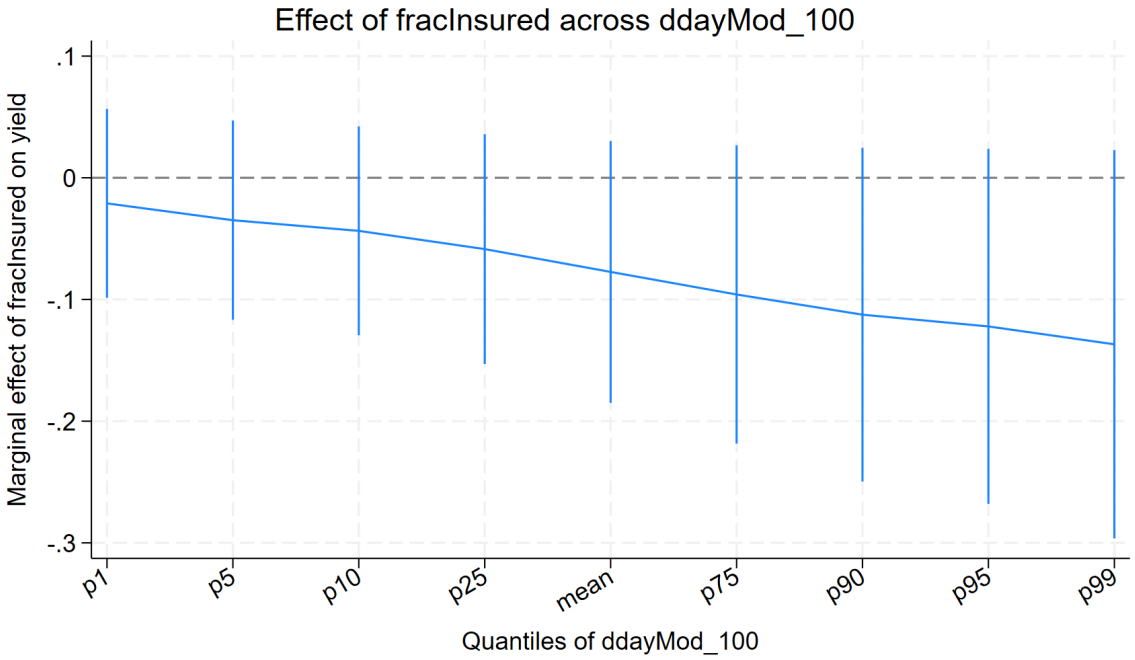


Figure 29: The overall marginal effects of change in fracInsured across distribution of number of moderate degree days, considering the standalone fracInsured variable as well as its interaction with weather variable in the regression. All other variables are at their mean values.

point increase in insurance coverage raises yield by about 0.13 percent, and this effect strengthens to about 0.24 percent at the 99th percentile. The figure suggests that the yield-enhancing role of insurance is more pronounced under wetter conditions.

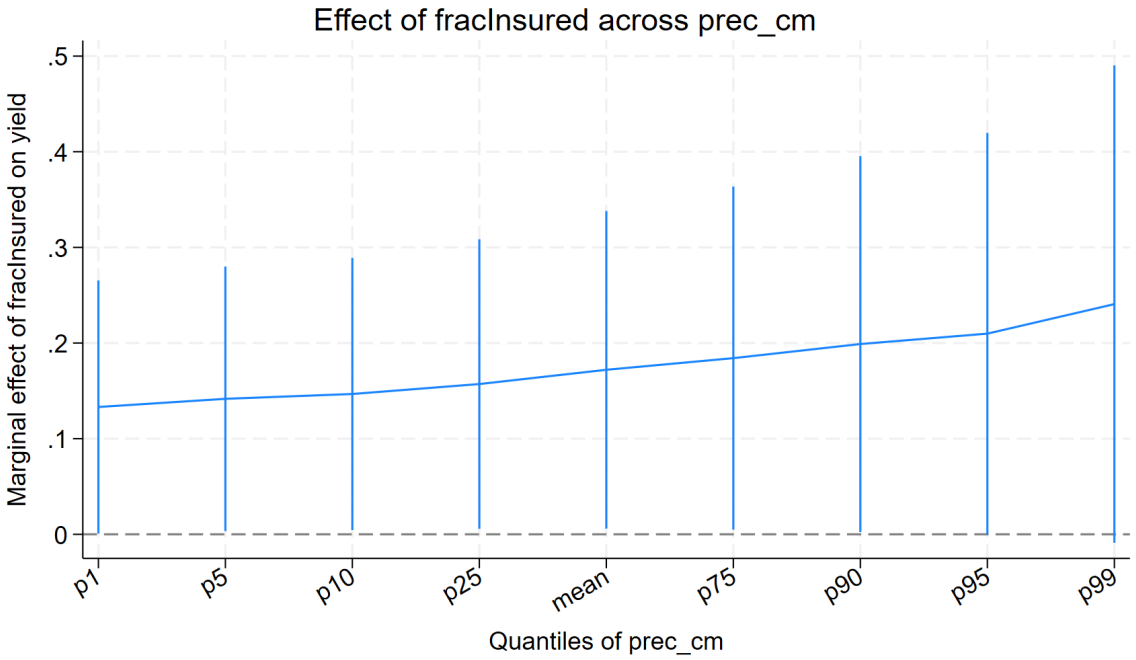


Figure 30: The overall marginal effects of change in fracInsured across distribution of amount of precipitation, considering the standalone fracInsured variable as well as its interaction with weather variable in the regression. All other variables are at their mean values.

Estimates of the marginal effect of fracInsured at different percentiles of squared precipitation in Figure 31 show that the marginal effect of insurance on yields is slightly positive at lower quantiles of squared precipitation, but not statistically significant. As precipitation squared increases, the effect decreases and turns slightly negative at the upper quantiles (p95–p99), suggesting diminishing or even adverse returns to insurance under very high levels of precipitation variability. The confidence intervals overlap zero across the entire range, indicating no robust evidence of significant effects.

Taken together, these results highlight a dual role of insurance subsidies. While they buffer yields under moderate stress and average rainfall, they may simultaneously exacerbate sensitivity to extreme weather, consistent with the hypothesis that subsidized insurance reduces incentives for long-term resilience investments.

5.1.3 Performance of the IV in Estimations

To check the relevance of the instrument for model (2), we conducted the Kleibergen-Paap rk LM statistic of under-identification, which reports the value of 69.457, with Chi-sq(1) P-val = 0.0000. This highly significant result rejects the null hypothesis of under-identification. Second, we checked the strength of the instrument and calculated the Kleibergen-Paap rk Wald F statistic, which results in values 413.118 and 82.474, respectively. Considering the simulated thresholds of maximum tolerated size distortion for F-statistic developed by Stock and Yogo (2002), our F statistics are well above the 10% Stock-Yogo threshold of 16.38. This shows that our instrument is strong, even in the strictest

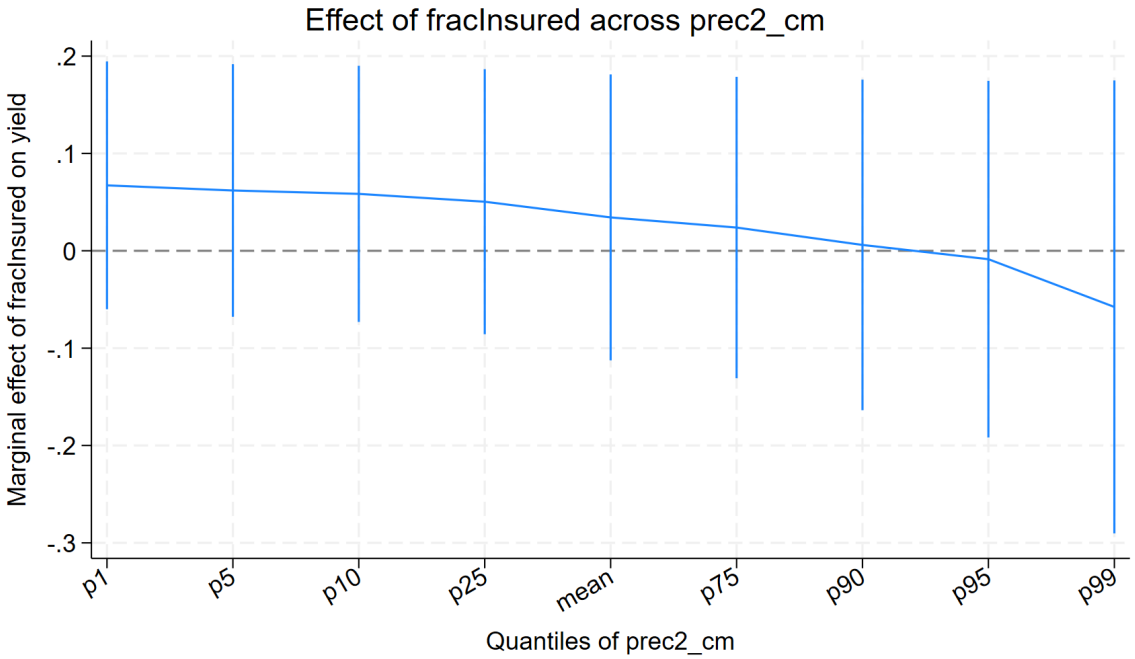


Figure 31: The overall marginal effects of change in fracInsured across distribution precipitation squared, considering the standalone fracInsured variable as well as its interaction with weather variable in the regression. All other variables are at their mean values.

category of Stock and Stock and Yogo (2002). In model (4), the under-identification test (Kleibergen-Paap rk LM statistic) reports value of 72.737 with Chi-sq(1) P-val = 0.0000. Again, this statistic rejects the null hypothesis of under-identification when including the weather variable interactions in the model. The Kleibergen-Paap rk Wald F statistic for the strength of instrument takes value of 16.347 is almost the 10% Stock-Yoko threshold of 16.38. The instrument is adequate but only just strong enough under the strictest criterion, leaving them robust to weak-IV inference.

In model (6), the under-identification test (Kleibergen-Paap rk LM statistic) reports value of 94.095 with Chi-sq(1) P-val = 0.0000, rejecting the null hypothesis of irrelevance of instrument. The Kleibergen-Paap rk Wald F statistic for weak instrument had value od 100.058, well above the the 10% Stock-Yogo threshold of 16.38. This proves the strength of the instrument. In model (8), the under-identification test (Kleibergen-Paap rk LM statistic) reports 3.897 with Chi-sq(1) P-val = 0.0484. This low static shows weaker results, but it still rejects the null hypothesis of irrelevance at the 5% level. But the Kleibergen-Paap rk Wald F statistic for weak instrument has value of 0.801, which is far below of even the 25% threshold (loosest threshold) of Stock-Yoko that is 5.53. This means that the instrument is weak for the specification of model (8): fraction land insured as the endogenous variable, and interaction of fraction insured with the weather variables included in the regression.

5.2 Sensitivity Analysis

Table 15 reports the IV–GMM estimates for corn when using sublia as a measure of intensity of insurance enrollment (columns 9 and 10), and fracInsured as a measure of intensity of enrollment

(columns 11 and 12). Table 15 demonstrates a robust negative association between intensity of insurance enrollment and yield outcomes, consistent with the hypothesis that subsidies may reduce incentives for costly adaptive investments. In the baseline specification without interactions (columns 9 and 11), the coefficients on sublia and fracInsured are negative and statistically significant, with a strong first stage (Kleibergen–Paap F = 63.402), giving confidence in the validity of the estimates. When interactions with weather variables are introduced (columns 10 and 12), the results show that subsidies amplify losses under extreme heat (negative $subliaHDD$ and $fracInsuredHDD$) while providing modest resilience under moderate heat (positive $subliaMDD$ and $fracInsuredMDD$). The precipitation terms reveal a convex relationship, where subsidies are associated with lower yields at moderate rainfall levels (negative $subliaprecipitation$) but this penalty diminishes at higher precipitation (positive $subliaprec^2$). All coefficients in column 12 are insignificant.

Table 16 shows the IV_GMM results for soybeans. In this case, the evidence is weaker and less consistent, reflecting both crop-specific agronomic differences and weaker first-stage identification. In the baseline IV specification, enrollment intensities again display a negative association with yield in models 13-15, but the instrument strength is marginal in models 13, 14 and 16 (Kleibergen–Paap F = 6.409 in model 13, 3.480 in model 15, 11.717 in model 15 and 0.526 in model 16), making these estimates a case for a weak IV. When interactions are added, the coefficients on subsidy-weather terms are small and statistically insignificant, suggesting that insured soybean yields do not exhibit the same heightened sensitivity to heat and precipitation as observed for corn. Instead, the positive direct effect of insured acreage (fracInsured) in model 16 may reflect management choices or residual selection into coverage rather than a robust weather interaction channel. Overall, while the soybean regressions preserve the expected signs on weather controls (moderate heat positive, extreme heat negative, precipitation positive, precipitation squared negative) the insurance interactions provide little systematic evidence. These results suggest that the subsidy-weather sensitivity mechanism is more pronounced in corn production, while soybean responses remain comparatively muted.

Comparing the two crops highlights an important asymmetry in the interaction between crop insurance subsidies and weather risk. Corn yields show a clear and robust pattern whereby subsidies increase vulnerability to extreme heat and excessive rainfall but provide some protection under moderate heat. By contrast, soybean yields display weaker and less systematic responses, with interaction terms generally imprecise and instrument strength considerably lower. This divergence likely reflects both agronomic differences and differences in insurance participation patterns across crops, as corn is more heat- and water-sensitive than soybeans (Licht and Archontoulis 2017). The results therefore suggest that the disincentive-to-adapt channel documented in the broader literature is particularly relevant for corn, while the effects for soybeans are more modest and less conclusive.

5.3 Counterfactual Cost of Subsidized Insurance for Counties Exposed to Extreme Heat

In this section, we calculate the financial burden of excessive sensitivity of crop yield for counties located in hot areas. To do so, we set the crop yield of counties exposed to extreme heat (Y^h) to the average crop yield of counties exposed to 0–50% extreme heat (Y^l), and compute the crop-specific cost of the shortfall for counties located in hot regions in each year:

Table 15: Results for crop of corn, for models with and without weather interaction.

	(9) *log_yield*	(10) *log_yield*	(11) *log_yield*	(12) *log_yield*
Sublia	-3.02***	-1.97		
	(1.13)	(4.28)		
Sublia × HDD		-0.044***		
		(0.016)		
Sublia × sq.precipitation		0.0024***		
		(0.00044)		
Sublia × MDD		0.053*		
		(0.028)		
Sublia × precipitation		-0.27***		
		(0.057)		
Last period yield,	0.00047***	0.00041***	0.00022***	0.0012
	(0.000080)	(0.000087)	(0.000067)	(0.00099)
MDD	0.0050***	0.0024**	0.0053***	-0.017
	(0.00030)	(0.0010)	(0.00024)	(0.033)
HDD	-0.0095***	-0.0069***	-0.0091***	-0.0032
	(0.00022)	(0.00080)	(0.00029)	(0.021)
Precipitation	0.017***	0.029***	0.019***	0.087
	(0.00085)	(0.0028)	(0.0013)	(0.11)
Sq. precipitation	-0.00015***	-0.00025***	-0.00016***	-0.00041
	(6.6e-06)	(2.1e-05)	(9.1e-06)	(0.00061)
Received price	-0.0011***	-0.00088**	-0.0010***	0.0037
	(0.00026)	(0.00035)	(0.00022)	(0.0055)
Corntime	0.023***	0.019***	0.024***	-0.016
	(0.0025)	(0.0032)	(0.0030)	(0.049)
FracInsured			-0.0034***	-0.0015
			(0.0012)	(0.019)
FracInsured × HDD				-0.00018
				(0.00040)
FracInsured × sq.precipitation				4.6e-06
				(0.000011)
FracInsured × MDD				0.00046
				(0.00063)
FracInsured × precipitation				-0.0012
				(0.0020)
Constant	3.71***	3.74***	3.67***	2.93**
	(0.065)	(0.18)	(0.049)	(1.33)
Observations	14,281	14,281	14,281	14,281
Number of panelID	653	653	653	653

Notes: Robust standard errors in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

County–crop fixed effects included; errors clustered at the county–crop level.

$$C_{jt}^{F} = \sum_{i} (Y_{jit}^{l} - Y_{jit}^{h}) \times A_{jit}^{h} \times EP_{jit} \tag{16}$$

Here, A_{jit}^{h} is the area planted in counties exposed to extreme heat, and summation is over counties.

Table 16: Yield regressions (dependent variable: log(yield))

	(13) *log_yield*	(14) *log_yield*	(15) *log_yield*	(16) *log_yield*
Sublia	-15.9**	-67.8		
	(6.85)	(43.9)		
Sublia × HDD		-0.022		
		(0.028)		
Sublia × sq.precipitation		-0.00077		
		(0.0019)		
Sublia × MDD		0.19*		
		(0.11)		
Sublia × precipitation		0.24		
		(0.30)		
Last period yield,	0.00049	0.00027	0.0018***	-0.00080
	(0.00031)	(0.00049)	(0.00059)	(0.00081)
MDD	0.0043***	-0.0034	0.0051***	0.0080
	(0.00054)	(0.0048)	(0.00026)	(0.019)
HDD	-0.0085***	-0.0095**	-0.0061***	0.048
	(0.00061)	(0.0037)	(0.00047)	(0.054)
Precipitation	0.012***	0.0032	0.021***	0.13
	(0.0013)	(0.012)	(0.0043)	(0.10)
Squared precipitation	-0.000096***	-0.000057	-0.00016***	-0.00054
	(9.0e-06)	(0.000069)	(0.000032)	(0.00070)
Received price	-0.00012	-0.00029	-0.0011*	0.0033*
	(0.00026)	(0.00043)	(0.00059)	(0.0017)
Soybeantime	0.043***	0.063*	0.041***	-0.036*
	(0.015)	(0.034)	(0.012)	(0.021)
FracInsured			-0.014***	0.12**
			(0.0050)	(0.060)
FracInsured × HDD				-0.00077
				(0.00080)
FracInsured × sq.precipitation				5.8e-06
				(0.000012)
FracInsured × MDD				-0.00080
				(0.00031)
FracInsured × precipitation				-0.0017
				(0.0017)
Constant	2.85***	4.82***	2.72***	-5.13
	(0.20)	(1.52)	(0.12)	(3.60)
Observations	13,908	13,908	13,908	13,908
Number of panelID	638	638	638	638

Notes: Robust standard errors in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

County–crop fixed effects included; errors clustered at the county–crop level.

6 Evidence on Channels for Change in Cop Yield

7 Decomposition of Channels of Change in Crop Yield

In this section, I use instrumental variables to causally identify the two channels for change in crop yield as illustrated in Section 2.3. I estimate a two-stage least squares (2SLS) model with multiple

endogenous regressors. Rather than estimating separate instrumental variables regressions, I specify a single structural equation in which all endogenous variables are instrumented jointly using the full set of excluded instruments.

Structural equation. Let y_{it} denote the outcome of interest, $x_{1,it}$ and $x_{2,it}$ two endogenous regressors, w_{it} a vector of exogenous controls, and u_{it} an error term. The structural (second-stage) equation is given by:

$$y_{it} = \beta_1 x_{1,it} + \beta_2 x_{2,it} + w_{it}'\gamma + u_{it}. \tag{17}$$

First-stage equations. Identification relies on a set of excluded instruments $z_{1,it}$ and $z_{2,it}$. In the 2SLS framework, each endogenous regressor is projected onto the full set of instruments and exogenous controls. Accordingly, we estimate one first-stage regression for each endogenous variable:

$$x_{1,it} = \pi_{11} z_{1,it} + \pi_{12} z_{2,it} + w_{it}'\delta_1 + v_{1,it}, \tag{18}$$

$$x_{2,it} = \pi_{21} z_{1,it} + \pi_{22} z_{2,it} + w_{it}'\delta_2 + v_{2,it}. \tag{19}$$

Although $z_{1,it}$ and $z_{2,it}$ may be motivated by their relevance for specific endogenous regressors, both instruments are included in each first-stage equation. This allows the data to determine the partial correlations between instruments and endogenous variables, and ensures consistency under joint endogeneity.

Second stage. The fitted values from the first-stage regressions, $\hat{x}_{1,it}$ and $\hat{x}_{2,it}$, are then used to estimate the structural relationship:

$$y_{it} = \beta_1 \hat{x}_{1,it} + \beta_2 \hat{x}_{2,it} + w_{it}'\gamma + \varepsilon_{it}. \tag{20}$$

Standard errors are computed using robust (or clustered) variance estimators. Instrument relevance is assessed using first-stage statistics, and overidentifying restrictions are tested where applicable.

8 Spatial Regression Model

To study the impact of crop insurance enrollment on agricultural yields, spatial modeling is essential because the underlying economic processes exhibit substantive spatial dependence. Farmers operate in spatially connected environments where decisions, risks, and outcomes spill across boundaries: adoption of insurance can influence neighboring farmers through peer effects, information exchange, or shared perceptions of risk. Moreover, yields are the equilibrium outcome of spatial interaction processes, including pest pressure, weather patterns, soil characteristics, and local management norms; factors that are documented in Meena et al. (2024) read Jan 21, 2026. Ignoring these channels leads to biased estimates because spatial correlation in both the treatment (insurance uptake) and the outcome (yield) violates the independence assumptions of standard regression approaches. Additionally, some spatial dependence arises from non-behavioral sources, such as data aggregation across counties or regions that share ecological and climatic gradients. For these reasons, spatial regression models (like the spatial lag and spatial autoregressive models) are necessary to correctly identify the impact of insurance enrollment on yield while accounting for the spatial structure of decisions and outcomes

[9].

Ignoring spatial lags in the relationship between crop insurance enrollment and agricultural yields effectively means ignoring the substantive spatial interactions that shape farm outcomes. Farmers' yields and insurance decisions are spatially correlated because they face shared weather shocks, soil characteristics, pest pressures, and information networks. If a model fails to incorporate this spatial dependence, it creates an omitted variable problem: the influence of neighboring counties' insurance uptake or yields is left out of the specification even though it systematically affects outcomes. As a result, ordinary least squares estimates become biased and inconsistent, since the spatially lagged variables are correlated with the error term. This can lead to incorrect conclusions—incorrect coefficient magnitudes, wrong signs, misleading standard errors, spurious statistical significance, and poor model fit. For policy questions, such as assessing whether insurance promotes or distorts yield outcomes, these errors can substantially misguide interpretation and policy design.

In the crop insurance–yield relationship, ignoring spatial error correlation does not bias the OLS coefficient on insurance enrollment, but it does make the estimator inefficient. Inefficiency means that although OLS still delivers the correct average effect, it fails to use all the information in the data because the residuals contain spatial structure that the model ignores. As a result, the estimated standard errors are too small or too large, confidence intervals are misleading, and statistical significance tests become unreliable. In practice, this means we we think insurance has a precisely estimated effect on yield when in fact the true uncertainty is much larger due to spatially correlated shocks (such as common weather events or pest outbreaks) that affect clusters of neighboring counties. Thus, inefficiency undermines inference even when the point estimate itself is correct.

Spatial Cross–Regressive (SLX) model. Following the Spatial Cross–Regressive (SLX) model of Florax and Folmer (1992) and its modern treatment in Halleck Vega and Elhorst (2015), I model spatial spillovers by allowing only the covariates, but not the dependent variable, to be spatially lagged. Specifically, I estimate

$$y_{ict} = X_{ict}\beta + WX_{ict}\theta + \alpha_i + \delta_t + \varepsilon_{it}. \tag{21}$$

where y_{ict} denotes crop yield in county i, crop c and year t, X_{ict} is a vector of county-level covariates including insurance adoption, weather conditions, and prices, and W is a row-standardized spatial weights matrix. The term WX_{ict} captures spatial spillovers arising from neighboring counties' covariates. States fixed effects α_i absorb time-invariant heterogeneity across states, while year fixed effects δ_t control for common aggregate shocks. Because the instrumental variables are time-invariant, specifications with county or county-by-crop fixed effects would absorb the excluded instruments and are therefore not identified. I therefore use state and year fixed effects and cluster standard errors at the county level in both OLS and IV specifications all along.

This specification allows for flexible local spillovers in observable determinants of yields while avoiding the strong equilibrium assumptions and potential endogeneity associated with spatial lag or spatial error models. As emphasized by Halleck Vega and Elhorst (2015), the SLX model provides a

[9] I appreciate Zhuyin Li and Pedro Amal at the Center for Spatial data Science of UChicago for our discussions. I thank Professor Rossi-Hansberg for letting me sit in his class on Spatial Economics

transparent and robust framework for identifying spatial interactions when spillovers operate through covariates rather than outcomes.

Spatial Durbin model. For the robustness check, I also modeled the spatial propagation using the spatial Durbin model. The Spatial Durbin Model allows for endogenous spatial dependence through the spatial lag of the dependent variable while simultaneously modeling spillovers in covariates (LeSage and Pace 2009). Let the spatially lagged dependent variable be Wy, and spatially lagged explanatory variables be WX. Then I have:

$$y_{ict} = \rho W y_{ict} + X_{ict}\beta + WX_{ict}\theta + +\alpha_i + \delta_t + \nu_{it}. \tag{22}$$

where ρ captures endogenous spatial dependence, and all other terms are similar to the SLX model.

Weight Matrix I calculate time-invariant "internal-only" mid west weights, meaning that I only used counties in the study area of 8 states for the calculation. With this choice, I modeled the spillovers only within the study region. I modeled the spatial dependence using a contiguity-based spatial weights matrix. Let W denote an $N \times N$ spatial weights matrix constructed at the county level using "Queen contiguity", where $w_{ij} = 1$ if counties i and j share a common boundary or vertex, and $w_{ij} = 0$ otherwise. The matrix is row-standardized such that each row sums to one,

$$\tilde{w}_{ij} = \frac{w_{ij}}{\sum_j w_{ij}}, \tag{23}$$

so that spatial lags can be interpreted as weighted averages of neighboring outcomes. Because each county borders only a small number of neighbors, the resulting weights matrix is sparse, with the vast majority of entries equal to zero. This sparsity substantially reduces computational burden and is standard in spatial econometric applications using large geographic units such as U.S. counties.

As an alternative specification, I also modeled the spatial dependence using a "Rook contiguity" spatial weights matrix. If counties i and j share a common boundary (edge), the corresponding weight takes value of 1 and 0 otherwise. Compared to Queen contiguity, this definition imposes a stricter notion of neighborhood by excluding corner-touching units. I row-standardized this matrix as well, using the same method for Queen weight matrix.

Spatial Panel Data Setup I first organize the data set at the county-crop-year level and ensure that geographic identifiers are consistent across the outcome/covariate file and the spatial weights. The spatial weights matrix W is time-invariant at the county level, row-standardized, and applied year-by-year to compute spatial lags of the regressors, WX_{ict}, by multiplying W with the county-level covariates within each year and crop, using the exact county ordering implied by W. Finally, I merge these spatially lagged covariates back to the long panel and estimate the SLX specification with state and year fixed effects and standard errors clustered at the county-crop level.

Spatial Regression Results I estimate the equations 21 and 22 for $y_{it} \in (\text{yield}_{it}, \ln(\text{yield}_{it}))$ and $\mathbf{X}_{it}$ contains the weather controls and two measures of enrollment in insurance; fraction of planted acres insured (level and log), and insurance liability per acre. I start with reporting simple OLS results, absent spatial effects and instrumental variables, then report results for 2sls-IV and spatial models with and without IV. For the IV-method, I report results in two cases: 1) just identified: two endogenous

variables - two IVs, and 2) over-identified: one endogenous variable - two IVs, one at a time, without including the other endogenous variable in the model.

Table 17: Simple OLS results in absence of IV or spatial correlations

	(1) yield	(2) ln(yield)	(3) yield (ln endog)	(4) ln(yield) (ln endog)
lnfracInsured			-3.8479***	-0.0440***
			(-9.8184)	(-9.5809)
lnsublia			-14.635***	-0.1678***
			(-25.435)	(-24.011)
fracInsured	-0.0629***	-0.0007***		
	(-4.1551)	(-3.9658)		
sublia	-226.82***	-2.3499***		
	(-20.305)	(-17.468)		
ddayMod_100	0.3126***	0.0038***	0.3674***	0.0044***
	(17.331)	(13.886)	(21.504)	(17.445)
ddayHot_100	-0.5132***	-0.0065***	-0.5247***	-0.0065***
	(-22.256)	(-20.387)	(-22.685)	(-21.467)
prec_cm	1.1352***	0.0131***	1.1097***	0.0127***
	(22.032)	(19.201)	(21.887)	(19.174)
prec2_cm	-0.0088***	-0.0001***	-0.0086***	-9.697e-05***
	(-23.245)	(-19.700)	(-23.211)	(-19.642)
price_received	-0.9126***	-0.0117***	-0.9201***	-0.0117***
	(-248.02)	(-326.36)	(-241.85)	(-321.00)
==========	==========	==========	==========	==========
No. Observations	30003	30003	30003	30003
R-squared	0.8425	0.8770	0.8465	0.8817
P-value (F-stat)	0.0000	0.0000	0.0000	0.0000
==========	==========	==========	==========	==========
Time Fixed Effects	Year	Year	Year	Year
Location Effects	State	State	State	State
Cluster	County-crop	County-crop	County-crop	County-crop

Table 18: Robustness check: Simple OLS results in absence of IV or spatial correlations and with White standard errors.

	(1) yield	(2) ln(yield)	(3) yield (ln endog)	(4) ln(yield) (ln endog)
lnfracInsured			-3.8479***	-0.0440***
			(-13.478)	(-13.571)
lnsublia			-14.635***	-0.1678***
			(-48.632)	(-46.644)
fracInsured	-0.0629***	-0.0007***		
	(-7.3008)	(-7.3976)		
sublia	-226.82***	-2.3499***		
	(-38.267)	(-33.059)		
ddayMod_100	0.3126***	0.0038***	0.3674***	0.0044***
	(26.267)	(25.217)	(31.380)	(29.564)
ddayHot_100	-0.5132***	-0.0065***	-0.5247***	-0.0065***
	(-31.119)	(-31.372)	(-32.250)	(-32.136)
prec_cm	1.1352***	0.0131***	1.1097***	0.0127***
	(18.628)	(20.110)	(18.634)	(20.125)
prec2_cm	-0.0088***	-0.0001***	-0.0086***	-9.697e-05***
	(-19.244)	(-20.529)	(-19.290)	(-20.592)
price_received	-0.9126***	-0.0117***	-0.9201***	-0.0117***
	(-367.03)	(-406.56)	(-372.72)	(-416.27)
No. Observations	30003	30003	30003	30003
R-squared	0.8425	0.8770	0.8465	0.8817
F-statistic	1.145e+04	1.526e+04	1.18e+04	1.596e+04
P-value (F-stat)	0.0000	0.0000	0.0000	0.0000
Time Fixed Effects	Year	Year	Year	Year
Location Effects	State	State	State	State
Cluster	No	No	No	No

Table 19: Robustness check: Simple OLS results in absence of IV or spatial correlations and with HAC-robust standard errors.

	(1) yield	(2) ln(yield)	(3) yield (ln endog)	(4) ln(yield) (ln endog)
lnfracInsured			-3.8479***	-0.0440***
			(-3.7273)	(-3.3019)
lnsublia			-14.635***	-0.1678***
			(-9.8113)	(-9.6004)
fracInsured	-0.0629**	-0.0007**		
	(-2.4975)	(-2.3043)		
sublia	-226.82***	-2.3499***		
	(-12.895)	(-10.944)		
ddayMod_100	0.3126***	0.0038***	0.3674***	0.0044***
	(3.6266)	(3.2547)	(4.2482)	(3.7013)
ddayHot_100	-0.5132***	-0.0065***	-0.5247***	-0.0065***
	(-3.7235)	(-3.7970)	(-3.8940)	(-3.8385)
prec_cm	1.1352***	0.0131***	1.1097***	0.0127***
	(6.9163)	(6.3430)	(6.6369)	(6.0114)
prec2_cm	-0.0088***	-0.0001***	-0.0086***	-9.697e-05***
	(-8.4162)	(-8.2954)	(-7.9023)	(-7.7295)
price_received	-0.9126***	-0.0117***	-0.9201***	-0.0117***
	(-12.808)	(-17.277)	(-13.159)	(-17.726)
No. Observations	30003	30003	30003	30003
Cov. Est.	Driscoll-Kraay	Driscoll-Kraay	Driscoll-Kraay	Driscoll-Kraay
R-squared	0.8425	0.8770	0.8465	0.8817
F-statistic	1.145e+04	1.526e+04	1.18e+04	1.596e+04
P-value (F-stat)	0.0000	0.0000	0.0000	0.0000
Time Fixed Effects	Year	Year	Year	Year
Location Effects	State	State	State	State
Cluster	No	No	No	No

Table 20: 2sls with two endogenous variables and two IVs without spatial effects- The results of this just-identified model was very insignificant. I tried 2 IVs for one endogenous at a time instead.

	(1) yield, endog in levels	(2) ln(yield), endog in levels	(3) yield, endog in logs	(4) ln(yield), endog in logs
lnfracInsured			-304.91	-1.3044
			(-0.0769)	(-0.0837)
lnsublia			-187.23	-0.7803
			(-0.0711)	(-0.0756)
fracInsured	0.6421	0.0027		
	(0.3597)	(0.1722)		
sublia	313.55	1.5653		
	(0.6695)	(0.3522)		
ddayMod_100	0.4148***	0.0045***	0.4466	0.0046
	(2.8199)	(3.4021)	(0.2658)	(0.7056)
ddayHot_100	-0.6212***	-0.0074***	-0.6234	-0.0075***
	(-9.5494)	(-11.864)	(-1.5242)	(-4.5499)
prec_cm	1.5503***	0.0160***	-2.0112	0.0008
	(3.1600)	(3.5815)	(-0.0447)	(0.0048)
prec2_cm	-0.0119***	-0.0001***	0.0150	-7.59e-06
	(-3.4704)	(-3.8641)	(0.0439)	(-0.0057)
price_received	-0.8876***	-0.0115***	-1.2542	-0.0131
	(-16.941)	(-24.859)	(-0.2710)	(-0.7187)
Estimator	IV-2SLS	IV-2SLS	IV-2SLS	IV-2SLS
No. Observations	29002	29002	29002	29002
P-value (F-stat)	0.0000	0.0000	0.0000	0.0000
Time Fixed Effects	Year	Year	Year	Year
Location Effects	State	State	State	State
Cluster	Yes	Yes	Yes	Yes
Distance IV	Yes	Yes	Yes	Yes
Weather Station IV	Yes	Yes	Yes	Yes

Table 21: 2sls with two endogenous variables and one endogenous variable of sublia or lnsublia. to add: over-identification test

	(1) yield, endog in levels	(2) ln(yield), endog in levels	(3) yield, endog in logs	(4) ln(yield), endog in logs
sublia	173.98	0.9795		
	(0.9189)	(0.4583)		
lnsublia			15.186	0.0857
			(0.8440)	(0.4410)
ddayMod_100	0.3686***	0.0043***	0.3204***	0.0041***
	(8.9372)	(9.2330)	(6.4894)	(7.9618)
ddayHot_100	-0.6285***	-0.0075***	-0.6401***	-0.0075***
	(-12.230)	(-12.634)	(-9.4232)	(-10.147)
prec_cm	1.3921***	0.0153***	1.4454***	0.0156***
	(9.8092)	(10.035)	(7.0319)	(7.1542)
prec2_cm	-0.0109***	-0.0001***	-0.0113***	-0.0001***
	(-9.8608)	(-9.8105)	(-7.0229)	(-6.9909)
price_received	-0.9058***	-0.0116***	-0.8982***	-0.0115***
	(-129.71)	(-147.04)	(-61.941)	(-73.105)
Estimator	IV-2SLS	IV-2SLS	IV-2SLS	IV-2SLS
No. Observations	29002	29002	29002	29002
P-value (F-stat)	0.0000	0.0000	0.0000	0.0000
Time Fixed Effects	Year	Year	Year	Year
Location Effects	State	State	State	State
Cluster	Yes	Yes	Yes	Yes
Distance IV	Yes	Yes	Yes	Yes
Weather Station IV	Yes	Yes	Yes	Yes

Table 22: 2sls with two endogenous variables and one endogenous variable of fracInsured or lnfracInsured. to add: over-identification test

	(1) yield, endog in levels	(2) ln(yield), endog in levels	(3) yield, endog in logs	(4) ln(yield), endog in logs
fracInsured	-0.5633	-0.0033		
	(-0.8691)	(-0.4480)		
lnfracInsured			-23.104	-0.1300
			(-1.0687)	(-0.4932)
ddayMod_100	0.3209***	0.0041***	0.3297***	0.0041***
	(7.9881)	(8.9313)	(10.165)	(11.022)
ddayHot_100	-0.6200***	-0.0074***	-0.6394***	-0.0075***
	(-13.147)	(-13.690)	(-11.575)	(-11.234)
prec_cm	1.2225***	0.0144***	1.1851***	0.0142***
	(10.959)	(12.142)	(9.3754)	(9.9780)
prec2_cm	-0.0097***	-0.0001***	-0.0093***	-0.0001***
	(-13.291)	(-14.514)	(-10.301)	(-10.701)
price_received	-0.9227***	-0.0117***	-0.9251***	-0.0117***
	(-54.041)	(-61.412)	(-57.149)	(-60.996)
Estimator	IV-2SLS	IV-2SLS	IV-2SLS	IV-2SLS
No. Observations	29002	29002	29002	29002
P-value (F-stat)	0.0000	0.0000	0.0000	0.0000
Time Fixed Effects	Year	Year	Year	Year
Location Effects	State	State	State	State
Cluster	Yes	Yes	Yes	Yes
Distance IV	Yes	Yes	Yes	Yes
Weather Station IV	Yes	Yes	Yes	Yes

9 Dynamic Effects of Upward Subsidy Shocks in Farm Bills

Comments from Christian Matthes (November 3, 2025):

1. Use long LP as it is the best practice.
2. Have subsidy of each year as the main variable, and amount of subsidy shocks in the previous Farm Bills as control variable.
3. He can join my committee, but will not write a recommendation letter.
4. The choice of spatial LP and any other configurations, depends on my IV variable, and whether it will cover it or not.

I borrowed the Local Projection (LP) method from the macroeconomics literature (Jordà (2023) and Jordà and Taylor (2025)) to study the dynamic effects of subsidies on farm outcomes. LP is a semiparametric method that imposes relatively weak assumptions on data and shape of the response. This choice comes at the cost that LP will be less efficient than models that impose more structure, such as Vector Autoregression (VAR). Using LP for estimating the dynamic effect of subsidies puts this paper in a recent literature on studying the dynamic effects of policy shocks like Baquie et al. (2025) that study the long-term effect of industrial policy instruments on competitiveness of the targeted product and Paredes et al. (2022) that studies the impact of introducing the tax incentive scheme for corporate R&D in firms from different R&D intensity sectors.

I am interested in how the Governmental intervention in the crop insurance market by providing subsidy affects the average outcome in the future relative to a baseline of no-intervention. The impulse response to the intervention is

$$\mathcal{R}_{s\to y}(h,\delta) \equiv \mathbb{E}[y_{t+h} \,|\, s_t = s_0 + \delta;\, \mathbf{x}_t] - \mathbb{E}[y_{t+h} \,|\, s_t = s_0;\, \mathbf{x}_t]\,, \qquad h = 0, 1, \ldots, H. \tag{24}$$

where δ is the size of the intervention, or *dose*, and s_0 is a baseline level. I calculated the heteroskedasticity-robust and serial correlation-robust standard errors within each cluster of county-

crop (Arellano 1987).

When it comes to the selection of variables of past periods as explanatory variables (lag of variables), there is no well-established method to select how many lags to include in the LP (Jordà and Taylor 2025, p. 65). The selection criteria are invalid when residuals are autocorrelated, as in LPs for $h > 1$. However, since the LP in the first horizon h = 1 is equivalent to the corresponding equation in a VAR, a natural approach is to use information criteria to determine the lag length similar to VAR method for that case, and then use the same lag length at subsequent horizons (Jordà and Taylor 2025, p. 65).

Knowing this, I chose the optimal number of lags based on the VAR literature, and using the *pvarsoc* command in *STATA*. It reports the model overall coefficient of determination, Hansen (1982) J statistic, its corresponding p-value, as well as the moment model selection criteria developed by Andrews and B. Lu (2001) based on the J statistic. Andrew and Lu's criteria are all based on Hansen's J statistic, which requires the number of moment conditions to be greater than the number of endogenous variables in the model (Abrigo and Love 2016). Tables 23 and 24 report the Modified Bayesian Information Criterion (MBIC), Modified Akaike Information Criterion (MAIC), and Modified Hannan-Quinn Information Criterion (MQIC) for corn and soybeans, respectively. These parameters penalize model complexity (e.g., the number of lags or instruments) to help select the optimal lag length or model specification in GMM-based systems like the panel VAR used in pvarsoc. In Tables 23 and 24, all MBIC, MAIC and MQIC decline monotonically from 1 to 3 lags and reach their minima at the third lag. Because the local projection at $h = 1$ is equivalent to the corresponding VAR, it is standard to select the dynamic order from a (P)VAR at $h = 1$ and carry that order to all horizons (as also recounted by Jordà and Taylor (2025)).

Table 23: Panel VAR lag order selection for corn (Sample: 1993–2012)

lag	CD	J	J pvalue	MBIC	MAIC	MQIC
1	0.8996546	1591.311	0	1479.255	1567.311	1537.693
2	0.9092503	1100.909	2.4e-232	1026.205	1084.909	1065.164
3	0.9202226	181.7747	3.10e-38	144.4229	173.7747	163.902
4	-0.9204785	.	.	.	.	.

No. of obs = 11361
No. of panels = 634
Ave. no. of T = 17.920

In order to model the dynamic effects of crop subsidies, I consider a typical LP of the form

$$\ln(Y_{ci,t+h}) = \theta_h \Gamma_{ci,t} + \sum_{p=1}^{3} \rho_{h,p} \ln(Y_{ci,t-p}) + \sum_{p=1}^{3} \kappa_{h,p} \Gamma_{ci,t-p} + \boldsymbol{\gamma}_h' \mathbf{x}_{ci,t} + \beta_{ci} + \delta_t + \nu_{ci,t+h}, \quad h = 0, 1, \ldots, H. \tag{25}$$

where:

- $\ln(Y_{ci,t+h})$ denotes the log of the outcome variable (e.g., crop yield) for crop c in county i at horizon $t + h$.

Table 24: Panel VAR lag order selection for soybeans (Sample: 1993–2012)

lag	CD	J	J pvalue	MBIC	MAIC	MQIC
1	0.9095104	1244.541	4.4e-259	1132.776	1220.541	1190.985
2	0.9152186	955.4655	6.1e-201	880.9558	939.4655	919.7617
3	0.9311757	199.2435	5.46e-42	161.9887	191.2435	181.3917
4	0.9009046	.	.	.	.	.

No. of obs = 11089
No. of panels = 619
Ave. no. of T = 17.914

- $\Gamma_{ci,t}$ is the endogenous policy or treatment variable of interest (e.g., insurance subsidy rate) observed at time t.
- $\Gamma_{ci,t-p}$ for $p = 1, 2, 3$ are the lagged values of the endogenous variable. The optimal number of lags is from Tables 24 and 23.
- $\ln(Y_{ci,t-p})$ for $p = 1, 2, 3$ are the lagged dependent variables, included to control for dynamic persistence. The optimal number of lags is from Tables 24 and 23.
- $\mathbf{x}_{ci,t}$ is a vector of control covariates (e.g., weather variables, input prices, or crop shares).
- $\boldsymbol{\gamma}_h$ is the vector of coefficients on the control variables at horizon h.
- θ_h measures the contemporaneous response of the dependent variable at horizon h to the shock $\Gamma_{ci,t}$.
- $\rho_{h,p}$ and $\kappa_{h,p}$ capture the effects of the p-th lag of the dependent and endogenous variables, respectively.
- β_{ci} represents county-crop fixed effects that absorb time-invariant heterogeneity.
- δ_{ct} represents crop-time fixed effects that absorb aggregate crop-specific shocks common to all counties.
- $\nu_{ci,t+h}$ is the error term at horizon h.
- $h = 0, 1, \ldots, H$ indexes the projection horizon.

Given the endogenous nature of the treatment, we estimate the LP using instrumental variables method, LP-IV, introduced by Jordà, Schularick, and Taylor (2015) in their study of the effect of low interest rate environments on mortgage lending and house prices.

Using the IVs in the LP setting still requires satisfying two conditions of relevance and independence (as first evaluated in Section 5). The exogeneity condition in an LP setting is slightly different than usual due to the dynamic structure of the problem, and it takes the form of (Jordà 2023, sec. 7):

$$\text{Lead-lag exogeneity:} \quad E(\nu_{t+h}, \mathbf{z}_t^P) = 0 \quad \text{for all h.}$$

in which, ν_{t+h} is the error term in h^{th} period impulse response, and $\mathbf{z}_t$ is the vector of instrumental variables. Because lags of the endogenous variable is included in $\mathbf{x}_{ijt}$, potential incidental parameter

biases could arise in short panels with a highly serially correlated endogenous variable. Per recommendation of Jordà (2023, p. 626), we used an Arellano-Bond estimator to address this serial correlation.

9.1 Results-Dynamic Effects of Upward Subsidy Shocks in Farm Bills

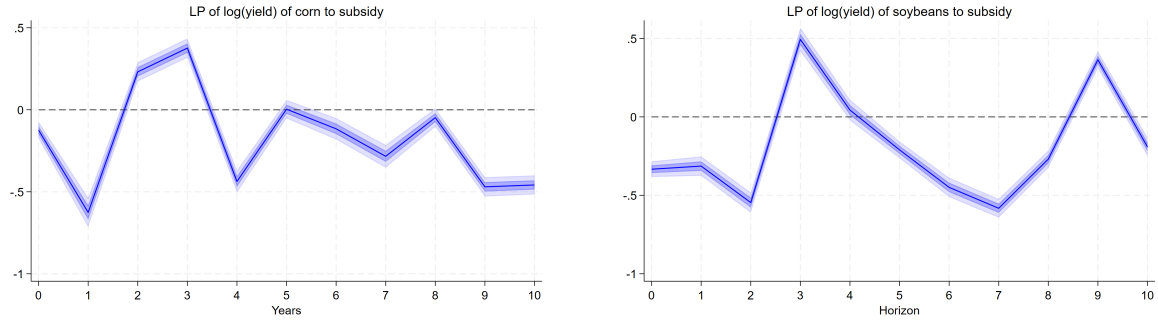
I used the *locproj* STATA package developed in Ugarte-Ruiz (2025b) and Ugarte-Ruiz (2025a) to model the LP for panel data in presence of endogeniety.

Figure 32 shows results for local project for a 10-year period. The graph shows two point-wise confidence intervals at the 95% and at the 66%, with the later to roughly shows the ±1 standard error band. Notice that since the impulse response coefficients are correlated with one another Jordà (2009) , I cannot make inferential statements about the shape of impulse response based on the error bands. In order to asses the overall significance of the impulse response, I did a joint test of the null hypnosis that all the response coefficients are zero (as recommended by Jordà (2023)).[10]

Overall Significance of the Impulse Response Since the impulse response coefficients are correlated with one another Jordà (2009) , I cannot make inferential statements about the shape of impulse response based on the error bands. In order to asses the overall significance of the impulse response, I did a joint test of the null hypnosis that all the response coefficients are zero (as recommended by Jordà (2023)). I assess the overall significance of the impulse response by estimating a stacked local projection in which the shock is interacted with horizon dummies, and then conducting a joint Wald test of the null that all horizon-specific coefficients are zero (cf. Jordà (2009); Jordà (2023)). Specifically, letting $\widehat{\boldsymbol{\theta}} = (\hat{\theta}_0, \ldots, \hat{\theta}_H)'$ denote the estimated effects of $1\{H = h\} \times \text{subsidy}_t$ and $\widehat{\Sigma}$ their covariance matrix (estimated with clustered or Driscoll–Kraay standard errors), the test statistic

$$W = \widehat{\boldsymbol{\theta}}^\top \widehat{\Sigma}^{-1} \widehat{\boldsymbol{\theta}} \sim \chi^2_{H+1}$$

under $H_0 : \theta_0 = \cdots = \theta_H = 0$. Rejecting H_0 implies that the impulse response is jointly different from zero across horizons.



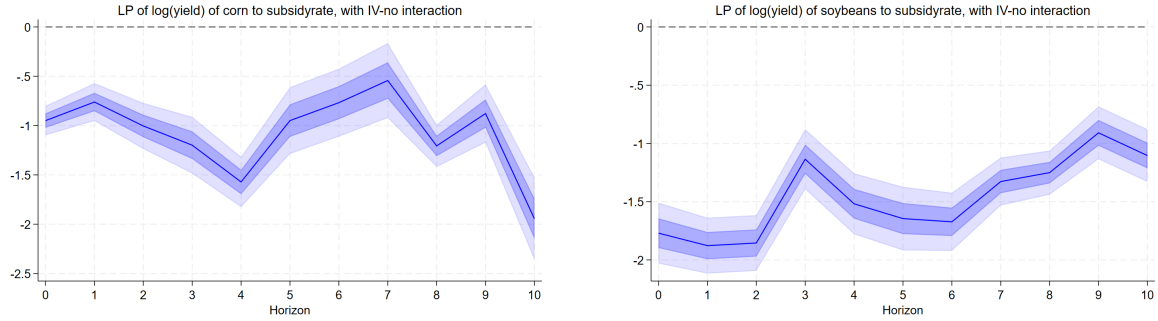
(a) Corn: LP of log(yield) to subsidy

(b) Soybean: LP of log(yield) to subsidy

Figure 32: Local projections responses (horizons 0–10) with controls and FE.

[10] I thank Christian Matthes for his insights about applications of LP. I thank Zhuyin Li for the fruitful discussions about the prerequisites of using LP in research.

Results in presence of Instrumental Variable When considering an IV for resolving the endogeneity problem, the impulse response changes to 33.



(a) Corn: LP of log(yield) to subsidy

(b) Soybean: LP of log(yield) to subsidy

Figure 33: Local projections responses (horizons 0–10) with Instrumental Variable of Tsiboe and Turner (2023a) and controls and FE.

Figures ?? and ?? show results for local project in 10-year period. The error bands provide a measure of estimation uncertainty. Notice that since the impulse response coefficients are correlated with one another Jordà (2009) , we cannot make inferential statements about the shape of impulse response based on the error bands. In order to asses the overall significance of the impulse response, we did a joint test of the null hypnosis that all the response coefficients are zero (as recommended by Jordà (2023)). to do: a joint test of the null hypothesis that all the response coefficients are zero has p-value = xx.[11]

We also tested for structural differences before and after the 1994 reform of the Federal Crop Insurance Program, we interact the shock variable with a post-1994 dummy. When taking sublia as the measure of intensity of insurance enrollment, the resulting impulse response function (Figure 36) shows that yield losses following subsidy increases are significantly larger and more persistent after 1994, consistent with greater behavioral distortion under the expanded subsidy regime. However, when taking fracInsured as the measure of intensity of insurance enrollment (Figure 37), the pattern is similar but slightly flatter and shifted upward, meaning the negative long-run yield response is weaker after 1994. That implies the 1994 reform may have reduced the marginal distortion of additional enrollment intensity by broadening participation and changing subsidy design.

Figure 38 presents impulse response functions of outcome to changes in sublia as a measure of insurance enrollment intensity. Top panel shows the impulse response of yields to changes in sublia before 1995 is negative and persistent, indicating that higher intensity was associated with slightly lower yields over subsequent years. This suggests limited short run adaptation and possible inefficiencies in production before the 1994 reform that made crop insurance participation mandatory. Middle panel shows that after 1995, the response remains negative but is smaller in magnitude and shorter in duration. This attenuation may reflect structural changes in the Federal Crop Insurance Program, such as broader participation and revised subsidy rates, which reduced the heterogeneity in responses across regions. Finally,,the bottom panel is the difference between post-1995 and pre-1995 responses isolating the policy induced change. The small positive coefficients after the fifth year imply that relative to the earlier period, post reform insurance subsidies had a somewhat less negative (or slightly more stabilizing) effect on yields. This suggests improved program design and risk management practices.

[11] I thank Christian Matthes for his insights about applications of LP. I thank Zhuyin Li for the fruitful discussions about the prerequisites of using LP in research.

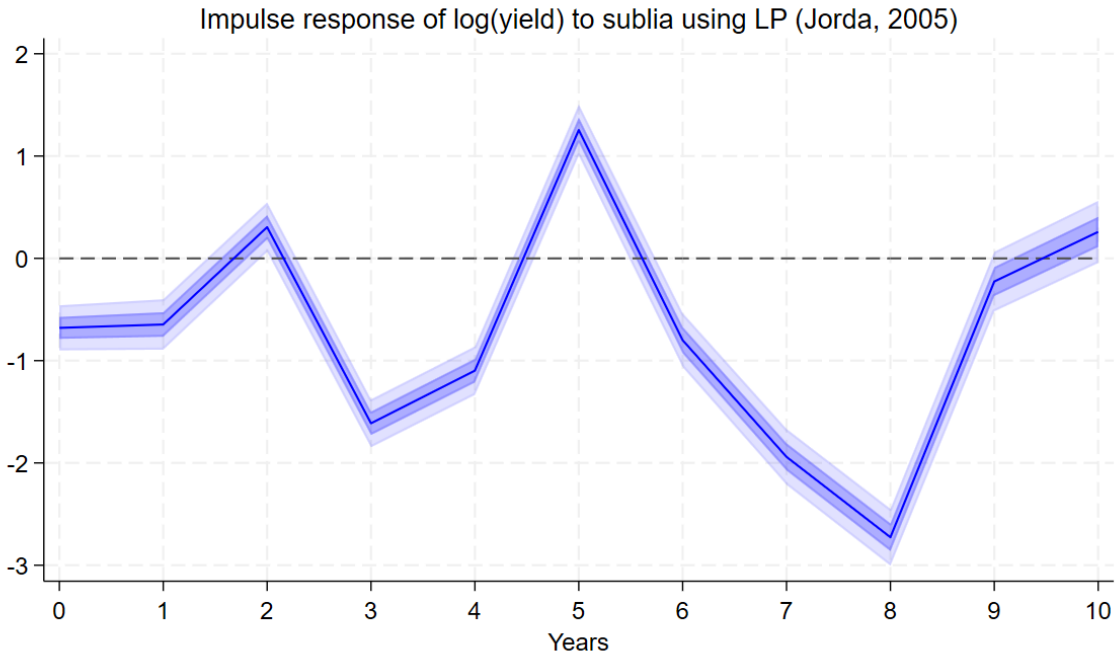


Figure 34: LP with two lag values of yield, sublia and price received as explanatory variables. The graph shows two point-wise confidence intervals at the 95% and at the 66%, with the later to roughly shows the ±1 standard error band.

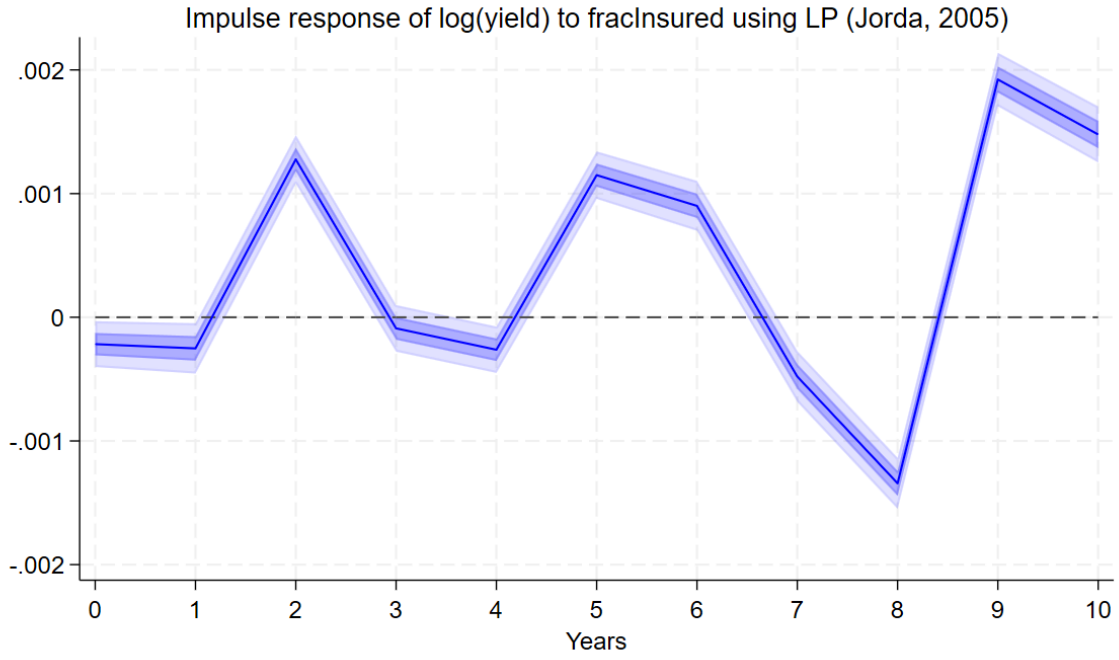


Figure 35: LP with two lag values of yield, fracInsured and price received as explanatory variables. The graph shows two two point-wise confidence intervals at the 95% and at the 66%, with the later to roughly shows the ±1 standard error band.

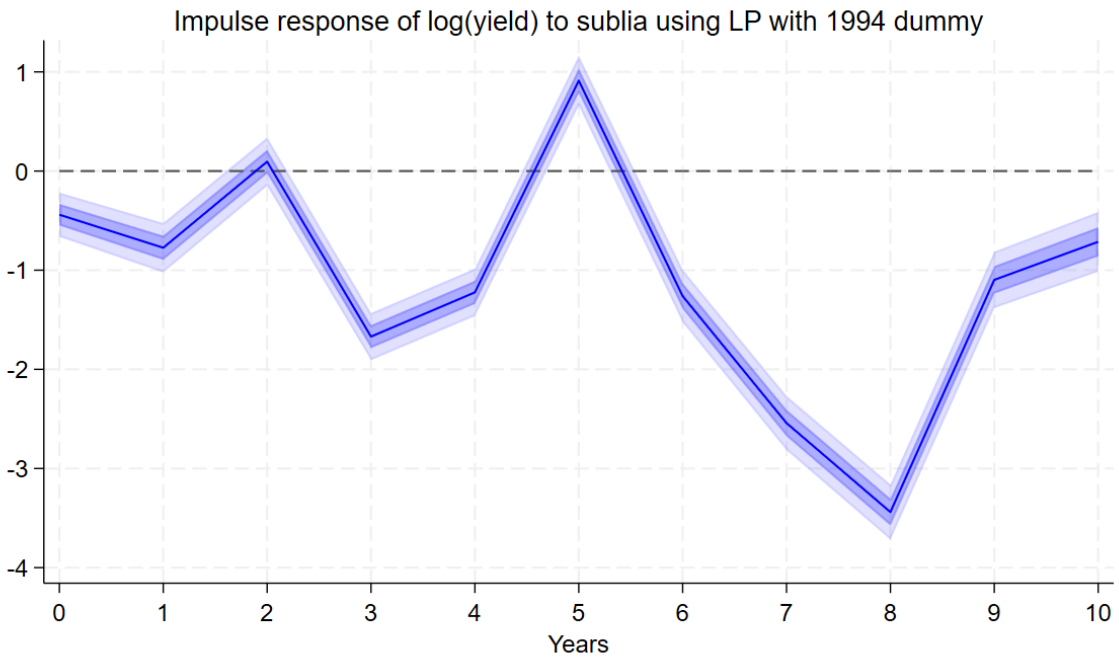


Figure 36: LP with two lag values of yield, sublia and price received as explanatory variables. The graph shows two point-wise confidence intervals at the 95% and at the 66%, with the later to roughly shows the ±1 standard error band.

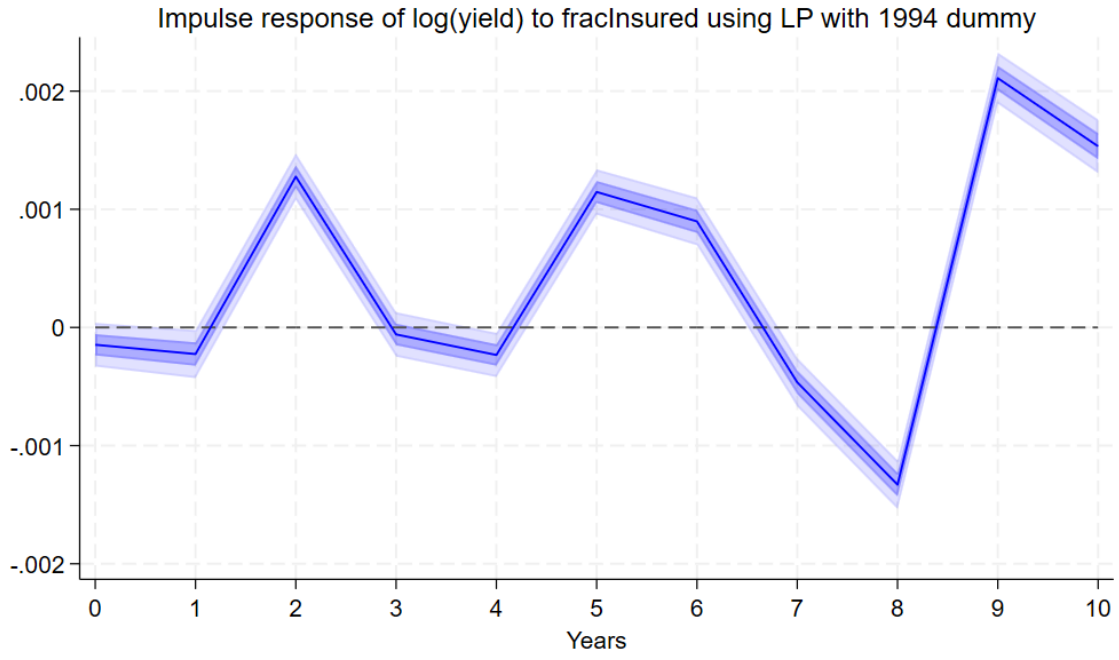


Figure 37: LP with two lag values of yield, fracInsured and price received as explanatory variables. The graph shows two two point-wise confidence intervals at the 95% and at the 66%, with the later to roughly shows the ±1 standard error band.

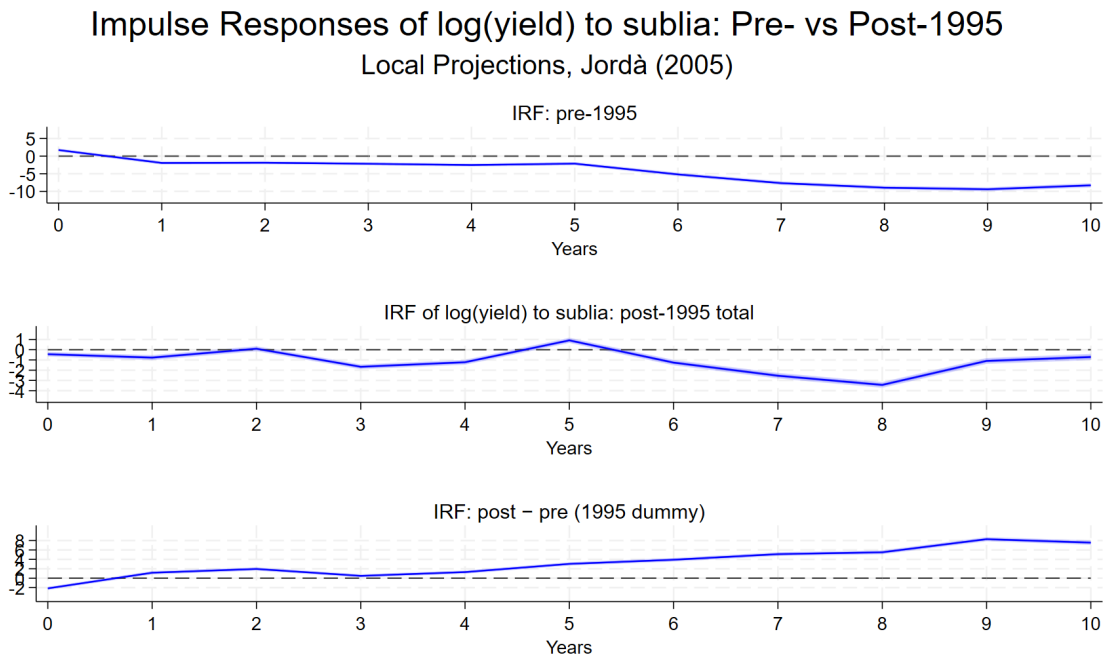


Figure 38: LP with two lag values of yield, fracInsured and price received as explanatory variables. The graph shows two two point-wise confidence intervals at the 95% and at the 66%, with the later to roughly shows the ±1 standard error band.

In the top panel of Figure 39, before 1995, a rise in the fraction of insured acreage is followed by a slight, statistically insignificant decline in yields, which then stabilizes. This negative effect may reflect moral hazard or adverse selection in the early years of the insurance program. Farmers who purchased insurance were often those facing higher production risks, and insurance may have reduced incentives to exert full effort or adopt yield-stabilizing practices. In the middle panel, after 1995, the yield response remains small and near zero, but it becomes less negative and more short-lived. This suggests that as crop insurance coverage expanded and the program matured, its distortive effects on yield outcomes diminished. Broader participation and improved underwriting likely mitigated selection biases and strengthened the link between insurance and risk management rather than moral hazard. Finally, in the bottom panel, the differential effect shows a slightly positive and persistent response, implying that, relative to the pre-reform period, yields became modestly higher following increases in insurance coverage. This shift is consistent with improved program design, such as tighter compliance, data-driven premium rating, and the diffusion of adaptive technologies, that enhanced farmers' ability to manage risk without reducing productivity.

9.2 Event Study of Farm Bills as Policy Shocks

In this section, I modified the LP by defining the percent change in county-crop level subsidies between two subsequent 1994, 1996, 2000, 2008, 2014 and 2018 Farm Bills as policy shocks. I used 3 lags for the dependent and shock variables as suggested in Tables 25 and 26.

The two graphs in Figure 40 display local projection (LP) impulse responses of crop yields to subsidy shocks associated with successive U.S. Farm Bills. For corn, the impulse response fluctuates around zero, with alternating positive and negative coefficients across horizons, and modest confidence intervals—indicating that yield responses to subsidy shocks are small and statistically weak.

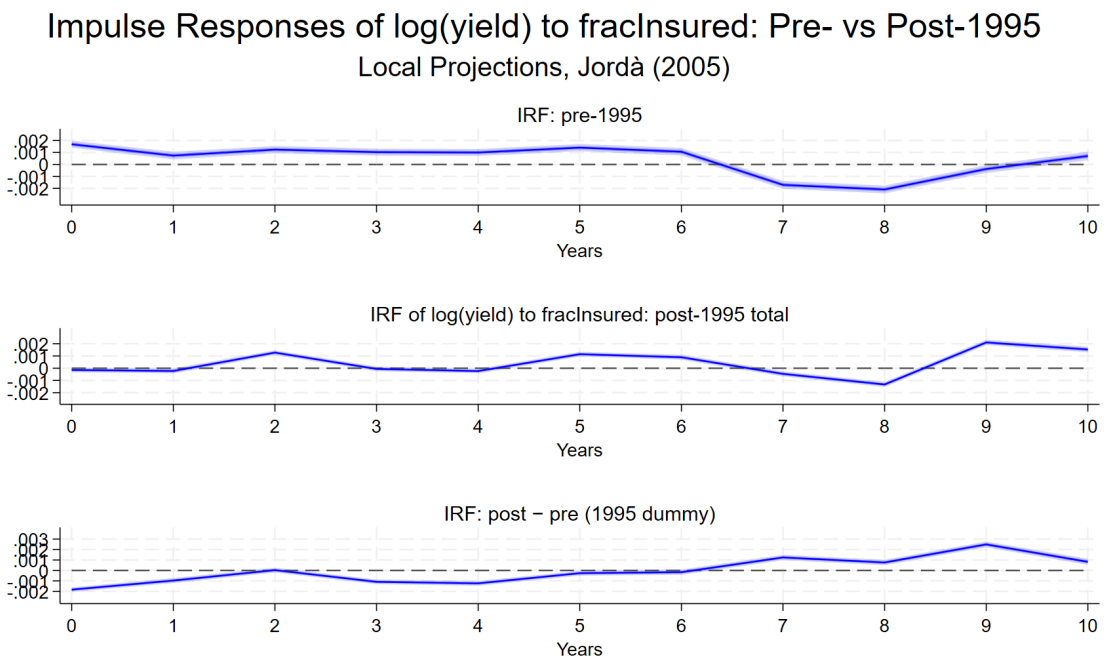


Figure 39: LP with two lag values of yield, fracInsured and price received as explanatory variables. The graph shows two two point-wise confidence intervals at the 95% and at the 66%, with the later to roughly shows the ±1 standard error band.

Table 25: Panel VAR lag order selection for corn with FB shocks (Sample: 1993–2012)

Lag	CD	J	J pvalue	MBIC	MAIC	MQIC
1	0.1137823	1751.253	0	1639.246	1727.253	1697.646
2	0.0256316	1054.1	3.1e-222	979.4286	1038.1	1018.362
3	-0.2004721	489.9997	9.7e-105	452.6638	481.9997	472.1304
4	0.4324079	.	.	.	.	.

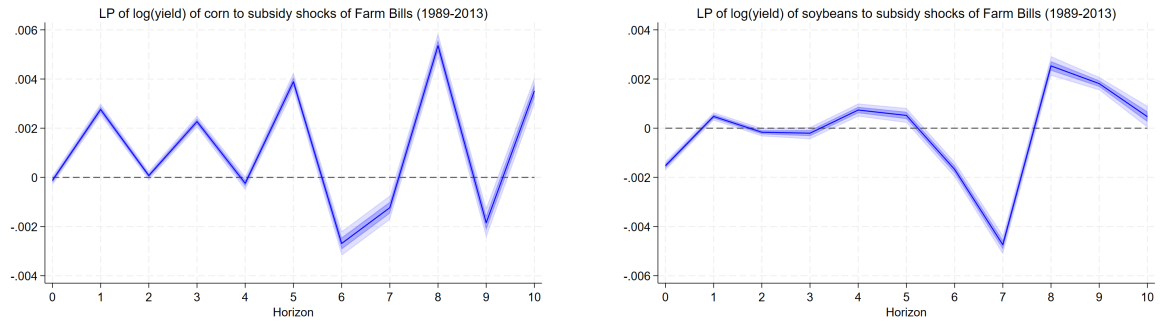
No. of obs = 11361
No. of panels = 633
Ave. no. of T = 17.877

Table 26: Panel VAR lag order selection for soybeans with FB shocks (Sample: 1993–2012)

Lag	CD	J	J pvalue	MBIC	MAIC	MQIC
1	0.2900373	1064.687	2.3e-220	952.9978	1040.687	1011.148
2	0.3108157	718.0961	9.1e-150	643.6364	702.0961	682.403
3	0.2879469	268.5619	6.51e-57	231.3321	260.5619	250.7154
4	0.4176169	.	.	.	.	.

No. of obs = 11020
No. of panels = 619
Ave. no. of T = 17.803

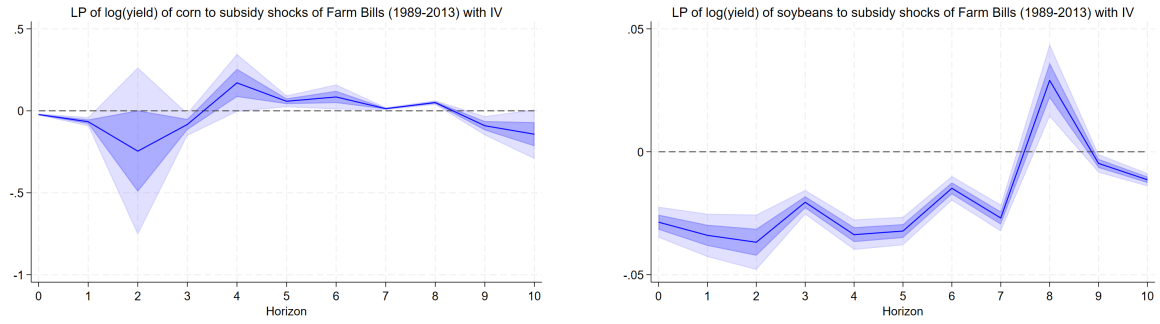
This oscillation could reflect short-term adjustments in production practices, insurance coverage, or acreage allocation following the enactment of new Farm Bill provisions, but without a clear long-term trend. For soybeans, the estimated effects are even smaller, with a mostly flat response and only a slight uptick around horizon 8, suggesting limited yield sensitivity to subsidy changes. The narrow confidence intervals imply precise but near-zero effects. Taken together, these patterns suggest that in the Midwest, while Farm Bill subsidy changes affected insurance incentives and risk coverage, did not translate into persistent yield-level responses. Instead, the observed fluctuations likely capture short-run adaptation and risk-management behavior rather than a causal impact of subsidy reform on productivity.



(a) Corn (b) Soybeans

Figure 40: Local projections for corn and soybeans under Farm Bill shocks without IV, without interaction with climate variables, with three lags of log(yield) and subsidy shocks.

In the next step, I used the IV developed by Tsiboe and Turner (2023a) for running the LP-IV as discussed by Jordà, Schularick, and Taylor (2015).



(a) Corn (b) Soybeans

Figure 41: Local projections for corn and soybeans under Farm Bill shocks with textcitetsiboe2023a 's IV

Instrument Relevance. To verify the strength of the instrument `tau_final`, we estimate the first-stage regression of the endogenous variable of shock to subsidies in the Farm Bills on the instrument and all control variables under a fixed-effects specification:

$$\texttt{fbshock}_{it} = \alpha_i + \gamma_t + \pi\, \texttt{tau_final}_{it} + \mathbf{X}'_{it}\boldsymbol{\delta} + \varepsilon_{it},$$

where α_i denotes county fixed effects and $\mathbf{X}_{it}$ includes price, weather, and lagged yield controls.

Table 27: First-stage fixed-effects regression of `fbshock` on `tau_final` and controls

	Coefficient	Std. Err.	t-stat	p-value
`tau_final`	64.338	4.446	14.47	0.000
`price_received`	0.497	0.007	62.77	0.000
`ddayMod_100`	-0.098	0.007	-13.67	0.000
`ddayHot_100`	-0.197	0.008	-25.51	0.000
`prec_cm`	-0.245	0.044	-5.51	0.000
`prec2_cm`	0.002	0.000	6.10	0.000
`corntime`	-0.340	0.029	-11.89	0.000
First-stage F-statistic: $F = 209.38$				

Instrument relevance requires that $\pi \neq 0$, meaning that the instrument is strongly correlated with the endogenous regressor after conditioning on the controls. The coefficient on `tau_final` is large and highly significant ($t = 14.47$, $p < 0.001$), indicating strong correlation between the instrument and the policy shock variable. The corresponding first-stage F-statistic is computed as:

$$F = \left(\frac{\hat{\pi}}{SE(\hat{\pi})}\right)^2 = \left(\frac{64.34}{4.45}\right)^2 = 209.38,$$

which far exceeds the conventional weak-instrument threshold of $F > 10$ (Staiger and Stock 1997), confirming that `tau_final` is a highly relevant instrument for `fbshock`. The main Stata output is summarized in Table 27.

This strong first-stage relationship ensures that `tau_final` provides meaningful exogenous variation in `fbshock`, thereby satisfying the relevance condition for valid instrumental-variable estimation.

10 Dynamic Spatial Effects of Farm Bills

In section 9, I used county fixed effect to absorb all time-invariant spatial heterogeneity like soil quality in modeling the local projection. But the model still assumed that after controlling for fixed effect, the error term is spatially independent across counties. It means that any correlation between neighboring counties' shocks, such as drought spillovers, policy diffusion or local markets, is treated as noise, and hence, not modeled.

In order to bring in information about the spatial propagation, spillovers, or cross-county feedback, I add an extra layer spatial analysis to the model in this section. I follow the spatial local projection method applied by Brady (2011) and Gonne and Hubert (2020) for studying the dynamic effect of subsidy shocks, taking into account the spatial overlaps. I define the spatial autoregressive model by including the spatially lag dependent variable as a regressor, and call it hereafter the *spatial regressor*.[12] This spatially lag variable captures the direct effect of the spatial interaction in determining the dependent variable. As Brady (2011, p. 215) put, spatial correlation can be thought of an omitted variable problem, and adding a regressor that captures the spatial correlation between regions, forms the spatial autoregressive model.

[12] Brady (2011, p. 215) introduces this style of simplifying the expressions, to prevent confusion with the common terms in the time series analysis.

Spatial Weight Matrix Brady (2011, p. 215) defines the spatial weight matrix W by contiguity, and assumes that the spatial propagation does not fade with distance. This simplifying assumption is common in the spatial literature and helps with simplifying the estimation of spatial models (Anselin and Bera 1998).

I use a row-standardized inverse-distance matrices; $w_{ii} = 0$. Common distance-decay choices choose one and expand it:

$$\text{Power (inverse distance):} \quad w_{ij} \propto d_{ij}^{-p}, \qquad \text{Negative exponential:} \quad w_{ij} \propto e^{-\phi d_{ij}}.$$

Using multiple decays ("*NSLX*") is possible by including separate $W^{(1)}$ and $W^{(2)}$ built with different (p, ϕ).

SLX model Following the Spatial Cross–Regressive model of Florax and Folmer (1992) and its modern treatment in Halleck Vega and Elhorst (2015), I exclude a spatial lag of the dependent variable and allow only spatially lagged regressors. For horizon h in a local-projection setting:

$$y_{i,t+h} - y_{i,t-1} = \alpha_i + \tau_t + \beta_h D_{it} + \theta_h (WD_{it}) + \boldsymbol{\gamma}_h^\top \mathbf{X}_{it} + \boldsymbol{\eta}_h^\top (W\mathbf{X}_{it}) + \varepsilon_{i,t+h} \tag{26}$$

in which, WD_{it} captures spatial spillovers of the (endogenous) policy/treatment; and $W\mathbf{X}_{it}$ are optional spatial lags of exogenous controls. α_i and τ_t are unit and time fixed effects; errors are $\varepsilon_{i,t+h}$.

I also estimate the dynamic spatial marginal effects presented in Equation 27 per (Anselin 2024d)'s specification. This is a methodological contribution of this paper. check it with Luc Anselin.

$$\mathbf{IMP} = \frac{\partial \mathbb{E}(\mathbf{Y})}{\partial \mathbf{x}_k'} = \begin{bmatrix} \frac{\partial \mathbb{E}(y_1)}{\partial x_{1k}} & \frac{\partial \mathbb{E}(y_1)}{\partial x_{2k}} & \cdots & \frac{\partial \mathbb{E}(y_1)}{\partial x_{nk}} \\ \frac{\partial \mathbb{E}(y_2)}{\partial x_{1k}} & \frac{\partial \mathbb{E}(y_2)}{\partial x_{2k}} & \cdots & \frac{\partial \mathbb{E}(y_2)}{\partial x_{nk}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \mathbb{E}(y_n)}{\partial x_{1k}} & \frac{\partial \mathbb{E}(y_n)}{\partial x_{2k}} & \cdots & \frac{\partial \mathbb{E}(y_n)}{\partial x_{nk}} \end{bmatrix} = \mathbf{S}^{-1}\boldsymbol{\beta}_k, \tag{27}$$

10.1 Specification Tests of the Dynamic Spatial Estimation

This section assesses whether the empirical model satisfies the assumptions required for valid inference and, where violations are likely, documents the robust procedures I use.

Null Hypothesis: There is no spatial spillover from the insurance subsidy policy.

$$\text{Null Hypothesis:} \quad H_0 : \theta_h = 0$$

It means the spatially lagged treatment WD_{it} does not affect outcomes in other counties (no cross-county policy spillovers).

One common specification test in the spatial modeling literature is Moran's I (Moran 1950). In the context of the spatial local projection, Moran's I serves as a useful but limited specification diagnostic. A failure to reject the null of zero spatial autocorrelation in residuals suggests that the model captures the main spatial structure adequately. However, rejecting the null must be interpreted cautiously, because Moran's I does not identify which spatial process is misspecified, and it could arise from a missing spatial lag of the dependent variable, a spatial error process, or omitted spatially correlated

regressors (as in an SLX-type model). Moreover, Moran's I has power against many alternative forms of misspecification, including non-normality and heteroskedasticity, so rejection alone does not confirm spatial dependence. Thus, while Moran's I is a good general misspecification test, it is not diagnostic in isolation (Anselin 2024e).

Due to this limitation, I used the Lagrange Multiplier (LM) Tests for specification. In spatial specification, one-directional tests like LM error test, LM lag test, or LM SLX (aka LM-WX) tests, have power against other alternatives. It means they reject they Null, when the cause for rejecting Null is not what researchers are actually looking for. This happens beccause these tests are affected by the presence of other alternatives. For example, it is possible to have a situation that the real case is an *error* case, researchers run a *lag* test, and it's get rejected; and researcher conclude that it is a lag, although it is not. This is where this notion of robust test comes into play (Anselin 2024a). What is the written formal reference for this paragraph? To address this possible misinterpretation, Bera and Yoon (1993) proposed a correction of LM tests under local misspecification, and calls it *robust LM-test*. This test is robust to local[13] missepecification under another alternative. Still this test suffers from a phenomenon called *Two out of Three Problem* (Anselin 2024b), which means it is not possible to develop three-way LP tests for lag, error and SLX (aka WX).

Summary Impact Measures Following LeSage and Pace (2009), I estimated two dynamic spatial summary measures as follows:

- **ADI** = Average Direct Impact of x_i on y_i
  - Average of the trace of the impact matrix IMP
  - Formula:
$$ADI = \frac{1}{n}\text{tr}(S^{-1}\beta_k) = \frac{1}{n}\sum_{i=1}^{n}\text{IMP}_{ii}$$

- **ATI** = Average Total Impact
  - Average of all elements of the impact matrix IMP
  - Multiplier: $\dfrac{1}{1-\rho}$
  - Formula:
$$ATI = \frac{1}{n}\mathbf{1}_n' S^{-1}\mathbf{1}_n\beta_k = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\text{IMP}_{ij}$$

10.2 Choice of Spatial Model

- I did not use a Spatial Lag Model, as it is a cross-sectional model (Anselin 2024c).
- I did not use Spatial Error Model, as the data set is so large, that it is not prone to this error. for the spatial correlation of error terms in measurement. Spatial structure in the covariance matrix. Use HAC standard errors with OLS.
- Very useful: SLX model, rediscovered by Vega and Elhorst (2015) of the Spatial Cross-Regressive Model (Florax and Folmer, 1992). Only spatially lagged endogenously variable. No spatially

[13] Here local means for small amounts of error in misspecification.

lagged dependent variable. Use of inverse of distance. Distance Decay graphs with NSLX parameters: inverse distance power function and negative exponential distance.

- Classic Spatial Durbin Model : has spill-over effects
- General Nesting Spatial Model (GNS), by Elhorts (2014): the specification doesn't work. We cannot meaningfully estimate it.

I used the specification test proposed in Bera, Doğan, et al. (2020) for the specification tests of the spatial panel data model.

10.3 Price Volatility

RMA uses a measure of price volatility based on the Black-Scholes Model. This model provides a formula that translates options prices (the amount the market charges to 'lock-in' a future price) into an implied volatility of the price of the commodity. This price volatility is used in the calculation of RMA's premium rates for revenue coverage. The result is that the premium rate RMA charges to lock-in a future (harvest time) price through crop insurance is equivalent to what the market charges to lock in a price through an options contract (United States Department of Agriculture, Risk Management Agency 2025).

11 Difficulty of Access to Public Service

To quantify geographic access to USDA Service Centers, I construct a continuous measure of administrative proximity based on Voronoi (Thiessen) polygons. Each office defines a Voronoi cell consisting of all locations that are closer to that office than to any other. I assign counties to offices using the centroid-based Voronoi partition and compute the road-network distance from each county centroid to its nearest USDA office. This Voronoi-weighted distance captures meaningful variation in administrative accessibility while avoiding arbitrary boundary definitions. The approach follows a growing literature that uses Voronoi tessellations or Thiessen polygon areas to model spatial access or local market areas—such as Seim (2006), who uses Voronoi regions to characterize competitive territories in empirical entry games; Holmes (2011), who relies on nearest-store market areas in analyzing the diffusion of Walmart; and Deschênes and Greenstone (2007), who use Thiessen polygons to interpolate climate exposure across U.S. counties. This continuous geographic access measure provides a natural and transparent way to characterize administrative reach, travel costs, and service accessibility in the federal crop insurance program.

12 Impact of Downsizing the USDA on Output

Historically, the Department of Agriculture's (USDA) Farm Service Agency has provided personalized service to farmers through thousands of field offices scattered across the country. These county offices run various commodity, loan, conservation, and emergency disaster assistance programs. In response to the Government Performance and Results Act of 1993 and the Vice-President's National Performance Review, the Secretary of Agriculture initiated a departmental reorganization to refocus and simplify the Department's headquarters structure, improve accountability and service to customers, reform the Department's field structure, and reduce costs. Several components of the Secretary's reorganization required enabling legislation, which the Congress provided in the Reorganization Act of

1994. In 1994, USDA was required to combine its field offices and reduce personnel and duplicative overhead expenses. Since then, the Farm Service Agency has closed more than 370 county offices and cut its county office staff by about 28 percent. These reductions were achieved primarily by closing and consolidating smaller county offices and by reducing staff at larger county offices (U.S. Government Accountability Office 1999).

In terms of data, the US Office of Personnel Management (OPM) reports head counts of employees, while the Office of Management and Budget (OMB) reports total hours worked as measures of federal workforce (Wilson 2023).

Data on county level number of staff of different USDA offices back to 1996: https://www.fedscope.opm.gov/.

13 Impact of Crop Insurance on Retirement of Farmers

This section studies the impact of enrollment in the federal crop insurance in retirement benefits of farmers. To my knowledge, there is no prior study of this question.

This impact can be three-fold: the first channel is through the income-smoothing channel: The idea of this section is that the FCIP stabilizes net farm income across years. If stabilization raises average reported self-employment earnings over a worker's top 35 years, Primary Insurance Amount (PIA) under Social Security can rise. Conversely, if insurance mainly shifts income intertemporally without changing lifetime covered earnings, effects on PIA are small; the gain is more from lower variance (allowing later retirement or reduced precautionary saving). The second channel is through the wealth accumulation and upgrading the saving portfolio of farmers. When the stabilized cash flow can increases contributions to tax-advantaged retirement accounts (SEP-IRA, SIMPLE-IRA, solo 401(k)), or accelerates debt pay-down, it results in raising net wealth. At the same time, a greater downside protection can reduce precautionary cash balances and reallocate to long-term retirement assets. The final channel is through the decision over the timing of retirement. By reducing risk of business failure, FCIP may delay farm exit/retirement or permit later Social Security claiming (higher monthly benefits). Alternatively, if FCIP raises expected returns, some may reach target wealth sooner and retire earlier.

To measure these effects, I used Difference-in-differences around ARPA 2000 using pre-policy county differences in baseline insurability (e.g., crop mix, historical loss experience) as exposure.

I expect to find the following

- Positive but modest effect on Social Security benefits if FCIP raises long-run average covered earnings; stronger where pre-policy income volatility caused frequent low-earning years.
- Higher private retirement saving for insured farms with binding liquidity constraints pre-FCIP.
- Later claiming/exit if risk reduction keeps older farmers operating; earlier retirement is also plausible where insurance enables target-wealth completion—your event-study on exit/claiming will tell which margin dominates.

Ergodicity.[14] **Setup.** Let a farm household's retirement wealth evolve multiplicatively:

$$W_{t+1} = W_t(1 + r_t) + c_t,$$

where r_t is the net portfolio return and $c_t \geq 0$ is the end-of-year contribution funded from farm net income Y_t. With proportional saving $c_t = \alpha Y_t$ and small returns, the long-run (time-average) growth rate of multiplicative wealth is

$$g = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \log\left(1 + r_t + \frac{c_t}{W_t}\right) \approx \mu - \tfrac{1}{2}\sigma^2 + \phi(\bar{c}/W),$$

where $\mu = \mathbb{E}[r_t]$, $\sigma^2 = \mathrm{Var}(r_t)$, and $\phi(\bar{c}/W)$ captures the contribution channel (higher steady contributions raise g nonlinearly, while contribution gaps lower it). In non-ergodic (multiplicative) settings, volatility penalizes time-average growth through $-\frac{1}{2}\sigma^2$ and through contribution interruptions.

Crop insurance. Enrollment in FCIP changes the distribution of disposable farm income by (i) lowering variance $\mathrm{Var}(Y_t)$ and the probability of deep shortfalls, and (ii) imposing a premium π (net of subsidy). Let participation reduce the variance of contribution-relevant cash flow by $\Delta\sigma_Y^2 > 0$ and the probability of a *contribution gap* event ($c_t = 0$) by $\Delta p_{\text{gap}} > 0$. Then the *change in time-average growth* of retirement wealth is approximately

$$\Delta g \approx \underbrace{\tfrac{1}{2}\beta^2 \Delta\sigma_Y^2}_{\text{variance relief}} + \underbrace{\gamma\, \Delta p_{\text{gap}}}_{\text{fewer missed contributions}} - \underbrace{\eta\, \pi}_{\text{premium drag}},$$

where β maps income volatility into contribution or portfolio volatility, $\gamma > 0$ is the growth penalty per missed contribution (depending on horizon and compounding), and η is the marginal impact of a one-dollar premium on g. FCIP raises time-average retirement wealth if the risk-reduction terms exceed the premium drag. This prediction holds independently of expected-utility curvature and provides a clear rationale for subsidies when contribution volatility is first-order.

Data. ARMS (farm-household finances, insurance use, retirement saving, age), RMA Summary of Business (county-crop coverage, liabilities, loss ratios), NASS yields; optionally SSA microdata (restricted) for claiming and PIA.

1) Identify exogenous variation in FCIP intensity.

- Event studies around **ARPA 2000** (large subsidy increase); interact with pre-policy county insurability (crop mix, historical losses).
- Instruments from **coverage-level subsidy schedule** kinks (price variation by coverage band) or product rollouts (revenue vs. yield plans).
- Border or timing discontinuities in availability or take-up.

2) Estimate the three inputs to Δg.

- **Variance relief ($\Delta\sigma_Y^2$):** Regress (by operator) the variance of real net farm income on FCIP participation or intensity using panel methods (or estimate at county×crop and map to households). Report within-farmer changes pre/post shock.

[14] I heard about the idea of Ergodicity in insurance from Ahmad Salahnejhad, when sharing his insights from his field of actuarial science.

- **Contribution gaps (Δp_{gap}):** Model $1[c_t = 0]$ (no IRA/SEP contribution year) or "downshift to cash/withdrawal" as a function of FCIP intensity; estimate marginal effects using the same instrument.
- **Premium drag (π) and (η):** Use RMA farmer-paid premium (after subsidy) per insured acre times insured acres; map to household budget. Calibrate η from the elasticity of contributions with respect to disposable income (from ARMS).

3) Map to Δg and interpret magnitudes.

$$\widehat{\Delta g} = \tfrac{1}{2}\hat{\beta}^2 \widehat{\Delta \sigma_Y^2} + \hat{\gamma}\,\widehat{\Delta p_{\text{gap}}} - \hat{\eta}\,\hat{\pi}.$$

- Estimate β from a first-stage linking income volatility to contribution or portfolio volatility.
- Calibrate γ by simulating how a missed contribution at age a reduces retirement wealth at R given observed returns (simple compounding).
- Report effects for ages **55+** (retirement-relevant window) and heterogeneity by leverage or liquidity.

4) Secondary outcomes (mechanisms).

- **Claiming/exit:** hazard models of Social Security claiming or farm exit versus FCIP intensity.
- **Portfolio shift:** risk share of financial assets; debt paydown.

5) Robustness.

- Placebo tests on younger cohorts; pre-trends; alternative volatility measures (winsorized, GARCH-style).
- Sensitivity to λ (marginal cost of public funds) if translating Δg into a social objective.

---

14 Effect of Subsidies on Crop Diversification and Irrigation

I used a between-crop strategy using the universe of insurance data at the county-crop level and compare the irrigation and diversifying choices among crops enrolled in different percentages of insurance.

15 Optimal Level of Subsidy

The FCIP initially started off without premium subsidies [15]. But farmers were found to be unwilling to purchase unsubsidized multiperil crop insurance. The Department of Agriculture initiated the premium subsidies to encourage the participation of farmers (Glauber 2004)read jan. 21.2026. In this section, I address the economic efficiency of the subsidized crop insurance, considering the costs and benefits of its provision. On the one hand, government provision of insurance in the context of market

[15] I appreciate Professor Ali Hortaçsu for letting me attend his IO-2 class to refresh my memory for the demand estimation concepts.

imperfection and high correlated risk facilitates insurance that private markets do not otherwise provide. This is the case, as according to Quiggin (1994) read Jan 2026, one problem that distinguishes crop insurance from all other types of insurance, is small potential gain from pooling of similar risks; a problem that makes it financially infeasible to provide for the private sector. While in cases like life insurance, individual risks are uncorrelated, all farms in each district are likely to experience adverse conditions at the same time, and correlations are likely to be high even for large regions. On the other hand, the asymmetric information problems may generate costs that outweigh the benefits. Social costs might arise if farmers alter their behavior in ways that lower expected profits net of insurance indemnities.

To evaluate the welfare impact of the subsidy, I quantify the costs and benefits of farms enrolling in different percentages of insurance coverage. The primary benefit is better income smoothing as more protection is offered against uncertainty. The primary cost is the financial cost of subsidies. However, secondary benefits and costs may arise if diversification affects costs and expected yields. Taking them into account is similar to the job market paper of Adam Solomon (Klosin and Solomon 2024)[16]. The optimal policy is a simple variant of (Baily 1978)'s formula for the optimal level of social insurance: the optimal subsidy level equates the marginal gains from a smoother consumption path with the marginal cost, measured by the behavioral response in effort.

I estimate the optimal level of crop insurance subsidies using various measures of demand in the vein of Barry K Goodwin (1993) and E. O'Donoghue (2014). In the empirical literature, Hennessy and Babcock (1997), Wang et al. (1998), Adhikari, Knight and Belasco (2012) estimate the welfare benefits from risk protection provided by the crop insurance (Jialing Yu 2017).

In this section, I first calculate the optimal level of subsidy using the a reduced form sufficient statistics. Then I build a structural model to incorporate systemic risk, spatial correlation, and credit/adoption externalities that are important in agriculture but hard to capture with pure reduced form. Similar to the previous parts of this paper, I use county-level data in the analysis. This means treating each county as a single farm, as notes by Barry K Goodwin (1993).

15.1 Sufficient Statistics

In this section, I apply the sufficient statistics approach (Baily 1978; Chetty and Finkelstein 2013; Lashkaripour 2021; Lashkaripour and Lugovskyy 2023) to calculate the optimal level of crop subsidy. Applying sufficient statistics would allow this study to estimate the welfare effects of policy changes in terms of estimable elasticities, and performing policy evaluation without making parametric assumptions or estimating the structural primitives of fully specified models. However, this modeling is possible with two assumptions: first, the policy change is small. This would allow using first-order approximations to be precise. And second, government policy is the only wedge between the actual equilibrium and the first-best equilibrium. This means there are no externalities or internalities other than of government that would change by behavioral responses to policy reforms (Kleven 2021, page 517). Working with sufficient statistics also limits the scope to analyze marginal changes in policy. An alternative traditional approach to finding the optimal point is to identify the structural primitives and perform welfare analysis by simulating alternative policies. In principle, that structural approach

[16] Adam Solomon generously shared with me the replication package of his job market paper before its publication. I thank him for his trust and support.

is an ideal model, but in practice, it is difficult to capture all the primitives of the model (Chetty and Finkelstein 2013, p. 143). Hence, I work with sufficient statistics rather than a fully specified structural model.

The closest prior work to my analysis is the Ph.D. dissertation of Jialing Yu (2017, ch. 4), that characterizes the marginal welfare effect consequences and optimal policy design for the FCIP. Jialing Yu (2017) reports that under current policy settings, the net marginal welfare effects of coverage levels and subsidy rates are negative, and both are greater than their optimal levels. However, if only consider the welfare of the farmers, the marginal welfare benefit of risk protection approximately equals to the welfare cost of moral hazard, indicating that the risk protection level seems to be appropriate in this case. The innovative use of Bootstrap method to introduce uncertainties in parameters, stands out in Jialing Yu (2017)'s work.

In modeling the welfare implications of insurance subsidies and farmers' choices about them, I take into account that most US corn and soybean acres are insured under revenue-based policies, and changes in insurance design or subsidies can shift coverage levels and total insured acreage. However, the relative share of corn vs. soybeans is only modestly affected, as both crops are similarly responsive to insurance incentives (Maisashvili, H. L. Bryant, and Jones 2020; Maisashvili, Fischer, and H. Bryant 2023; Jisang Yu, A. Smith, and Sumner 2018). Some studies note that insurance may encourage expansion onto lower-quality land, potentially affecting crop rotation and environmental outcomes, but the overall impact on crop mix remains small (Claassen, Langpap, and Wu 2017; Jisang Yu, A. Smith, and Sumner 2018).

Given this, I start off with modeling the insurance enrollment without an acreage substitution between corn and soybeans (model A). I estimate robust, IV-backed crop-specific demand for insurance; and compute the optimal subsidy rate s^*, with sensitivity check on λ, administrative costs. Then I run the quick upper-nest test. If α_k are small and counterfactual mix shifts are less than 1 pp, I explain why nested logit is necessary for my question. If unexpectedly, α_k are meaningful or policy scenarios target one crop asymmetrically, then I promote the nest model.

15.2 Model A (v1)- Crop-specific Demand and Optimal Subsidy in Absence of Acreage Substitution

I index counties by c, years by t, and crops by $k \in \{\text{corn}, \text{soybeans}\}$. For each crop k, the potential market size in a county-year is the planted area A_{ctk}. Farmers choose whether to insure each acre under a subsidized contract. Let P_{ctk} denote the gross actuarial premium per acre and let $s \in [0, 1)$ be the premium subsidy rate; the price farmers pray is $(1 - s)P_{ctk}$. Insured acres are $Q_{ctk}(s)$ and the insured share is $S_{ctk}(s) = Q_{ctk}(s)/A_{ctk}$.

Demand. I model the insured share for each crop separately using an acre–level logit as follows:

$$S_{ctk}(s) = \frac{exp(\nu_{ctk}(s))}{1 + exp(\nu_{ctk}(s))} = \frac{\exp\Big(\alpha_{ck} + \tau_{tk} - \theta_k \ln(1 - s) - \phi_k \ln P_{ctk} + X'_{ctk}\beta_k\Big)}{1 + \exp\Big(\alpha_{ck} + \tau_{tk} - \theta_k \ln(1 - s) - \phi_k \ln P_{ctk} + X'_{ctk}\beta_k\Big)}, \tag{28}$$

where α_{ck} and τ_{tk} are county and year fixed effects, X_{ctk} collects controls (weather, loss experience, income, rotation indicators, etc.), $\theta_k > 0$ captures sensitivity to the net price $\ln(1 - s)$, and $\phi_k > 0$

captures sensitivity to the premium level (risk/plan mix). Identification of θ_k requires exogenous variation in net price. I instrument $\ln(1-s)$ using IVs developed in the previous sections. Premium variation is absorbed by $\ln P_{ctk}$ and the fixed effects.

Welfare. Let $v_{ctk}(s)$ denote the logit index in 28. Under the usual logit aggregation, expected consumer surplus per potential acre is

$$CS_{ctk}(s) = \ln\big(1 + \exp\{v_{ctk}(s)\}\big),$$

which we treat in money–metric units (normalizing the utility scale). For a change in the subsidy from a reference level s_0 to s, the county–crop–year welfare is

$$W_{ctk}(s) \;=\; A_{ctk}\Big[CS_{ctk}(s) - CS_{ctk}(s_0)\Big] - (1+\lambda)\, s\, P_{ctk}\, Q_{ctk}(s) - \kappa_A\, Q_{ctk}(s), \tag{29}$$

where λ is the marginal excess burden of public funds and κ_A collects any per–insured–acre resource cost (e.g., administrative overhead one chooses to treat as real resource use). Aggregate welfare for crop k is $W_k(s) = \sum_{c,t} W_{ctk}(s)$. Because we model crops as independent markets (no acreage substitution), overall welfare is $W(s) = W_{\text{corn}}(s) + W_{\text{soybeans}}(s)$ and the crop–specific optimal subsidies solve

$$s_k^* \in \arg\max_{s\in[0,1)} W_k(s) \qquad (k = \text{corn, soybeans}).$$

This is a transparent "sufficient–statistics / inclusive–value" approach: estimate causal demand in (28) with respect to the net price, then conduct a grid search over s using (29). It admits straightforward sensitivity analysis in λ and κ_A and can be implemented crop–by–crop with county and year fixed effects.

Interpretation. In this specification each crop is a separate market in which the subsidy shifts the insured share but does not alter total planted acres. This is appropriate when insurance–induced reallocation between corn and soy is empirically modest and the main action is in insurance coverage intensity within each crop.

move this sentence to a better place: I develop equations for the welfare consequences of the FCIP program that are a function of high-level elasticities. Risk protection is usually attributed to risk spreading and risk pooling(Nelson and Loehman 1987).

Predetermined parameters from the literature. I chose policy primitives based on the previous literature as follows. The marginal excess burden of public funds, λ, takes the values of 0.25. This means than every \$1 of government spending costs society \$1.25, once distortion effect of tax is included. I chose this value, as the public finance literature typically estimates the marginal excess burdern to be between 0.15 and 0.30 for the US. To do: for robustness check, I test results for values of λ equal to 0 (no excess burden), 0.15,0.30 and 0.50 (conservative). Similarly, I set the per-acre administrative cost of insurance provision, κ, to 0. This choice was because papers studying optimal subsidy often set it so as a baseline. The empirical estimates of true resource cost is \$1-5 per insured acre.To do: For robustness check, set kappa to 1, 3, and 5.

Table 4-1. Summary of the parameterization for MC simulation

Parameter	Distribution assumption	Distribution parameters		Sources/Methods
RISK ATTITUDE				
CARA: γ	Uniform: Unif(a,b)	a=0.00001, b=0.00005		Hennessy and Babcock (1997)
CRRA: $\bar{\gamma}$	Normality	Mean=0.77 Std. Err.=0.64		Barham et al. (2014)
LOADING FACTOR ι				
CAT loads	Uniform	a=0.07, b=0.46		Coble et al. (2010)
A&O cost loads	Normality	Mean=0.170 Std. Err.=0.0138 95% Conf. Interval: (0.143, 0.197)		A&O data: Shields (2010 and 2015), Bootstrapping
DEADWEIGHT LOSS: φ	Uniform	a=0.2, b=0.5		Alston and Hurd (1990)
Elasticities		Mean	Std. Err.	
Moral hazard: First-order moment $\eta_{R,b}$,		-0.817	0.001	Chapter 3
Moral hazard: Second-order moment $\eta_{\sigma,b}$		0.241	0.183	Chapter 3
Adverse selection: $\eta_{\mathrm{l},\tau}$		-0.0966	0.0464	Chapter 3
Effect of subsidy rate: $\eta_{\tau,s}$		2.83	442	O'Donoghue (2014)
Summary statistics				
Profit of low state: $\bar{\pi}_L$		390.348	0.378	Bootstrapping
Coefficient of variance: c_v		0.134	0.009	Bootstrapping
Profit gap between two states: $\bar{\Delta}$		147.143	0.117	Bootstrapping
Inverse mill's ratio of low state: λ_L		2.690	0.003	Bootstrapping
Probability of not getting indemnified: $\bar{e}$		0.949	0.000	Bootstrapping
Correlation coefficient: ρ_1		0.003	0.002	Bootstrapping
Correlation coefficient: ρ_2		-0.009	0.002	Bootstrapping
Expected indemnity payment of the marginal insureds: τ_b		26.623	0.048	Bootstrapping

Note: The standard errors of summary statistics and elasticities are bootstrap standard errors.

Figure 42: Summary of the parameterization for MC simulation (Table 4 in Jialing Yu (2017))

In the welfare simulations, I set the marginal excess burden of public funds to $\lambda = 0.25$, following (Jialing Yu 2017) and classic estimates in (Ballard, Shoven, and Whalley 1985; Browning 1987; Hendren and Sprung-Keyser 2020). I set the per-acre administrative resource use $\kappa_A = 0$ as a baseline; Section xx reports robustness to $\kappa_A > 0$. I evaluate welfare over a policy-relevant subsidy grid $s \in [0, 0.90]$ in one-percentage-point increments.

E. O'Donoghue (2014)

15.3 Model A (v2): Crop-Specific Insurance Demand and Optimal Subsidy

This section develops a crop-specific model of insurance demand and welfare grounded in the sufficient-statistics approach to social insurance (Baily 1978; Chetty and Finkelstein 2013) and the random utility framework for discrete choice demand (Berry and Haile 2021, sec. 3.2). The goal is to recover the welfare effects of premium subsidies using estimable demand elasticities and fiscal parameters, without fully modeling production or consumption.

I index counties by c, years by t, and crops by $k \in \{\text{corn}, \text{soybeans}\}$. The potential market size in a county–year–crop is the planted acreage A_{ctk}. Each acre chooses between enrollment ($j = 1$) and non-enrollment ($j = 0$) in a subsidized insurance contract. Let P_{ctk} be the gross actuarially fair premium per acre, and $s \in [0, 1)$ the subsidy rate, so the farmer-paid premium is $(1 - s)P_{ctk}$. Insured acreage is $Q_{ctk}(s)$, with share $S_{ctk}(s) = Q_{ctk}(s)/A_{ctk}$.

Demand. I model insurance enrollment using an acre-level logit, derived from a parametric random utility model. Let the utility of insurance option j to acre i in market (c, t, k) be:

$$u_{ijctk} = x'_{jctk}\beta_k - \theta_k \ln p_{jctk} + \xi_{jctk} + \varepsilon_{ijctk}, \tag{30}$$

where $p_{jctk} = (1 - s)P_{ctk}$ is the net premium, x_{jctk} includes observed covariates (weather, loss

experience, income proxies, etc.), ξ_{jctk} is an unobserved demand shifter common across acres, and ε_{ijctk} is an idiosyncratic shock.

From Utility to Aggregate Demand. Assuming the idiosyncratic shocks ε_{ijctk} follow an i.i.d. Type I Extreme Value distribution, the individual utility specification in Equation 30 yields a closed-form expression for the aggregate probability of insurance enrollment[17]. Specifically, the share of insured acres in market (c, t, k) corresponds to the logit choice probability associated with the utility difference between enrollment and non-enrollment, as shown in Equation 31. Assuming $\varepsilon_{ijctk} \sim$ Type I Extreme Value implies the logit choice probability:

$$S_{ctk}(s) = \frac{\exp(v_{ctk}(s))}{1 + \exp(v_{ctk}(s))}, \tag{31}$$

with the indirect utility index:

$$v_{ctk}(s) = \alpha_{ck} + \tau_{tk} - \theta_k \ln(1 - s) - \phi_k \ln P_{ctk} + X'_{ctk}\beta_k. \tag{32}$$

Here, α_{ck} and τ_{tk} are county–crop and year–crop fixed effects that absorb latent demand shifters ξ_{ctk}, including local insurance culture and agronomic conditions. Parameters $\theta_k > 0$ and $\phi_k > 0$ capture sensitivity to the net and gross premiums, respectively. This structure nests naturally within the canonical discrete choice model of Berry and Haile (2021), with a binary product set and a well-defined outside option (non-enrollment). Identification of θ_k exploits exogenous variation in $\ln(1 - s)$, instrumented using policy-driven subsidy differences across counties and crops. Variation in gross premiums is absorbed by $\ln P_{ctk}$ and fixed effects.

Welfare. Under the logit structure, expected consumer surplus per potential acre is given by the log-sum formula

$$CS_{ctk}(s) = \ln(1 + \exp(v_{ctk}(s))), \tag{33}$$

which has a money-metric interpretation under quasi-linear utility (Chetty and Finkelstein 2013). For a change in the subsidy from a reference level s_0 to s, county–crop–year welfare is

$$W_{ctk}(s) = A_{ctk}\Big[CS_{ctk}(s) - CS_{ctk}(s_0)\Big] - (1 + \lambda)\, s\, P_{ctk} Q_{ctk}(s) - \kappa_A Q_{ctk}(s), \tag{34}$$

where λ denotes the marginal excess burden of public funds and κ_A captures any real per-acre administrative resource cost of insurance provision. Aggregate welfare for crop k is $W_k(s) = \sum_{c,t} W_{ctk}(s)$. Because Model A treats planted acreage as exogenous with respect to insurance incentives, crops are modeled as independent markets. Overall welfare is therefore

$$W(s) = W_{\text{corn}}(s) + W_{\text{soybeans}}(s),$$

and the crop-specific optimal subsidy satisfies

$$s_k^* \in \arg\max_{s \in [0,1)} W_k(s), \qquad k \in \{\text{corn}, \text{soybeans}\}.$$

[17] The i.i.d. extreme value assumption should be interpreted as applying to unobserved heterogeneity in insurance valuation across infinitesimal acreage units within a county–crop–year market, conditional on observed covariates and fixed effects, rather than as literal independence of farm-level decisions.

Model A implements a sufficient-statistics approach to social insurance in which the welfare effects of subsidies are fully summarized by insurance demand responses and fiscal parameters. The model abstracts from acreage reallocation across crops, which is appropriate when insurance-induced substitution between corn and soybeans is empirically small. In this setting, the primary margin of adjustment is insurance coverage intensity within each crop, rather than crop choice itself.

Mapping Model A to the Sufficient-Statistics Framework. The welfare analysis in Model A can be interpreted through the sufficient-statistics approach to optimal social insurance developed by Baily (1978) and Chetty (2006) Read on Feb 2, 2026. In this framework, optimal policy depends only on observable behavioral responses and fiscal parameters, rather than on the full set of structural primitives governing preferences, technology, or risk.

In Model A, the subsidy s affects welfare through two channels. First, it raises insurance enrollment, generating benefits from improved risk protection that are captured by consumer surplus. Second, it imposes a fiscal cost through government-financed premium subsidies. These components map directly into Chetty's notation as follows.

The behavioral benefit from a subsidy s in a county–crop–year (c, t, k) is given by expected consumer surplus per potential acre,

$$B_{ctk}(s) = A_{ctk} \cdot CS_{ctk}(s) = A_{ctk} \cdot \ln(1 + \exp(v_{ctk}(s))) \,, \tag{35}$$

where $v_{ctk}(s)$ is the logit mean utility index defined in Equation (31). This term summarizes the welfare gains from increased insurance take-up induced by the subsidy.

The fiscal cost of the subsidy is

$$C_{ctk}(s) = s \cdot P_{ctk} \cdot Q_{ctk}(s) = s \cdot P_{ctk} \cdot A_{ctk} \cdot S_{ctk}(s), \tag{36}$$

where $Q_{ctk}(s)$ denotes insured acreage and $S_{ctk}(s)$ the insured share. Accounting for the marginal excess burden of public funds, the social cost of financing the subsidy is $(1 + \lambda) \cdot C_{ctk}(s)$.

Aggregating across counties and years yields total welfare for crop k:

$$W_k(s) = \sum_{c,t} \left[B_{ctk}(s) - (1 + \lambda) C_{ctk}(s) - \kappa_A Q_{ctk}(s) \right] ,$$

which corresponds exactly to the sufficient-statistics welfare objective in Chetty (2006).

Sufficient-Statistics Characterization of the Optimal Subsidy. The optimal subsidy s_k^* maximizes $W_k(s)$ and therefore satisfies the first-order condition

$$\frac{dW_k(s)}{ds} = 0. \tag{37}$$

Differentiating the welfare expression yields

$$\sum_{c,t} \left[A_{ctk} \cdot \frac{dCS_{ctk}(s)}{ds} - (1 + \lambda) \left(P_{ctk} \cdot Q_{ctk}(s) + s \cdot P_{ctk} \cdot \frac{dQ_{ctk}(s)}{ds} \right) - \kappa_A \cdot \frac{dQ_{ctk}(s)}{ds} \right] = 0. \tag{38}$$

Under the logit structure, the marginal consumer-surplus term satisfies

$$\frac{dCS_{ctk}(s)}{ds} = S_{ctk}(s) \cdot \frac{dv_{ctk}(s)}{ds},$$

so that the optimal subsidy equates the marginal welfare gain from increased insurance coverage to the marginal social cost of public funds. Importantly, this condition depends only on observable enrollment responses $\frac{dQ_{ctk}(s)}{ds}$ and known fiscal parameters $(P_{ctk}, \lambda, \kappa_A)$, rather than on unobserved preferences or risk aversion.

Consistent with the sufficient-statistics logic of Chetty (2006), Model A therefore permits welfare analysis and optimal policy design using estimated insurance demand responses alone, without requiring explicit measurement of consumption smoothing or structural utility parameters.

15.4 Control Function IV for Nonlinear Insurance Demand in Model A

In the crop–specific demand model, the insured share for crop k in county c and year t is

$$S_{ctk} = G\big(\alpha_{ck} + \tau_{tk} - \theta_k \ln(1 - s_{ct}) - \phi_k \ln P_{ctk} + X'_{ctk}\beta_k + u_{ctk}\big),$$

where $G(\cdot)$ is the logit link, s_{ct} is the premium subsidy rate, P_{ctk} is the gross actuarial premium, and X_{ctk} collects exogenous controls. The parameter of interest is θ_k, the causal response of insurance demand to the net price for farmers, $\ln(1 - s_{ct})$.

Since counties with unobserved risk, information, or political influence may both receive higher subsidies and display different insurance demand, $\ln(1 - s_{ct})$ is endogenous. Let z_{ct} denote the set of instruments that shift the subsidy schedule but are plausibly orthogonal to the unobserved demand shock u_{ctk} in the short run. In linear models I developed and estimated in Section xx, I used two–stage least squares (2SLS). In a nonlinear logit or fractional logit, however, 2SLS is no longer valid. A convenient solution is the *control–function* (or two–stage residual inclusion) approach; see Wooldridge (2015) read Jan 31, 2026 for formal treatments and applications. Even when the CF approach is identical to a 2SLS estima tor, the CF perspective has a couple of attractive features. First, the CF approach produces a simple Hausman (1978) test that compares OLS and 2SLS, and the test is easily made robust to heteroskedasticity and cluster correlation (including serial correlation in a panel data setting) of unknown form. Second, the CF approach parsimoniously handles fairly complicated models that are nonlinear in endogenous explanatory variables (Wooldridge 2015, p. 421).

Step 1: first–stage for the endogenous regressor Project the endogenous net price on instruments and exogenous variables:

$$\ln(1 - s_{ct}) = Z'_{ct}\pi_1 + X'_{ctk}\pi_2 + \alpha_c + \tau_t + v_{ct},$$

where Z_{ct} includes z_{ct} and any other excluded instruments, and (α_c, τ_t) are county county or state? in the ols and spatial regressions, I did state fixed effects and year fixed effects. Under the usual IV relevance and exogeneity conditions, the residual $\hat{v}_{ct}$ is an estimate of the component of $\ln(1 - s_{ct})$ that is correlated with the structural error u_{ctk}.

Step 2: structural demand with a control function The control–function approach augments the structural logit with the first–stage residual:

$$S_{ctk} = G\big(\alpha_{ck} + \tau_{tk} - \theta_k \ln(1 - s_{ct}) - \phi_k \ln P_{ctk} + X'_{ctk}\beta_k + \rho_k v_{ct} + \varepsilon_{ctk}\big).$$

If the joint distribution of (u_{ctk}, v_{ct}) is such that $E[u_{ctk} \mid \ln(1 - s_{ct}), Z_{ct}, X_{ctk}] = h(v_{ct})$ and $h(\cdot)$ is well approximated by a linear term, then conditioning on v_{ct} renders $\ln(1 - s_{ct})$ *conditionally exogenous*. Estimating this equation by maximum likelihood (fractional logit) yields consistent estimates of $(\theta_k, \phi_k, \beta_k, \rho_k)$.

The coefficient ρ_k is an endogeneity parameter: a joint test of $H_0 : \rho_k = 0$ is an exogeneity test in the spirit of Rivers and Vuong (1988). If ρ_k is statistically and economically small, the simpler single–equation logit without a control function may be adequate. If ρ_k is non-zero, then the control–function logit recovers the causal effect of the net price on the insured share, while preserving the nonlinear probability structure used later for welfare calculations.

For policy counterfactuals (changes in s holding instruments fixed), one treats $\hat{v}_{ct}$ as fixed county–year heterogeneity: the net price $\ln(1 - s)$ is moved along the estimated demand curve, while the control function continues to absorb the endogenous component of the unobserved demand shock.

15.5 Results of Model A

I first estimate a baseline model without the use of instrumental variables to show raw correlations and fit. Then I estimate the model with the IVs to report the causal estimates. Finally, I test the adequacy of Model A in modeling, and based on that, decide to add the Model B below or not.

Figure 43 summarize the welfare implications of varying the premium subsidy rate s under the crop-specific Option A model, where insurance demand is estimated separately for corn and soybeans and no acreage substitution across crops is allowed. In this environment, changes in the subsidy affect welfare exclusively through: (i) induced take-up within each crop; (ii) the associated change in consumer surplus; and (iii) the fiscal and administrative cost of subsidizing premiums. The graph 43a plots total welfare,

$$W(s) = W_{\text{corn}}(s) + W_{\text{soybeans}}(s),$$

when a single subsidy rate applies to both crops. Welfare is normalized to zero at $s = 0$. The curve declines sharply and approximately linearly as the subsidy increases. Over the range $s \in [0, 0.9]$, total welfare becomes increasingly negative and does not exhibit an interior maximum. Therefore, the welfare-maximizing subsidy lies at the lower boundary, $s^* = 0$. This pattern arises because the fiscal cost of subsidies grows proportionally with take-up, whereas the consumer-surplus gain from additional insured acres grows slowly. Demand elasticities for both crops are modest, so increases in the subsidy induce relatively small increases in insured acreage. As a result, the welfare gain from improved risk smoothing is dominated by the cost of public funds, particularly once the marginal excess burden of taxation is incorporated ($\lambda = 0.25$ in the baseline). The nearly linear welfare decline suggests that each additional percentage point of subsidy destroys welfare at roughly the same marginal rate, with little curvature in the welfare function. In sufficient-statistics terms, the ratio of fiscal cost to consumer-surplus gain exceeds one for all positive subsidy rates.

Figure 43b decomposes total welfare into its corn and soybean components. Both crops exhibit the same qualitative pattern: welfare is decreasing in the subsidy rate, with no interior optimum within the empirically relevant range. However, soybeans show a smaller welfare decline than corn. Several mechanisms may explain this difference:

1. **Different demand elasticities.** If soybean insurance demand is less sensitive to s, increases in the subsidy induce fewer additional insured acres, moderating the growth in government cost.

2. **Different premium levels.** If soybean premiums P_{ctk} are typically lower, the fiscal cost of subsidizing each acre is correspondingly lower.

3. **Different acreage bases.** Corn occupies more acres than soybeans in many regions. A uniform subsidy thus generates larger total fiscal expenditures for corn simply because more insured acres receive the subsidy.

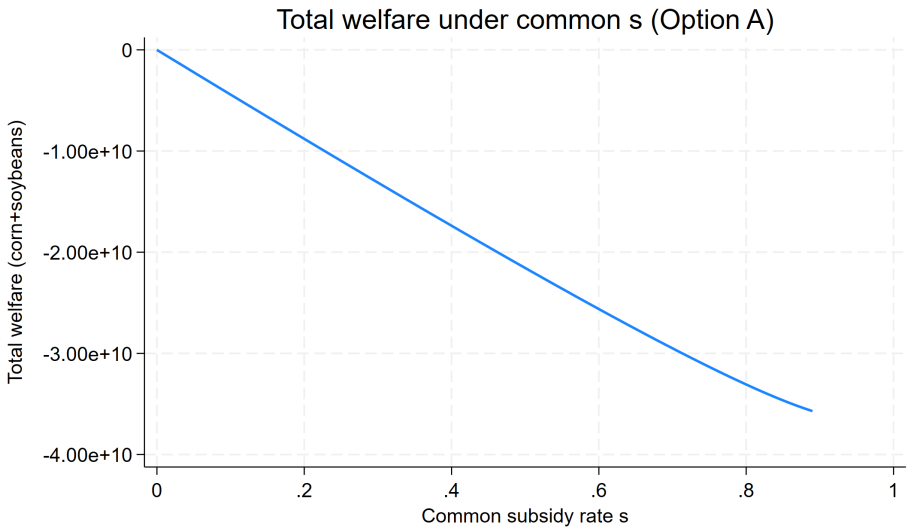
Despite these differences in magnitude, neither welfare curve exhibits a turning point. In both crops, the marginal cost of public funds exceeds the marginal consumer-surplus benefit of additional insurance take-up. This is consistent with recent empirical evidence showing that insurance demand for major U.S. row crops is relatively inelastic, so large subsidies primarily transfer public resources to farmers without generating proportionate increases in risk protection.

Taken together, the figures indicate that under the assumptions of Model A, a uniform subsidy applied to corn and soybeans is welfare-reducing at all positive levels. The results imply that federal premium subsidies during the study period exceeded the welfare-maximizing level, even before incorporating moral hazard, adverse selection, or environmental externalities. Including such externalities would only strengthen the case against substantial subsidies.

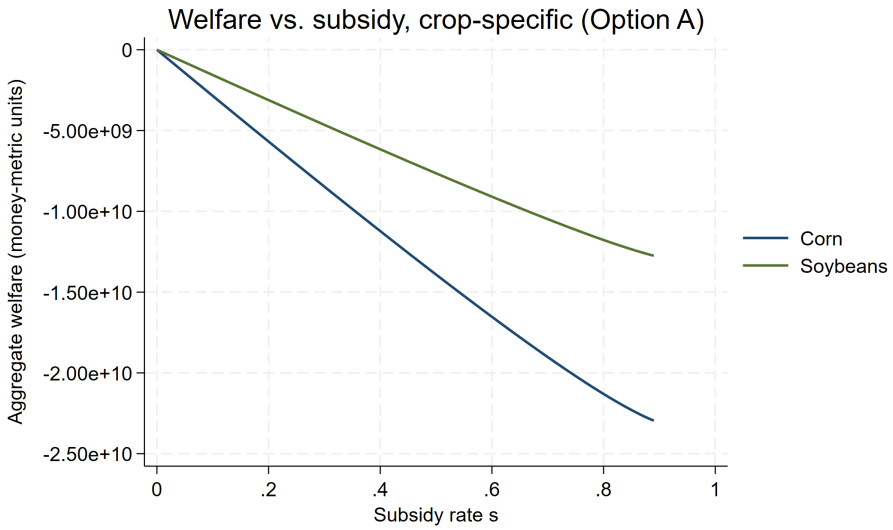
15.6 Results of Model A with IV

Figure 44 reports results for the sufficient statistics exercise. The first panel plots crop-specific welfare as a function of the subsidy rate s using the control function IV (CF-IV) estimates of insurance demand. Because the CF-IV procedure corrects for the endogeneity of the net price $\ln(1-s)$, the implied demand elasticities differ markedly from the baseline no–IV specification. For both corn and soybeans, welfare initially declines as subsidies increase from zero, reflecting rising fiscal costs and only moderate consumer–surplus gains. However, for soybeans the welfare schedule reaches an interior minimum around $s \approx 0.55$ and then begins to rise at higher subsidy levels. This non–monotonicity indicates that, once endogeneity is corrected, the model implies a relatively strong insurance demand response at low net prices, generating substantial consumer–surplus improvements that partially offset the government cost at high subsidy rates. The resulting curvature underscores how optimal–subsidy calculations can change when the price endogeneity in insurance participation is properly addressed.

The second panel aggregates welfare across both crops under a common subsidy rate imposed on the entire market. Unlike the crop–specific case, the combined schedule is dominated by the much larger corn sector and therefore exhibits a more monotonic pattern. Total welfare falls sharply from $s = 0$ to approximately $s = 0.40$, where it reaches its minimum, and then flattens without displaying the strong rebound observed for soybeans alone. At high subsidy levels, the continued welfare losses in corn outweigh the modest gains implied by the soybean demand estimates, producing a relatively smooth and predominantly downward–sloping welfare profile. These results highlight that welfare conclusions can differ substantially between crop–specific and uniform subsidies, and that the CF–IV correction affects the two cases in distinct ways due to differences in market size and estimated demand elasticities.

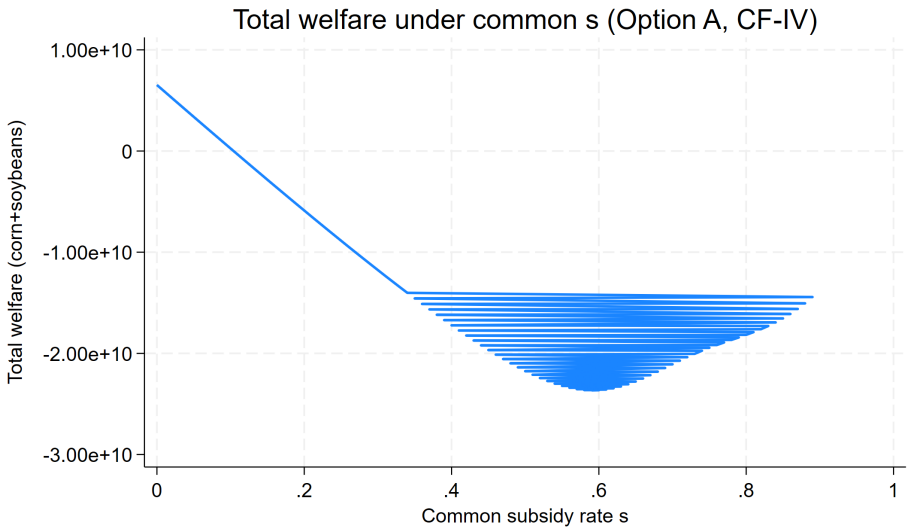


(a) Two crops combined

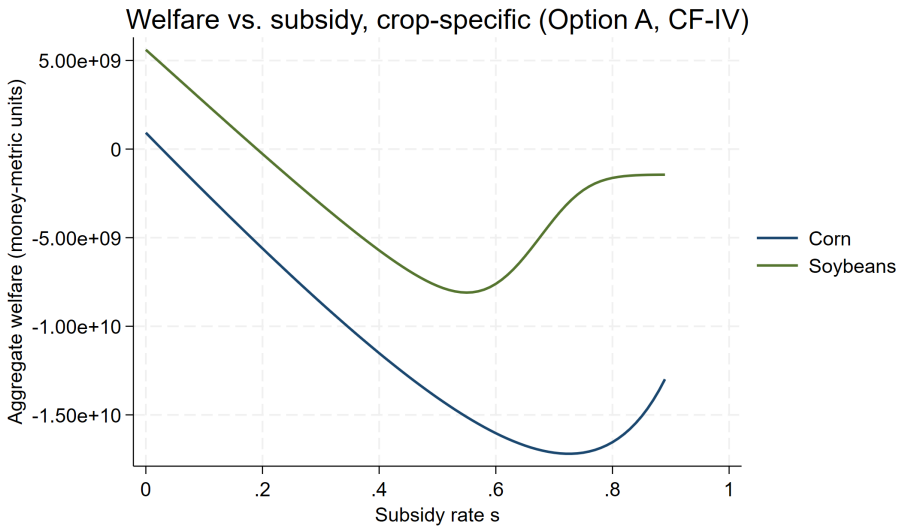


(b) Crop-specific

Figure 43: The change with welfare measured with consumer surplus with change in crop insurance subsidy. The county-level data is from U.S. Department of Agriculture, Risk Management Agency (RMA) (2025).



(a) Two crops combined



(b) Crop-specific

Figure 44: The change with welfare measured with consumer surplus with change in crop insurance subsidy. The county-level data is from U.S. Department of Agriculture, Risk Management Agency (RMA) (2025).

15.7 Model B(v4): Acreage Substitution and Spatial Equilibrium with Crop Insurance

This section extends Model A by allowing farmers to reallocate acreage across crops in response to insurance incentives and closes the model with a competitive land market. The resulting framework delivers a spatial equilibrium in which crop insurance subsidies affect insurance take-up, crop allocation, and land rents within counties.

Environment. Consider counties $c \in \mathcal{C}$, crops $k \in \mathcal{K}$, and years t. Each county has a fixed agricultural land endowment $\bar{A}_{ct}$ that can be allocated across crops:

$$\sum_{k \in \mathcal{K}} A_{ckt} = \bar{A}_{ct}.$$

Output prices p_{kt} are taken as given. Yield per acre for crop k in county c is stochastic and given by

$$y_{ckt} = \mu_{ckt} + \eta_{ckt},$$

where η_{ckt} captures weather-related production risk.

Crop insurance. For each (c, k, t), farmers may purchase insurance coverage indexed by $d \in \mathcal{D}$, where $d = 0$ denotes no insurance. Let $I_{ckt}(d, y)$ denote indemnity payments and $\pi^{AF}_{ckt}(d)$ the actuarially fair premium. A subsidy schedule $s_{ckt}(d)$ reduces the farmer-paid premium to

$$P_{ckt}(d) = (1 - s_{ckt}(d))\, \pi^{AF}_{ckt}(d).$$

In addition, farmers face a non-price insurance access cost τ_{ict} that varies across counties due to administrative frictions and insurance infrastructure.

Per-acre profit. Let r_{ct} denote the competitive land rental rate in county c. Per-acre profit from planting crop k and choosing insurance coverage d is

$$\pi_{ickt}(d) = p_{kt} y_{ckt} - g_{ckt} - r_{ct} + I_{ckt}(d, y_{ckt}) - P_{ckt}(d) - \tau_{ict} \mathbf{1}\{d \neq 0\},$$

where g_{ckt} denotes non-land production costs.

Insurance choice. Farmers are risk averse with Bernoulli utility $u(\cdot)$. Conditional on planting crop k in county c, the farmer chooses insurance coverage to maximize expected utility. Let

$$\bar{V}_{ckt}(d) = \mathbb{E}\left[u\left(\pi_{ickt}(d)\right)\right]$$

denote the systematic component of expected utility. Assuming i.i.d. Type-I extreme value preference shocks over insurance choices, the probability of purchasing insurance for crop k in county c is

$$\sigma_{ckt} = 1 - \frac{\exp\left(\bar{V}_{ckt}(0)\right)}{\sum_{d \in \mathcal{D}} \exp\left(\bar{V}_{ckt}(d)\right)}.$$

Crop choice and acreage allocation. Let the expected indirect utility from planting crop k in county c incorporate the optimal insurance decision:

$$W_{ckt} = \mathbb{E}\left[\max_{d \in \mathcal{D}} u\left(\pi_{ickt}(d)\right)\right].$$

Assuming random utility over crop choices, acreage shares follow a multinomial logit structure:

$$\frac{A_{ckt}}{\bar{A}_{ct}} = \frac{\exp\left(\alpha_k + \lambda W_{ckt}\right)}{\sum_{m \in \mathcal{K}} \exp\left(\alpha_m + \lambda W_{cmt}\right)},$$

where α_k captures crop-specific amenities and $\lambda > 0$ governs the responsiveness of acreage allocation to expected returns.

Land market clearing and spatial equilibrium. Land is allocated competitively within each county. The land rental rate r_{ct} adjusts such that total acreage demand equals available land:

$$\sum_{k \in \mathcal{K}} A_{ckt} = \bar{A}_{ct}.$$

A spatial equilibrium is a set $\{A_{ckt}, \sigma_{ckt}, r_{ct}\}$ such that insurance choices are optimal, acreage allocations satisfy crop choice probabilities, and land markets clear in every county.

An increase in insurance subsidies $s_{ckt}(d)$ raises insurance take-up by increasing the expected utility of insured states. Through the crop choice margin, higher insurance generosity increases the relative attractiveness of insured crops, shifting acreage allocation and bidding up local land rents. Consequently, insurance subsidies generate both production responses and capitalization effects in equilibrium.

15.8 Model B(v3): Acreage Substitution and the Nested Logit Framework

To extend Model A and account for potential crop-level reallocation in response to variation in insurance subsidies, I adopt a nested logit framework. In this structure, farmers make two sequential choices: first, they allocate total planted acreage across crops based on the relative attractiveness of insurance-inclusive bundles; second, they decide whether to insure each acre within the chosen crop. This model captures both the extensive margin (crop choice) and the intensive margin (insurance enrollment), offering a unified framework for analyzing policy incentives.

Are random coefficient model and LATE substitutes? Which one should I use? In random coefficients model, I assume stationarity assumption for the coeffcients. In random coefficient models, the partial effects of the endogenous explanatory variables can vary across individual units in unobserved ways. Estimating such models has fallen somewhat out of favor in empirical research yet they (implicitly) play a role in the recent program evaluation literature, where treatment effects are assumed to be heterogeneous. In the last decade or so, the focus in the treatment effects literature has been on quantities that are identified using standard IV methods under general assumptions, with the "local average treatment effect" (LATE) introduced in Imbens and Angrist (1994) being the most popular (Wooldridge 2015, p. 421).

While Model A treats planted acreage as fixed and focuses exclusively on the insurance decision, such a restriction may understate the full behavioral response to differential subsidies. Insurance generosity varies modestly across crops and locations, and if farmers perceive higher net returns to one crop due to more favorable insurance terms, they may shift acreage accordingly. Model B accommodates this possibility.

Lower Nest: Insurance Enrollment Conditional on Crop Choice For each crop $k \in \{\text{corn}, \text{soybeans}\}$, I retain the acre-level logit from Model A to model insurance demand conditional on crop planted:

$$S^{\text{ins}}_{ctk}(s) = \frac{\exp\left(\nu^{\text{ins}}_{ctk}(s)\right)}{1 + \exp\left(\nu^{\text{ins}}_{ctk}(s)\right)},$$

where

$$\nu^{\text{ins}}_{ctk}(s) = \alpha_{ck} + \tau_{tk} - \theta_k \ln(1-s) - \phi_k \ln P_{ctk} + X'_{ctk}\beta_k.$$

This stage remains unchanged from Model A. It captures how insured share within crop responds to price, premium, and controls. The inclusive value for insurance in money-metric utility units is defined as:

$$IV^{\text{ins}}_{ctk} = \ln\left(1 + \exp\left\{\nu^{\text{ins}}_{ctk}(s)\right\}\right).$$

Upper Nest: Crop Choice Based on Insurance-Inclusive Utility Let total planted acreage $A_{ct} = \sum_k A_{ctk}$ be fixed within a county-year. I model the share of land allocated to crop k using a logit response to inclusive utility:

$$\frac{A_{ctk}}{A_{ct}} = \frac{\exp\left(\gamma IV^{\text{ins}}_{ctk} + \alpha_c + \delta_t + Z'_{ctk}\eta\right)}{\sum_j \exp\left(\gamma IV^{\text{ins}}_{ctj} + \alpha_c + \delta_t + Z'_{ctj}\eta\right)},$$

where γ captures the degree to which insurance attractiveness affects crop allocation. When $\gamma = 0$, crop shares are unaffected by insurance and Model B collapses back to Model A. The vector Z_{ctk}

includes agronomic or economic shifters that influence crop profitability independently of insurance.

Welfare Insured acres are given by:

$$Q_{ctk}(s) = A_{ctk} \cdot S^{\text{ins}}_{ctk}(s).$$

Welfare from insurance provision is then:

$$W_{ctk}(s) = A_{ctk} \cdot \left[CS^{\text{ins}}_{ctk}(s) - CS^{\text{ins}}_{ctk}(s_0)\right] - (1+\lambda)sP_{ctk}Q_{ctk}(s) - \kappa_A Q_{ctk}(s),$$

where

$$CS^{\text{ins}}_{ctk}(s) = \ln\left(1 + \exp\left\{\nu^{\text{ins}}_{ctk}(s)\right\}\right).$$

Total welfare aggregates across counties, crops, and years. I numerically solve for the optimal subsidy s^* that maximizes this total welfare, accounting for both crop allocation and insurance decisions.

Model B offers a flexible extension of the sufficient-statistics approach in Model A. It allows me to evaluate not only whether insurance subsidies increase take-up, but also whether they distort land use by inducing acreage shifts across crops. A test of $H_0 : \gamma = 0$ provides direct evidence on whether insurance-inclusive utility affects planting decisions. If the estimated γ is small and counterfactual acreage shifts are minimal, Model A may suffice. Otherwise, Model B provides a more comprehensive basis for policy evaluation.

15.9 Model B(v2-adapted to Model A)- Acreage Substitution and the Nested Logit Model

To allow for crop-level reallocation in response to insurance subsidy variation, I extend the crop-specific demand model in Model A into a nested logit structure. In this model, farmers jointly choose which crop to plant and whether to insure it, recognizing that crop choice may respond to relative insurance attractiveness across crops. This formulation permits insurance-induced changes in crop acreage shares, capturing intensive and extensive margin responses within a unified framework.

Model A treats planted acreage as exogenous with respect to insurance subsidies. While this assumption is standard in the crop insurance literature, insurance subsidies vary slightly across crops and counties, raising the possibility that relative insurance attractiveness may affect crop allocation. To allow for insurance-induced acreage substitution between corn and soybeans, I extend Model A to a nested logit framework in which farmers jointly determine crop allocation and insurance enrollment.

Lower Nest: Insurance Enrollment Conditional on Crop Choice. For each crop $k \in \{\text{corn}, \text{soybeans}\}$, county c, and year t, I model the conditional insurance enrollment share among planted acres using the same acre-level logit specification as in Model A:

$$S^{\text{ins}}_{ctk}(s) = \frac{\exp(\nu^{\text{ins}}_{ctk}(s))}{1 + \exp(\nu^{\text{ins}}_{ctk}(s))}, \tag{39}$$

where

$$\nu^{\text{ins}}_{ctk}(s) = \alpha_{ck} + \tau_{tk} - \theta_k \ln(1-s) - \phi_k \ln P_{ctk} + X'_{ctk}\beta_k. \tag{40}$$

The parameters retain the same interpretation as in Model A. As before, identification of θ_k relies on exogenous variation in the net price $\ln(1-s)$, instrumented using policy-driven variation in subsidy schedules.

From this index, I compute the insurance inclusive value,

$$IV^{\text{ins}}_{ctk} = \ln\big(1 + \exp\{\nu^{\text{ins}}_{ctk}(s)\}\big), \tag{41}$$

which represents the money-metric expected utility of insurance conditional on crop choice.

Upper Nest: Crop Acreage Allocation. Let total planted acreage in a county-year be fixed. Farmers allocate land across crops based on the relative attractiveness of each crop's insurance-inclusive value. I model crop acreage shares using a fractional logit:

$$s^A_{ctk} = \frac{A_{ctk}}{\sum_j A_{ctj}} = g\big(\gamma\, IV^{\text{ins}}_{ctk} + \alpha_c + \delta_t + Z'_{ctk}\eta\big), \tag{42}$$

where $g(\cdot)$ is the logit link, α_c and δ_t are county and year fixed effects, and Z_{ctk} includes crop-specific agronomic and economic controls. The coefficient γ captures the extent to which insurance attractiveness affects crop allocation. When $\gamma = 0$, acreage allocation is independent of insurance incentives and Model B collapses to Model A.

Welfare. Let insured acres be $Q_{ctk}(s) = S^{\text{ins}}_{ctk}(s) \cdot A_{ctk}$. County–crop–year welfare is

$$W_{ctk}(s) = A_{ctk}\Big[CS^{\text{ins}}_{ctk}(s) - CS^{\text{ins}}_{ctk}(s_0)\Big] - (1+\lambda)\, s\, P_{ctk}\, Q_{ctk}(s) - \kappa_A\, Q_{ctk}(s), \tag{43}$$

where

$$CS_{ctk}^{\text{ins}}(s) = \ln\bigl(1 + \exp\{\nu_{ctk}^{\text{ins}}(s)\}\bigr).$$

Aggregate welfare is obtained by summing over counties, years, and crops. I compute optimal subsidy rates by maximizing total welfare with respect to s, allowing acreage to adjust endogenously through the upper nest. Model B generalizes Model A by allowing insurance subsidies to affect both the intensive margin of insurance coverage within crops and the extensive margin of crop allocation. The nested structure provides a transparent test of whether insurance subsidies distort land use: a Wald test of $H_0 : \gamma = 0$ evaluates whether insurance incentives induce meaningful acreage substitution between corn and soybeans.

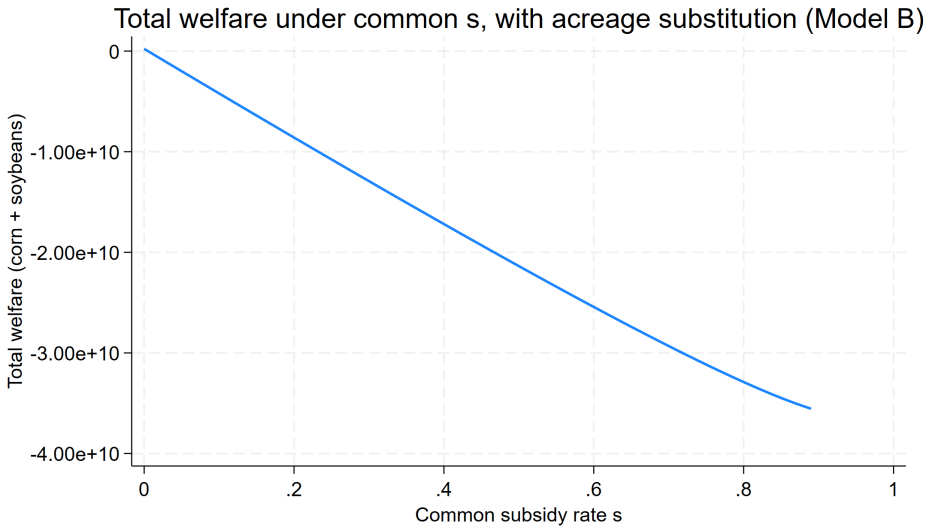
15.10 Results of Model B

Figure 45 shows the welfare simulations based on Model B—the specification that allows insurance subsidies to affect acreage allocation across crops—yield results that are remarkably similar to those from the baseline crop-specific model. Panel (a) of Figure **??** shows that total welfare declines monotonically as the common subsidy rate s increases, with no interior optimum; welfare losses become increasingly large for higher subsidy levels because government outlays scale directly with insured acreage while consumer-surplus gains remain small. Panel (b) illustrates the crop-specific welfare curves for corn and soybeans. Both crops exhibit strongly negative welfare slopes, although the magnitude is larger for corn due to its larger acreage base and higher baseline premium volume. Importantly, the welfare curves are nearly linear in s, and the difference between crops is driven almost entirely by the scale of the insured area rather than by any substitution response across crops. This pattern is consistent with the estimation results from the upper nest, which indicated that the insurance inclusive value IV_{ctk}^{ins} has no significant effect on acreage shares. As a consequence, the welfare implications under Model B closely mirror those from the simpler Model A: insurance subsidies generate substantial fiscal costs with limited offsetting consumer-surplus benefits, and allowing for acreage substitution does not materially change this conclusion.

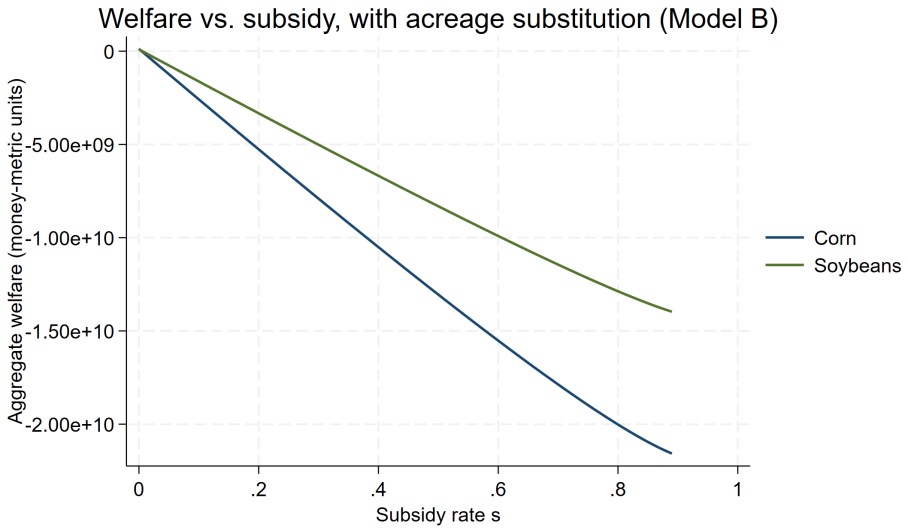
15.11 Test for Whether Nested Logit (Model B) is Needed After Estimating Model A

In order to assess whether crop insurance subsidies induce acreage substitution between corn and soybeans, we estimated a nested fractional–logit model in which the lower nest governs insurance take-up within crop and the upper nest allocates acreage shares across crops as a function of the inclusive value of insurance, IV_{ctk}^{ins}. In the upper nest we test the null hypothesis $H_0 : \gamma_{IV} = 0$, where γ_{IV} is the coefficient on IV_{ctk}^{ins}. The Wald test reported in Table 28 fails to reject this restriction ($\chi^2(1) = 0.99$, $p = 0.32$), implying that acreage shares do not respond systematically to the insurance inclusive value. Consistently, the information–criterion comparison in Table 29 shows that the unrestricted nested model (Model B) delivers virtually identical log-likelihood, AIC, and BIC values as the restricted specification (Model A) that excludes IV_{ctk}^{ins}. Taken together, these results provide no empirical evidence of subsidy-induced acreage substitution between corn and soybeans; in the rest of the analysis we therefore use the simpler crop-specific model without acreage substitution as our benchmark, and treat the nested specification as a robustness check.

This code implements a diagnostic test that evaluates whether the *upper nest* in a nested logit model is empirically needed. The logic is as follows:



(a) Two crops combined



(b) Crop-specific

Figure 45: The change with welfare measured with consumer surplus with change in crop insurance subsidy. The county-level data is from U.S. Department of Agriculture, Risk Management Agency (RMA) (2025).

Table 28: Nested test of acreage response to insurance inclusive value

Regressor	$\chi^2(1)$	p–value
IV^{ins}_{ctk}	0.99	0.3205

Table 29: Information–criterion comparison of acreage models

Model	N	ℓ(null)	ℓ(model)	df	AIC	BIC
NEST_B	5	−20,760.47	−20,550.02	676	42,452.04	42,188.02
RESTRICTED_A	5	−20,760.47	−20,550.69	676	42,453.39	42,189.37

- Model A assumes that crop insurance demand for corn and soy is independent across crops (no correlated unobservables across corn and soy).
- Model B adds an "upper nest" that allows correlation in unobserved utility through the inclusive value (IV) parameters.
- If the inclusive value terms for corn and soy are jointly insignificant, then Model A is sufficient; if they matter, Model B is required.

The Stata test (shown below) proceeds as follows:

1. Use a dataset where crop shares sum to one across {corn, soybeans}:

$$\texttt{share_corn} + \texttt{share_soy} = 1.$$

2. Estimate a fractional multinomial logit model with inclusive value terms:

$$\texttt{share_corn} = f(\texttt{c.IV_corn},\ \texttt{c.IV_soybeans},\ \text{county FE}, \text{year FE}),$$
$$\texttt{share_soy} = f(\texttt{c.IV_corn},\ \texttt{c.IV_soybeans},\ \text{county FE}, \text{year FE}).$$

3. Test whether the inclusive values matter jointly:

$$H_0 : \alpha_{\text{IV,corn}} = \alpha_{\text{IV,soybeans}} = 0.$$

If the p, value of the joint test is above 0.10 and the coefficients are economically tiny, then the inclusive values do not explain crop mix changes, and the simpler Model A is likely sufficient.
The Stata code:

```
* This code tests whether nested logit is needed after Model A.
* If the inclusive value parameters (upper nest) do not matter,
* then Model A is sufficient and Model B adds little.

* Crop shares summed to 1 across {corn, soy}
use crop_shares_with_IV.dta, clear
* Variables: county_id year share_corn share_soy IV_corn IV_soy

* Fractional multinomial logit with inclusive values
ssc install fmlogit, replace
fmlogit share_corn share_soy c.IV_corn c.IV_soy ///
    i.county_id i.year, basevar(share_soy) vce(cluster county_id)

* Test whether inclusive-value parameters matter jointly
test IV_corn IV_soy

* Interpretation:
* If p > 0.10 and effects are tiny, Model A is enough.
* If coefficients are large or significant, Model B (nested logit) is required.
```

```
* Back-of-envelope counterfactual:
* Raise s by +0.05 in the lower nest, recompute IVs,
* Plug into the model, predict new crop shares,
* Report share changes $\Delta$share_corn, $\Delta$share_soy.
```

Parameters not necessary for Model A (but needed for a structural model) The following parameters are unnecessary for Model A (which uses a revealed-preference, inclusive-value welfare formula), but are used in structural risk-utility models:

- Risk aversion (CARA/CRRA parameters), γ,
- Variance terms (c_v), profit gap Δ,
- Correlation parameters (ρ_1, ρ_2),
- Adverse-selection terms (inverse Mills ratio λ_L, marginal indemnity τ_b),
- Moral-hazard elasticities ($\eta_{R,b}$, $\eta_{\sigma,b}$),
- Adverse-selection elasticities ($\eta_{I,\tau}$),
- Subsidy-response elasticities ($\eta_{\tau,s}$).

These are required only if you build a full CARA-Normal or sufficient-statistics structural welfare model. They are not needed for the baseline Model A.

15.12 Discussion of Sufficient Statistics Modeling

Earlier work over roughly the same time period (1989-2012) using county-level data had shows that while increases in subsidies can induce farmers to enroll more land, they primarily encourage them to adopt higher levels of coverage on land already enrolled (E. O'Donoghue 2014).

According to E. O'Donoghue (2014), based on total premiums (the sum of the premium the producer pays for a policy plus associated premium subsidies), the average response for corn producers in the Midwest (IL, IN, IA, and OH) was to increase the demand for crop insurance when the price of crop insurance fell. That is, a 1-percent increase in subsidies per acre aof roughly 0.96 percent in total premium purchases. The average response of corn producers in the Lake States (MI, MN, and WI) and the Northern Plains (KS, NE, and SD) exhibited smaller changes, with total premium purchases increasing by 0.83 percent and 0.80 percent, respectively, following a 1-percent increase in subsidies.

According to E. O'Donoghue (2014), the average responses of soybean producers were similar but smaller than those of corn producers (not surprising, as corn producers often grow soybeans as well). For a 1-percent increase in subsidies, the average response of soybean producers was to increase total premium purchases by 0.84 percent in the Midwest, 0.59 percent in the Lake States, and 0.47 percent in the Northern Plains.

According to E. O'Donoghue (2014), the effects of subsidy increases are also revealed in other measures of demand for crop insurance during the period. Premiums per acre (total premiums divided by acres insured) and liabilities per acre (total liability divided by acres insured) also tended to rise, with premiums increasing the most. The change in demand measured as acres enrolled (both total and buy-up policies—those policies above the minimal, fully subsidized catastrophic coverage available to growers) typically showed very small, and statistically insignificant, movements in response to

subsidy changes. Further, these results were often negative, suggesting the potential for cross-product substitution effects. Together, these findings suggest that subsidies may not (during this time frame) have drawn many new acres into crop insurance, but they appear to have induced higher levels of coverage on enrolled acres.

According to E. O'Donoghue (2014), across all regions, for a 1-percent increase in subsidies, demand for crop insurance rose about 0.86 percent for corn, 0.74 percent for soybeans, and 0.64 percent for wheat in terms of total premiums. For liabilities, increases in demand ranged from roughly 0.2 percent for soybeans and wheat to 0.3 percent for corn. For acres enrolled, the rise in subsidies typically resulted in statistically insignificant changes in demand with magnitudes below 0.2 percent.

I used county-level data in this research. This means I could not control for the idiosyncratic behavior of individuals. So care must be taken in interpreting results, as what holds at the aggregate (county) level may mask considerable variation at the individual level. Furthermore, the analysis was undertaken one crop at a time and only with corn and soybeans, assuming that each enterprise on a farm is operated independently of the next. If cross-product substitution effects exist that are not taken into account here, results may change.

16 Predictors of Cause of Loss

This section models the relationship between remote sensing data and *cause of loss* for crop insurance. The analysis combined remote sensing and crop insurance loss data to model the relationship between observed field conditions and the official *Cause of Loss* categories reported by the USDA Risk Management Agency (RMA).[18]

To examine how economic conditions of farmers affects their decision to report losses to the RMA, I followed an approach that is widely used in economics literature to use the change in local rainfall as a proxy for local economic conditions (see Corno, Hildebrandt, and Voena (2020), Jayachandran (2006), Burke, Gong, and Jones (2015), Kaur (2019), Shah and Steinberg (2017) among many others). To construct a measure of local droughts, we use rainfall data produced by geographers at the University of Delaware ("UDel data"). The UDel data set provides estimates of monthly precipitation on a 0.5 × 0.5 degree grid covering terrestrial areas across the globe for the 1900–2017 period.[19]

High-resolution aerial imagery from the National Agriculture Imagery Program (NAIP) was used to characterize post-event field conditions such as flood ponding, lodging, or bare-soil exposure. Each NAIP image, typically captured once per year during the growing season, was processed to generate vegetation and water indices (e.g., NDVI, NDWI) and texture metrics that describe within-field variability and potential damage patterns. These high-resolution indicators were spatially joined to county-level RMA data on insured crop losses and corresponding CoL codes (e.g., drought, excess moisture, hail). This integration enabled the construction of a labeled dataset linking remotely sensed field signals to observed insurance loss outcomes.

To capture the timing and climatic drivers of each peril, the NAIP indicators were complemented with time-dense satellite and weather data—including MODIS and Sentinel-2 vegetation time series, Sentinel-1 SAR flood maps, PRISM/GRIDMET temperature and precipitation anomalies, drought

[18] I thank Gianluca Sperone for his help with working with remote sensing data sets, and Julia Koschinsky for facilitating my attendance in remote sensing classes at University of Chicago.

[19] https://psl.noaa.gov/data/gridded/data.UDel_AirT_Precip.html

indices (SPEI/PDSI), and NOAA hail and storm event data. Data preprocessing and feature extraction were conducted in Google Earth Engine, QGIS, and Python (geopandas, rasterio, xarray), while econometric and machine-learning models—such as multinomial logit, random forest, or gradient boosting—were implemented in Stata and Python (scikit-learn). Model outputs estimated the probability of each loss cause given the remotely sensed and meteorological conditions, offering a scalable method to attribute or validate reported crop insurance losses across U.S. counties. Use of texture, soil moisture and vegetation biomass for the same reason. USDA FSA data set for crop yield and other agricultural related data.

The modeling process begins by framing each cause of loss (CoL) as a discrete outcome that depends on observable environmental and agronomic factors. A multinomial logit model (MNL) is first used to estimate the probability that a given county–crop–year observation experiences a specific loss cause—such as drought, excess moisture, freeze, or hail—conditional on predictors derived from weather records and remote sensing data. In this econometric specification, the dependent variable is categorical (the dominant or most probable CoL), while the explanatory variables include precipitation and temperature anomalies, drought indices, soil moisture, vegetation health metrics (NDVI, NDWI), and fixed effects that control for crop and regional heterogeneity. The multinomial logit provides interpretable coefficients and marginal effects, allowing researchers to quantify how a one-unit change in a predictor (e.g., a one-standard-deviation increase in drought severity) shifts the probability distribution across loss categories. This interpretability makes it particularly suitable for hypothesis testing and policy discussion within an economics framework.

To enhance predictive accuracy and capture nonlinear relationships that the multinomial logit cannot model, the same feature set can be used in a machine-learning classifier, such as random forest, gradient boosting (e.g., XGBoost), or neural networks. These algorithms automatically learn complex interactions between weather and remote-sensing variables without prespecifying functional forms. A hybrid or "mixed" modeling strategy combines both approaches: the MNL serves as a benchmark for structural interpretation, while the machine-learning model refines prediction quality by capturing nonlinearities and variable interactions. In practice, the two models can be integrated by using the machine-learning model's predicted probabilities as additional covariates in the logit, or by comparing their out-of-sample predictive performance through cross-validation. This combination yields a framework that is both explainable—through the economic logic of the multinomial logit—and powerful in prediction—through the flexibility of machine learning—making it highly suitable for modern applied economic analysis of climate risk and insurance behavior.

For the U.S.—and especially the Midwest—previous studies have largely connected remote sensing to specific types of agricultural perils rather than developing unified multi-cause models of crop loss. For instance, the USDA's RF-CLASS system demonstrated the feasibility of satellite-based crop-loss mapping for decision support and has been used in flood monitoring applications across U.S. states. Similarly, several studies analyzed the lingering effects of the 2019 Midwest spring floods using Sentinel-1 SAR and multi-sensor time-series data, revealing substantial yield impacts and identifying spatial patterns of inundation and delayed planting. Research on prevented planting has also leveraged satellite-derived flood layers and cropland history to infer likely prevented-plant acreage beyond county aggregates, offering insights into insurance-relevant loss mechanisms.

Other peril-specific work focuses on hail and drought. Early hail-damage adjustment studies in Illinois employed aerial imagery to improve insurance assessments, and recent research—though mostly outside the U.S.—has used machine learning and high-resolution remote sensing to detect and quantify hail-induced crop damage. On the insurance data side, the USDA Risk Management Agency's Cause-of-Loss (CoL) dataset has been used to describe spatial and temporal patterns of indemnities, showing that drought, excess moisture, and hail dominate weather-related claims across the country, while smaller-scale studies (e.g., for Tennessee) modeled indemnities for drought versus excess moisture. Together, this literature provides a strong foundation for peril-specific analysis in the Midwest but highlights the absence of a single, county–crop–year framework integrating remote sensing and weather data to predict multiple loss causes simultaneously—making such a study both methodologically and regionally novel.

17 Insurance Product Substitution

Alternative Framing: Insurance Product Substitution An alternative lens for analyzing the optimal design of crop insurance subsidies is to focus on product-level substitution between different types of coverage—namely, yield-based insurance (YP) and revenue-based insurance (RP, RP-HPE, and ARP). Rather than modeling the intensity of insurance coverage for a given crop, this approach treats farmers as choosing among mutually exclusive insurance products, each offering distinct risk protection and subsidization profiles. Under this framing, subsidies act as relative price instruments that distort product choice, and optimal policy aims to balance efficiency and equity across coverage types.

Let $i \in \{0, \text{YP}, \text{RP}\}$ denote the insurance choice for each farmer or county-crop-year observation, where $i = 0$ denotes no insurance. Each alternative i yields an indirect utility:

$$U_{ctki} = V_{ctki} + \varepsilon_{ctki},$$

where V_{ctki} captures the observable component (including price, subsidy, crop yield volatility, etc.) and ε_{ctki} follows a Type I Extreme Value distribution. Assuming a multinomial logit structure, the probability of selecting insurance type i is:

$$P_{ctki} = \frac{\exp(V_{ctki})}{\sum_{j \in \{0,\text{YP},\text{RP}\}} \exp(V_{ctkj})}.$$

The deterministic component V_{ctki} may include crop-year fixed effects, premium cost net of subsidy $\log[(1-s_i)P_{ctki}]$, and product-specific controls (e.g., historical indemnity rates, farm revenue volatility). The price sensitivity coefficients can be used to compute the welfare effects of marginal subsidy changes, using sufficient statistics methods (Chetty and Finkelstein 2013).

This approach is well suited for analyzing the behavioral and welfare impacts of reforms that reallocate subsidies between YP and RP products—such as those considered in Jialing Yu (2017) and Glauber (2004). In contrast to Model A and B, which treat insurance intensity as the margin of adjustment, this product-level framing captures extensive-margin shifts in insurance contract choice. Estimating optimal subsidies in this framework involves simulating counterfactual market shares and expected indemnities under alternative subsidy schedules, and comparing total welfare across policy regimes.

The empirical implementation of this model requires detailed product-level insurance data, includ-

ing market shares, premiums, and indemnities by type. Such an analysis is complementary to the crop-specific models developed in Section 15, and may be especially relevant for evaluating policies that shift the mix of insurance products without necessarily expanding total coverage.

18 Economic Model

Environment and choice. I consider a farmer indexed by county c, crop i, and year t, who chooses insurance coverage intensity D_{ict}, which may be interpreted as the fraction of planted area insured, the coverage level, or any monotone index of insurance coverage.

Let random profit without insurance be

$$\pi^0_{ict} = p_{ict} y_{ict} - k_{ict},$$

where output price is p_{ict}, production cost is k_{ict}, and yield y_{ict} depends on weather shocks ω_{ict}, such as degree days and precipitation realizations.

With insurance coverage D_{ict}, profit becomes

$$\pi_{ict}(D) = \pi^0_{ict} + I_{ict}(D, \omega_{ict}) - P_{ict}(D),$$

where $I_{ict}(\cdot)$ denotes state-contingent indemnity payments and $P_{ict}(D)$ is the insurance premium.

Participation frictions and instruments. In addition to the premium, I assume that farmers incur non-price participation frictions associated with insurance use. These frictions take the form

$$\kappa_{ict}(D; Z_c) = D \cdot \tau(Z_c),$$

where Z_c is a vector of predetermined county-level shifters. Specifically,

$$Z_{1c} = \text{Distance}_c \quad \text{and} \quad Z_{2c} = \text{Stations}_c,$$

corresponding to distance to the nearest USDA service office and the number of weather stations in the county in 1989, respectively.

Distance to USDA offices captures administrative, informational, and transaction costs associated with enrollment and servicing, while historical weather-station density proxies for local information infrastructure and administrative capacity. The sign of $\tau(Z_c)$ is ultimately empirical.

Net profit can therefore be written as

$$\tilde{\pi}_{ict}(D) = \pi^0_{ict} + I_{ict}(D, \omega_{ict}) - P_{ict}(D) - D\tau(Z_c).$$

Expected-utility maximization. Farmers choose insurance coverage to maximize expected utility,

$$\max_{D \in [0,1]} \; \mathbb{E}[U(\tilde{\pi}_{ict}(D)) \mid \mathcal{I}_{ict}],$$

where $\mathcal{I}_{ict}$ denotes the information set available at the insurance sign-up date, including predetermined climate risk, historical weather variability, and prices.

First-order condition. For an interior solution, the first-order condition is

$$\mathbb{E}\left[U'(\tilde{\pi}_{ict}(D))\left(\frac{\partial I_{ict}}{\partial D}-\frac{\partial P_{ict}}{\partial D}-\tau(Z_c)\right)\right]=0.$$

Define the risk-adjusted marginal benefit and marginal cost of insurance as

$$MB_{ict}(D)\equiv\mathbb{E}\left[U'(\tilde{\pi}_{ict}(D))\,\frac{\partial I_{ict}}{\partial D}\right],\qquad MC_{ict}(D)\equiv\mathbb{E}\left[U'(\tilde{\pi}_{ict}(D))\,\frac{\partial P_{ict}}{\partial D}\right],$$

and define the marginal participation friction as

$$FC_{ict}(D;Z_c)\equiv\mathbb{E}[U'(\tilde{\pi}_{ict}(D))]\,\tau(Z_c).$$

The optimal insurance condition can then be written compactly as

$$MB_{ict}(D)-MC_{ict}(D)-FC_{ict}(D;Z_c)=0.$$

This expression highlights that optimal coverage equates the marginal, risk-adjusted benefits of insurance to the sum of marginal premium costs and marginal participation frictions.

Linear approximation and empirical first stage. To connect the theoretical model to the empirical specification, I take a first-order approximation of the optimality condition around a typical coverage level D^*. I assume that (i) marginal benefits and costs depend approximately linearly on observable risk and price shifters, and (ii) participation frictions enter linearly through $\tau(Z_c)$.

This yields the reduced-form approximation

$$D_{ict}=\pi_1 Z_{1c}+\pi_2 Z_{2c}+\gamma' W_{ict}+\alpha_s+\delta_t+u_{ict},$$

where W_{ict} includes growing-degree-day measures, precipitation moments, and output prices, while α_s and δ_t denote state and year fixed effects.

Economic interpretation. The coefficient π_1 captures the semi-elastic response of insurance coverage to administrative and transaction costs associated with distance to USDA offices. The coefficient π_2 reflects the effect of local information and administrative capacity, as proxied by historical weather-station density. The vector γ captures how baseline climate risk and price incentives shape insurance demand. This formulation provides the economic foundation for the first-stage regressions used in the empirical analysis.

19 Discussion

The results of both the sublia and fracInsured specifications highlight a consistent trade-off in the design of the Federal Crop Insurance Program (FCIP). Subsidies and insurance coverage reduce yield losses under moderate stress of heat and normal rainfall, but increase vulnerability under extreme conditions, particularly heat. This asymmetric effect reflects the weather trends in our study period: as shown in the stylized facts, the Midwest has experienced a rising number of moderate days that are relatively more conducive to soybean production, but also more frequent extreme heat events that disproportionately affect corn yields. In this context, our findings suggest that insurance subsidies

amplify short-term protection under moderate conditions, while leaving farmers more exposed to yield losses when weather crosses damaging thresholds.

The crop-specific results reinforce this pattern. Corn, which has higher yields but also greater yield variability (Figure 6), benefits more in the short run from insurance support under favorable weather but is especially vulnerable to extreme heat, consistent with its physiological sensitivity. Soybeans, by contrast, face relatively fewer extreme hot days and have lower variance, meaning that insurance dampens incentives for adaptation but in a more muted way. These dynamics mirror farmers' planting choices: despite relative price signals that should have favored soybeans (Figure 4), acreage decisions did not fully adjust, likely because insurance reduced the need to reallocate risk through crop switching.

Taking together, these results contribute to a growing literature (Annan and Schlenker (2015); Jisang Yu, A. Smith, and Sumner (2018); Chemeris, Liu, and Alan P. Ker (2022)) showing that while crop insurance stabilizes income and smooths consumption in the short run, it may undermine long-run resilience to climate change. The policy implication is that subsidy design must better account for heterogeneous responses across the climate distribution and across crops: without adjustment, crop insurance may inadvertently amplify the weather risks that it was designed to mitigate.

20 Limitation

In this paper, we did not consider the spatial effect of subsidies on acreage. As Barry K. Goodwin and Vincent H. Smith (2013) has shown, the geographic distributions of participation and relative returns to insurance are varied across the geography in the US. To the extent that the benefits of subsidized crop insurance are heterogeneous across regions and crops, distortions in production patterns may follow. Participation (in terms of total liability) is particularly high in the Corn Belt, the Mississippi Delta, and in the Upper Great Plains. At the same time, the relative returns to insurance as represented by the ratio of indemnities to subsidy-adjusted premiums, are lower in the Corn Belt but appear to be particularly high in Western Texas, the mid-south, and in certain portions of the Northern Plains. This suggests the potential for production distortions as certain crops and growing areas receive significantly higher subsidies. Next work could address this issue.

On a separate note, this paper took a "positive" approach to study the impact of subsidies on crop acreage, taking the existing policies as given. An alternative "normative" approach would be needed to study whether the presence of an insurance subsidy is an optimal welfare-improving policy, and if yes, what would the optimal level of subsidy be. Section 3 in the Handbook chapter of Chetty and Finkelstein (2013) could be a good starting point for answering this question.

In calculating the optimal level of insurance subsidy, I did not model the choice of farmers over percentage coverage of insurance, i.e. yield or revenue coverage that can take values of 55%, 60%, ..., 90%. A follow up research can take that dimension into account. Adjesy the three limitations below for this paper. These limitations are from the 2006 paper of Raj Chetty, "A general formula for optimal level of social insurance".
Fiscal externalities. I assumed that the only distortionary tax/subsidy in the economy is the UI benefit. In practice, many behaviors are taxed, and these taxes create bfiscal externalitiesQ that could

change the formula for the optimal benefit rate. For example, higher UI benefits will in general lower private savings, which could in turn reduce tax collections from capital gains or dividends. This tax revenue effect is not internalized by the agent and therefore affects the optimal benefit rate directly. It would be useful to determine the magnitude of such fiscal externalities too assess whether they affect the calculation of the optimal benefit rate significantly.

General equilibrium effects: In the models analyzed here, all behavioral responses to UI were solely determined by the agent. This assumption was important because the envelope conditions used to obtain the formula for optimal benefits relied on the idea that all endogenous variables in the model are chosen to maximize the agent's utility. Obtaining a reduced-form formula that takes equilibrium responses by firms into account would be useful.

Myopic agents. The envelope arguments above rely on the assumption that agents are optimizing. If agents experience large consumption drops during unemployment because they are myopic and do not save enough, the formula for the optimal benefit level may be different.

21 Conclusion

I have shown that aggregated annual rainfall and exposure to heat measures mask intra-seasonal heterogeneity in rainfall/heat exposure effects and attenuate estimated impacts of rainfall shocks on crop yield First line of conclusion of Farris et al. (2023).

This study investigated how federal crop insurance subsidies influence the sensitivity of US crop yields to weather extremes. Across specifications, two consistent patterns emerge. First, insurance subsidies reduce yields on average, a result robust in the baseline models where subsidies per dollar of liability (sublia) and the fraction of acreage insured (fracInsured) are treated as endogenous. Focusing on the number of heating degree days as the most important dimension of extreme weather, we observed that the effect of sublia on yield is negative across the distribution of hot degree days. The impact was small and mostly insignificant at lower quantiles, but it becomes larger and statistically significant at higher heat levels, reaching yield losses of about 7–9 percent per additional percentage point of adoption. At the 99th, 95th, 90th, 75th percentiles and average of hot degree days, a one percentage point increase in sublia decrease yield by 8.6%, 7.8%, 7.4%, 7% and 6.8%. This negative relationship supports the hypothesis that subsidies can discourage adaptive investment and shift cultivation toward more marginal land, thereby lowering resilience.

Crop-specific analyses reinforce these findings. For corn, the evidence is especially strong: subsidies magnify vulnerability to heat and rainfall extremes, with robust first-stage instruments supporting the reliability of these results. For soybeans, the effects are weaker and less systematic, with subsidy–weather interactions often statistically insignificant. This divergence highlights crop-specific agronomic differences in sensitivity to weather stress and the role of insurance. Taken together, the results indicate that while crop insurance reduces income volatility in the short run, it may inadvertently increase sensitivity of yield to weather extremes, particularly for corn, by altering both adaptive incentives and land-use decisions. From a policy perspective, these findings underscore the need to balance the stabilizing function of crop insurance with incentives for climate adaptation.

Designing subsidies that encourage resilience-enhancing practices such as adoption of drought-

tolerant varieties, conservation tillage, and improved water management, could mitigate the disincentive effects documented here and better align federal crop insurance with the realities of a changing climate.

22 Acknowledgement

The author reports no conflicts of interest in this research. She received funding from Indiana University Bloomington for presenting an earlier version of this paper at the 89th *Midwest Economics Association Conference* in 2025.

23 Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

24 References

References

Abrigo, Michael RM and Inessa Love (2016). "Estimation of panel vector autoregression in Stata". In: *The Stata Journal* 16.3, pp. 778–804.

Andrews, Donald WK and Biao Lu (2001). "Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models". In: *Journal of econometrics* 101.1, pp. 123–164.

Angel, Jonathan et al. (2018). "Midwest". In: *Impacts, Risks, and Adaptation in the United States: Fourth National Climate Assessment, Volume II*. Ed. by David R. Reidmiller et al. Washington, DC, USA: U.S. Global Change Research Program, pp. 872–940.

Angrist, Joshua D and Jörn-Steffen Pischke (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Annan, Francis and Wolfram Schlenker (2015). "Federal Crop Insurance and the Disincentive to Adapt to Extreme Heat". In: *American Economic Review* 105.5, pp. 262–266.

Anselin, Luc (2024a). *Applied Spatial Regression Analysis - Lecture: Robust Tests*. YouTube, viewed *minute 00:00*. URL: https://youtu.be/nWk2eBPEsR4?si=scT1to8JxGaElejQ.

— (2024b). *Applied Spatial Regression Analysis - Lecture: Robust Tests*. YouTube, viewed *minute 05:20*. URL: https://youtu.be/nWk2eBPEsR4?si=E8e-BJWACJlci0NG.

— (2024c). *Applied Spatial Regression Analysis - Lecture: Spatial Regression Specifications*. YouTube, viewed *minute 2:20*. URL: https://www.youtube.com/watch?v=yMwQ4MljEWc&list=PLzREt6r1NenlbfMWgHZ9aRpI

— (2024d). *Applied Spatial Regression Analysis - Lecture: Spatial Spillovers*. YouTube, viewed *minute 14:30*. URL: https://youtu.be/C-sBZ9XWC9g?si=iwqPTRCNpBQYddtg&t=876.

— (2024e). *Applied Spatial Regression Analysis - Lecture: Specification Tests*. YouTube, viewed *minute 35:46*. URL: https://youtu.be/XyIokAi01Ko?si=RCYHrXi_zI3ouBkS&t=2146.

Anselin, Luc and Anil K Bera (1998). "Spatial dependence in linear regression models with an introduction to spatial econometrics". In: *Statistics textbooks and monographs* 155, pp. 237–290.

Arellano, Manuel (1987). "Computing robust standard errors for within-groups estimators." In: *Oxford Bulletin of Economics & Statistics* 49.4.

Babcock, Bruce A and Chad E Hart (2005). "Influence of the premium subsidy on farmers' crop insurance coverage decisions". In.

Baily, Martin Neil (1978). "Some aspects of optimal unemployment insurance". In: *Journal of public Economics* 10.3, pp. 379–402.

Ballard, Charles L, John B Shoven, and John Whalley (1985). "General equilibrium computations of the marginal welfare costs of taxes in the United States". In: *The American Economic Review* 75.1, pp. 128–138.

Balli, Hatice Ozer and Bent E. Sørensen (2013). "Interaction Effects in Econometrics". In: *Empirical Economics* 45.1. Earlier draft (2010) available at https://www.uh.edu/~bsorense/Ozer-Balli_Sorensen.pdf, pp. 583–603. DOI: 10.1007/s00181-012-0604-2.

Baquie, Sandra et al. (2025). *Industrial Policies: Handle with Care*. International Monetary Fund.

Bekkerman, Anton, Eric J Belasco, and Vincent H Smith (2018). "Where the money goes: The distribution of crop insurance and other farm subsidy payments". In: URL: https://www.aei.org/wp-content/uploads/2018/01/Where-the-Money-Goes.pdf?x97961.

Belasco, Eric J and Kate Binzen Fuller (2022). "Who buys crop insurance? Predictors of the participation gap between organic and conventional farms". In: *Applied Economic Perspectives and Policy* 44.3, pp. 1554–1572.

Bellemare, Marc F and Casey J Wichman (2020). "Elasticities and the inverse hyperbolic sine transformation". In: *Oxford Bulletin of Economics and Statistics* 82.1, pp. 50–61.

Bera, Anil K, Osman Doğan, et al. (2020). "Specification tests for spatial panel data models". In: *Journal of Spatial Econometrics* 1.1, p. 3.

Bera, Anil K and Mann J Yoon (1993). "Specification testing with locally misspecified alternatives". In: *Econometric theory* 9.4, pp. 649–658.

Berry, Steven T and Philip A Haile (2021). "Foundations of demand estimation". In: *Handbook of industrial organization*. Vol. 4. 1. Elsevier, pp. 1–62.

Brady, Ryan R (2011). "Measuring the diffusion of housing prices across space and over time". In: *Journal of Applied Econometrics* 26.2, pp. 213–231.

Brazil Ministry of Agriculture, Livestock and Food Supply (Nov. 2016). *Agricultural Risk Management in Brazil*. Presentation at the World Trade Organization. Accessed via WTO website. Geneva. URL: https://www.wto.org/english/tratop_e/agric_e/brazil91116_e.pdf.

Browning, Edgar K (1987). "On the marginal welfare cost of taxation". In: *The American Economic Review*, pp. 11–23.

Bureau, U.S. Census (2000). *Census 2000 TIGER/Line Files*. https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html. Geographic boundary files in TIGER/Line ASCII format.

Byrnes, James P, David C Miller, and William D Schafer (1999). "Gender differences in risk taking: A meta-analysis." In: *Psychological bulletin* 125.3, p. 367.

Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller (2011). "Robust inference with multiway clustering". In: *Journal of Business & Economic Statistics* 29.2, pp. 238–249.

Chemeris, Anna, Yong Liu, and Alan P. Ker (2022). "Insurance subsidies, climate change, and innovation: Implications for crop yield resiliency". In: *Food Policy* 108.

Chetty, Raj (2006). "A general formula for the optimal level of social insurance". In: *Journal of Public Economics* 90.10-11, pp. 1879–1901.

Chetty, Raj and Amy Finkelstein (2013). "Social insurance: Connecting theory to data". In: *Handbook of public economics*. Vol. 5. Elsevier, pp. 111–193.

Chintapalli, Prashant and Christopher S Tang (2022). "Crop minimum support price versus cost subsidy: Farmer and consumer welfare". In: *Production and Operations Management* 31.4, pp. 1753–1769.

Chung, Wonho (2014). "Analysis of social welfare effects of crop revenue insurance". In: *Journal of Rural Development/Nongchon-Gyeongje* 37.4, pp. 75–95.

Claassen, Roger, Christian Langpap, and JunJie Wu (2017). "Impacts of federal crop insurance on land use and environmental quality". In: *American Journal of Agricultural Economics* 99.3, pp. 592–613.

Coate, Stephen (1995). "Altruism, the Samaritan's dilemma, and government transfer policy". In: *The American Economic Review*, pp. 46–57.

Coble, Keith H. and Barry J. Barnett (2013). "Why do we subsidize crop insurance?" In: *American Journal of Agricultural Economics* 95.2, pp. 498–504.

Corno, Lucia, Nicole Hildebrandt, and Alessandra Voena (2020). "Age of marriage, weather shocks, and the direction of marriage payments". In: *Econometrica* 88.3, pp. 879–915.

Cummins, J David and Olivier Mahul (2009). *Catastrophe risk financing in developing countries: principles for public intervention*. World Bank Publications.

Deschênes, Olivier and Michael Greenstone (2007). "The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather". In: *American economic review* 97.1, pp. 354–385.

Didan, K. (2021). *MODIS/Terra Vegetation Indices 16-Day L3 Global 500m SIN Grid V061*. Accessed: 2025-11-11. NASA Land Processes Distributed Active Archive Center (LP DAAC). DOI: 10.5067/MODIS/MOD13A1.061. URL: https://doi.org/10.5067/MODIS/MOD13A1.061.

Diekmann, Florian et al. (2012). "Judging farmers' willingness to trade distance and taxes for extension services". In: *Applied economic perspectives and policy* 34.3, pp. 454–471.

Dijkstra, Edsger W. (1959). "A note on two problems in connexion with graphs". In: *Numerische Mathematik* 1, pp. 269–271. DOI: 10.1007/BF01386390.

Edwards, Chris (2025). *Department of Agriculture Timeline*. Accessed: 2025-11-03. Cato Institute. URL: https://www.downsizinggovernment.org/agriculture/timeline.

Falco, Salvatore Di et al. (2014). "Crop insurance as a strategy for adapting to climate change". In: *Journal of Agricultural Economics* 65.2, pp. 485–504.

Farris, Jarrad et al. (2023). "Growing pains: Timing of in Utero rainfall shocks and child growth in rural Rwanda". In: *Economic Development and Cultural Change* 71.2, pp. 793–818.

Finkelstein, Amy and Kathleen McGarry (2006). "Multiple dimensions of private information: evidence from the long-term care insurance market". In: *American Economic Review* 96.4, pp. 938–958.

Florax, Raymond and Henk Folmer (1992). "Specification and estimation of spatial linear regression models: Monte Carlo evaluation of pre-test estimators". In: *Regional science and urban economics* 22.3, pp. 405–432.

Gardner, Bruce L and Randall A Kramer (1986). *Experience with crop insurance programs in the United States*.

Glauber, Joseph W (2004). "Crop insurance reconsidered". In: *American Journal of Agricultural Economics* 86.5, pp. 1179–1195.

— (2013). "The growth of the federal crop insurance program, 1990—2011". In: *American Journal of Agricultural Economics* 95.2, pp. 482–488.

Gonne, Nicolas and Olivier Hubert (2020). "Spatial distancing: air traffic, COVID-19 propagation, and the cost efficiency of air travel restrictions". In: *Covid Economics* 24, pp. 111–25.

Goodwin, Barry K (1993). "An empirical analysis of the demand for multiple peril crop insurance". In: *American journal of agricultural economics* 75.2, pp. 425–434.

Goodwin, Barry K and Ashley Hungerford (2015). "Copula-based models of systemic risk in US agriculture: Implications for crop insurance and reinsurance contracts". In: *American Journal of Agricultural Economics* 97.3, pp. 879–896.

Goodwin, Barry K and Alan P Ker (1998). "Nonparametric estimation of crop yield distributions: implications for rating group-risk crop insurance contracts". In: *American journal of agricultural economics* 80.1, pp. 139–153.

Goodwin, Barry K. and Vincent H. Smith (2013). "What Harm Is Done by Subsidizing Crop Insurance?" In: *American Journal of Agricultural Economics* 95.2, pp. 489–497.

Granger, Clive W. J. and Paul Newbold (July 1974). "Spurious Regressions in Econometrics". In: *Journal of Econometrics* 2.2, pp. 111–120.

Gustin, Georgina (2018). *U.S. Taxpayers on the Hook for Insuring Farmers Against Growing Climate Risks*. Inside Climate News. URL: https://insideclimatenews.org/news/31122018/crop-insurance-farm-bureau-taxpayer-subsidies-climate-change-risk-rising/.

Halek, Martin and Joseph G Eisenhauer (2001). "Demography of risk aversion". In: *Journal of Risk and Insurance*, pp. 1–24.

Halleck Vega, Solmaria and J Paul Elhorst (2015). "The SLX model". In: *Journal of Regional Science* 55.3, pp. 339–363.

Hansen, Lars Peter (1982). "Large sample properties of generalized method of moments estimators". In: *Econometrica: Journal of the econometric society*, pp. 1029–1054.

Hazell, Peter BR (1992). "The appropriate role of agricultural insurance in developing countries". In: *Journal of International Development* 4.6, pp. 567–581.

Hendren, Nathaniel, Camille Landais, and Johannes Spinnewijn (2021). "Choice in insurance markets: A Pigouvian approach to social insurance design". In: *Annual Review of Economics* 13.1, pp. 457–486.

Hendren, Nathaniel and Ben Sprung-Keyser (2020). "A unified welfare analysis of government policies". In: *The Quarterly journal of economics* 135.3, pp. 1209–1318.

Holmes, Thomas J (2011). "The diffusion of Wal-Mart and economies of density". In: *Econometrica* 79.1, pp. 253–302.

Jefferson, Thomas (Aug. 23, 1785). *Letter to John Jay*. The Avalon Project. URL: https://avalon.law.yale.edu/18th_century/let32.asp (visited on 12/15/2025).

— (Aug. 14, 1787). *Extract from Thomas Jefferson to George Washington*. Letter written in Paris; quoted in *Jefferson Quotes & Family Letters*. Monticello. URL: https://www.monticello.org/research-education/thomas-jefferson-encyclopedia/agriculture-quote/ (visited on 12/15/2025).

Jordà, Òscar (2009). "Simultaneous confidence regions for impulse responses". In: *The Review of Economics and Statistics* 91.3, pp. 629–647.

— (2023). "Local projections for applied economics". In: *Annual Review of Economics* 15.1, pp. 607–631.

Jordà, Òscar, Moritz Schularick, and Alan M Taylor (2015). "Betting the house". In: *Journal of international economics* 96, S2–S18.

Jordà, Òscar and Alan M Taylor (2025). "Local projections". In: *Journal of Economic Literature* 63.1, pp. 59–110.

Kleven, Henrik J (2021). "Sufficient statistics revisited". In: *Annual Review of Economics* 13.1, pp. 515–538.

Klosin, Sylvia and Adam Solomon (2024). "Optimal Insurance Scope: Theory and Evidence from US Crop Insurance". In: *Manuscript in preparation*.

LaPorte, Jonathan (Mar. 21, 2025). *The Corn–Soybean Ratio and Its Potential Impact on Farm Profits*. Michigan State University Extension. URL: https://www.canr.msu.edu/news/the-corn-soybean-ratio-and-its-potential-impact-on-farm-profits (visited on 09/21/2025).

Lashkaripour, Ahmad (2021). "The cost of a global tariff war: A sufficient statistics approach". In: *Journal of International Economics* 131, p. 103419.

Lashkaripour, Ahmad and Volodymyr Lugovskyy (2023). "Profits, scale economies, and the gains from trade and industrial policy". In: *American Economic Review* 113.10, pp. 2759–2808.

Lei, Xu and Zhang Qiao (2010). "Modeling agricultural catastrophic risk". In: *Agriculture and Agricultural Science Procedia* 1, pp. 251–257.

LeSage, James and Robert Kelley Pace (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC.

Licht, Mark and Sotirios Archontoulis (July 21, 2017). *Influence of Drought on Corn and Soybean*. Integrated Crop Management News, Iowa State University Extension and Outreach. URL: https://crops.extension.iastate.edu/cropnews/2017/07/influence-drought-corn-and-soybean.

Little, B et al. (2007). "Remote sensing and US crop insurance program integrity: Data mining satellite and agricultural data". In: *WIT Trans. Inf. Commun. Technol* 30, pp. 151–159.

Mahul, Olivier and Charles J Stutley (2010). *Government support to agricultural insurance: challenges and options for developing countries*. World Bank Publications.

Maisashvili, Aleksandre, Henry L Bryant, and Jason PH Jones (2020). "Implications of alternative crop insurance subsidies". In: *Journal of Agricultural and Applied Economics* 52.2, pp. 240–263.

Maisashvili, Aleksandre, Bart Fischer, and Henry Bryant (2023). "Crop insurance implications of permanently authorizing the emergency relief program". In: *Journal of Agricultural and Applied Economics* 55.1, pp. 171–193.

Meena, A et al. (2024). "Spatial Econometrics in Agriculture: Modelling Spatial Dependencies in Data". In: *Archives of Current Research International* 24.6, pp. 488–506.

Menne, Matthew J, Imke Durre, Bryant Korzeniewski, et al. (2012). "Global historical climatology network-daily (GHCN-Daily), Version 3". In: *(No Title)*.

Menne, Matthew J, Imke Durre, Russell S Vose, et al. (2012). "An overview of the global historical climatology network-daily database". In: *Journal of atmospheric and oceanic technology* 29.7, pp. 897–910.

Monke, Jim (May 2024). *Farm Bills: Major Legislative Actions, 1965–2024*. CRS Report R45210. Congressional Research Service. URL: https://sgp.fas.org/crs/misc/R45210.pdf.

Moran, Patrick AP (1950). "Notes on continuous stochastic phenomena". In: *Biometrika* 37.1/2, pp. 17–23.

National Sustainable Agriculture Coalition (2025). *Uninsured: Federal Crop Insurance Program Leaves Most Farms Unprotected*. Accessed: January 9, 2026. URL: https://sustainableagriculture.net/blog/uninsured-federal-crop-insurance-program-leaves-most-farms-unprotected/.

National Weather Service (2025). *NOAA Weather Radio County Coverage by State*. Accessed: January 7, 2026. National Oceanic and Atmospheric Administration. URL: https://www.weather.gov/nwr/counties.

Nelson, Carl H and Edna T Loehman (1987). "Further toward a theory of agricultural insurance". In: *American Journal of Agricultural Economics* 69.3, pp. 523–531.

Nickell, Stephen (1981). "Biases in dynamic models with fixed effects". In: *Econometrica: Journal of the econometric society*, pp. 1417–1426.

NOAA National Centers for Environmental Information (2025a). *GHCN-Daily Station Inventory (ghcnd-inventory.txt)*. https://www.ncei.noaa.gov/pub/data/ghcn/daily/ghcnd-inventory.txt. Accessed January 07, 2026.

— (2025b). *GHCN-Daily Station Metadata (ghcnd-stations.txt)*. https://www.ncei.noaa.gov/pub/data/ghcn/daily/ghcnd-stations.txt. Accessed January 07, 2026.

— (2025c). *README for Daily Global Historical Climatology Network (GHCN-Daily)*. https://www.ncei.noaa.gov/pub/data/ghcn/daily/readme.txt. Accessed on January 07, 2026.

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

OpenStreetMap contributors (2025). *OpenStreetMap*. https://www.openstreetmap.org. [Online; accessed 24-November-2025].

Paredes, Alexandre et al. (2022). "Does R&D tax credit impact firm behaviour? Micro evidence for Portugal". In: *Research Evaluation* 31.2, pp. 226–235.

Parliament of the Islamic Republic of Iran (May 1983). *Agricultural Products Insurance Law*. Official legislation. Adopted 21 May 1983 (31 Ordibehesht 1362); Published 26 July 1983 (4 Mordad 1362). URL: https://web.archive.org/web/20190629235755/http://rc.majlis.ir/fa/law/show/90745 (visited on 02/03/2025).

Pawlowski, Boguslaw, Rajinder Atwal, and Robin IM Dunbar (2008). "Sex differences in everyday risk-taking behavior in humans". In: *Evolutionary Psychology* 6.1, p. 147470490800600104.

Quiggin, John (1994). "The optimal design of crop insurance". In: *Economics of Agricultural Crop Insurance: Theory and Evidence*. Springer, pp. 115–134.

Regmi, Madhav, Jesse Tack, and Allen M. Featherstone (2023). "Does crop insurance influence crop yield impacts of warming temperatures? A farm-level analysis". In: *Journal of the Agricultural and Applied Economics Association* 2.4, pp. 808–822.

Roberts, Michael J, Erik J O'Donoghue, and Nigel D Key (2007). "Does Crop Insurance Affect Crop Yields?" In.

Rocchio, E. Laura (Apr. 2006). *Saving Millions in Government Dollars: Landsat Helps Fight Crop Insurance Fraud*. https://landsat.gsfc.nasa.gov/article/saving-millions-in-government-dollars-landsat-helps-fight-crop-insurance-fraud/?utm_source=chatgpt.com. NASA Landsat Science.

Roosevelt, Franklin D. (Feb. 18, 1937). *Message to Congress on Crop Insurance*. Online by Gerhard Peters and John T. Woolley. URL: https://www.presidency.ucsb.edu/node/209338 (visited on 12/15/2025).

Rosa, I. (2018). *Federal Crop Insurance: Program Overview for the 115th Congress*. Report R45193. Congressional Research Service.

Sabins Jr, Floyd F and James M Ellis (2020). *Remote sensing: Principles, interpretation, and applications*. Waveland Press.

Sadoulet, Elisabeth, Alain De Janvry, and Benjamin Davis (2001). "Cash transfer programs with income multipliers: PROCAMPO in Mexico". In: *World development* 29.6, pp. 1043–1056.

Schroeder, Jonathan et al. (2025a). *IPUMS National Historical Geographic Information System: Version 20.0*. Minneapolis, MN: IPUMS.

— (2025b). *IPUMS National Historical Geographic Information System: Version 20.0 [dataset]*. Dataset. Minneapolis, MN. DOI: 10.18128/D050.V20.0.

Seim, Katja (2006). "An empirical model of firm entry with endogenous product-type choices". In: *The RAND Journal of Economics* 37.3, pp. 619–640.

Smith, Vincent H and Joseph W Glauber (2012). "Agricultural insurance in developed countries: where have we been and where are we going?" In: *Applied Economic Perspectives and Policy* 34.3, pp. 363–390.

Staiger, Douglas and James H Stock (1997). "Instrumental Variables Regression with Weak Instruments". In: *Econometrica* 65.3, pp. 557–586.

Stigler, Matthieu and David Lobell (2020). "On the benefits of index insurance in US agriculture: a large-scale analysis using satellite data". In: *arXiv preprint arXiv:2011.12544*.

Stock, James H and Motohiro Yogo (2002). *Testing for weak instruments in linear IV regression*.

Tibshirani, Robert (1996). "Regression shrinkage and selection via the lasso". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 58.1, pp. 267–288.

Tsiboe, Francis and Dylan Turner (2023a). "Econometric identification of crop insurance participation". In: *Agricultural and Resource Economics Review* 52.3, pp. 476–497.

— (2023b). "The Crop Insurance Demand Response to Premium Subsidies: Evidence from US Agriculture". In: *Food Policy*.

U.S. Department of Agriculture, Risk Management Agency (2025). *RMA Information Reporting System (RIRS)*. Accessed January 8, 2026. U.S. Department of Agriculture. URL: hhttps://webapp.rma.usda.gov/apps/RIRS/.

U.S. Department of Agriculture, Risk Management Agency (RMA) (2025). *State/County/Crop Summary of Business*. U.S. Department of Agriculture. URL: https://www.rma.usda.gov/tools-reports/summary-of-business/state-county-crop-summary-business (visited on 11/08/2025).

U.S. Department of Labor, Bureau of Labor Statistics (2016). *Producer Price Index*. U.S. Department of Labor, Bureau of Labor Statistics. Accessed: September 21, 2025. URL: https://www.bls.gov/ppi/detailed-report/.

U.S. Geological Survey (1990). *County Boundaries of the United States, 1990. USGS*.

U.S. Government Accountability Office (May 1999). *Farm Service Agency: Characteristics of Small County Offices*. Letter Report GAO/RCED-99-162. Accessed October 30, 2025. U.S. Government Accountability Office (GAO). URL: https://www.govinfo.gov/content/pkg/GAOREPORTS-RCED-99-162/html/GAOREPORTS-RCED-99-162.htm.

Ugarte-Ruiz, Alfonso (July 2025a). *Locproj & Lpgraph: Stata Commands to Estimate Local Projections*. Working Paper WP 25-09. Accessed October 28, 2025. BBVA Research. URL: https://www.bbvaresearch.com/en/publications/locproj-lpgraph-stata-commands-to-estimate-local-projections/.

— (2025b). "LOCPROJ: Stata module to estimate local projections". In.

United States Department of Agriculture, National Agricultural Statistics Service (2025). *Data – Large Datasets*. https://www.nass.usda.gov/datasets/. Accessed 16 March 2025.

United States Department of Agriculture, Risk Management Agency (Feb. 2008). *Rate Methodology Handbook: Actual Production History (FCIC-11010 RMH-APH)*. Handbook FCIC-11010. Federal Crop Insurance Corporation (FCIC). URL: https://www.rma.usda.gov/sites/default/files/topics/ratemethodology.pdf.

— (June 2025). *RMA Volatility Factor Calculation Methodology*. Tech. rep. Accessed: 2025-11-19. Risk Management Agency (USDA). URL: https://www.rma.usda.gov/sites/default/files/2025-06/volatilitymethodology.pdf.

USDA Economic Research Service (Jan. 31, 2025). *Educational attainment for adults age 25 and older for the United States, States, and counties, 1970–2023. County-level Data Sets: Download Data*. URL: https://www.ers.usda.gov/data-products/county-level-data-sets/county-level-data-sets-download-data (visited on 02/24/2025).

USDA Farm Service Agency (2025). *USDA Service Center Locator*. U.S. Department of Agriculture. URL: https://offices.sc.egov.usda.gov/locator/app (visited on 11/24/2025).

Vergara, Oscar et al. (2008). "Modeling the potential impact of catastrophic weather on crop insurance industry portfolio losses". In: *American Journal of Agricultural Economics* 90.5, pp. 1256–1262.

Walters, Cory and Richard Preston (2018). "Net income risk, crop insurance and hedging". In: *Agricultural Finance Review* 78.1, pp. 135–151.

Williams, Richard (2012). "Using the margins command to estimate and interpret adjusted predictions and marginal effects". In: *The Stata Journal* 12.2, pp. 308–331.

Wilson, Carol (Sept. 2023). *Federal Workforce Statistics Sources: OPM and OMB*. CRS Report R43590. Accessed November 2025. Congressional Research Service, Library of Congress. URL: https://crsreports.congress.gov/product/pdf/R/R43590.

Woodard, Joshua D and Jing Yi (2020). "Estimation of insurance deductible demand under endogenous premium rates". In: *Journal of Risk and Insurance* 87.2, pp. 477–500.

Wooldridge, Jeffrey M (2015). "Control function methods in applied econometrics". In: *Journal of Human Resources* 50.2, pp. 420–445.

World Wide Fund for Nature (WWF) (2025). *Tackling the insurance protection gap: Leveraging climate mitigation and nature to increase resilience*. White paper. World Wide Fund for Nature.

Wright, Brian D and Julie A Hewitt (1994). "All-risk crop insurance: lessons from theory and experience". In: *Economics of agricultural crop insurance: theory and evidence*. Springer, pp. 73–112.

Yi, Fujin, Wuyi Lu, and Yingheng Zhou (2016). "Cash transfers and multiplier effect: Lessons from the grain subsidy program in China". In: *China Agricultural Economic Review* 8.1, pp. 81–99.

Yu, Jialing (2017). "Welfare Effects of the Federal Crop Insurance Program: The Sufficient Statistics Approach". In.

Yu, Jisang, Aaron Smith, and Daniel A Sumner (2018). "Effects of crop insurance premium subsidies on crop acreage". In: *American Journal of Agricultural Economics* 100.1, pp. 91–114.

Appendices

A Parameters of Interest

In this paper, we often referred to technical expressions in the context of agricultural insurance. To be on the same page, we provide an explanation for the major terms from a Handbook of the United States Department of Agriculture, Risk Management Agency (2008)[20].

Liability is a measure of the USDA's exposure to loss for a given farmer or group of farmers. Liability is the maximum amount of insurance coverage that a farmer has under their policy for a given crop and year. It represents the total insured value of the crop.

$$\text{Liability} = \text{Acres Planted} \times \text{Expected Yield} \times \text{Selected Coverage Level} \times \text{Base Price} \times \text{Price Election Percentage} \tag{A.1}$$

- Farmers do not pay the liability itself. They pay a premium minus the subsidy to receive insurance coverage.
- When a covered loss occurs, the insurer (usually a private company that provides the federal crop insurance) pays the farmer an indemnity.
- Indemnity payment is triggered when the actual value of production falls below the liability.
  - If actual revenue $<$ liability $\rightarrow$ indemnity payment.
  - If actual revenue $\geq$ liability $\rightarrow$ no indemnity payment.

Indemnity is the amount paid to a farmer who suffers a covered loss. Indemnity is paid when the value of production is less than the purchased liability. The amount of indemnity paid equals the liability minus the value of the production, where they are defined as follows:

$$\text{Indemnity} = \text{Liability} - \text{Value of Production} \tag{A.2}$$

$$\text{Value of Production} = \text{Acres Planted} \times \text{Actual Yield} \times \text{Base Price} \times \text{Price Election Percentage} \tag{A.3}$$

Finally, **premium**, is the price that farmers pay for insurance, and it equals

$$\text{Premium} = \text{Liability} \times \text{Premium Rate} \times \text{Adjustment Factor} \tag{A.4}$$

B Causality

We are interested in the causal relationship between enrollment of farmers in FCIC and their crop yield. For simplicity, let's assume enrollment in insurance is a dummy variable, and we denote it with C_i (see also Angrist and Pischke 2009, sec. 3.2). To address this question, we imagine two potential yield outcomes:

[20] See page 50.

$$\text{Potential Outcome} = \begin{cases} Y_{1i} & \text{if } C_i = 1, \\ Y_{0i} & \text{if } C_i = 0 \end{cases}$$

In this case, $Y_1 i$ and $Y_0 i$ are the farmer i's yield when insured and not insured, respectively. We would like to know the difference between $Y_1 i$ and $Y_0 i$; which is the causal effect of insurance enrollment for farmer i. The observed outcome, Y_i, can be written in terms of potential outcomes as

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})C_i$$

We aim to measure $Y_{1i} - Y_{0i}$ for those who purchased insurance. But the challenge is that we get to see one of Y_{1i} and Y_{0i}, but never both. To simplify the analysis, we assume that the conditional independence assumption (CIA) holds (see also Angrist and Pischke 2009, p. 53). Let us call all other covariates X_i. The CIA would assert that conditional on observed characteristics, X_i, selection bias for buying insurance disappears. Formally, this means

$$\{Y_{0i}, Y_{1i}\} \perp\!\!\!\perp C_i \mid X_i$$

Or alternatively, it is equivalent to

$$E[Y_i \mid X_i, C_i = 1] - E[Y_i \mid X_i, C_i = 0] = E[Y_{1i} - Y_{0i} \mid X_i]$$

Next, let us consider the insurance coverage as a continuous variable. The causal relationship between the level of coverage and the crop yield is likely to be different for each farmer (or county). We therefore use individual-specific functional notation,

$$Y_{\theta i} \equiv f_i(\theta),$$

to denote the potential yield that farmer i would obtain at coverage level θ. $f_i(\theta)$ answers the causal "what if" questions. The CIA in this more general setup becomes

$$\{Y_{\theta i}\} \perp\!\!\!\perp \theta_i \mid X_i \qquad \text{for all } \theta$$

In an observational study like ours, the CIA means that θ_i is "as good as randomly assigned", conditional on X_i. Conditional on X_i, the average causal effect of one extra percentage point of insurance coverage is $E[f_i(\theta) - f_i(\theta - 1) \mid X_i]$, while the average causal effect of 5 percentage (a typical coverage incremental step in the federal crop insurance program) increase in insurance coverage is $E[f_i(\theta) - f_i(\theta - 5) \mid X_i]$. The data reveals only $Y_i \equiv f_i(\theta_i)$, that is, $f_i(\theta)$ for $\theta = \theta_i$. But given the CIA, conditional on X_i, comparisons of average yield with different levels of insurance coverage have a causal interpretation. In other words, for any value of θ,

$$E[Y_i \mid X_i, \theta_i = \theta] - E[Y_i \mid X_i, \theta_i = \theta - 1] = E[f_i(\theta) - f_i(\theta - 1) \mid X_i] \tag{B.1}$$

Using the law of iterated expectations, the unconditional (overall) average causal effect would be

$$\begin{aligned} E\Big[\, E(Y_i \mid X_i, \theta_i = \theta) - E(Y_i \mid X_i, \theta_i = \theta - 1) \,\Big] &= E\Big[\, E\big(f_i(\theta) - f_i(\theta - 1) \,\big|\, X_i\big) \Big] \\ &= E[f_i(\theta) - f_i(\theta - 1)] \end{aligned} \tag{B.2}$$

C Omitted Variable Bias

A riskier crop or a riskier county has a higher premium rate for a given crop insurance choice (Coble and Barnett 2013). As a result, farmers tend to plant less risky crops and plant less in riskier counties, ceteris paribus. Riskiness is positively correlated with Γ_ijt and negatively correlated with Y_{ijt}. So, omitting the "riskiness" variable that has a positive correlation with Γ_ijt, would cause a downward bias in OLS estimate for coefficient of Γ_ijt. To verify this, let's define the long and short regressions as discussed by Angrist and Pischke (2009, pp. 59–60) as follows:

$$\textit{Long regression:} \quad \ln Y_{ijt} = \alpha_0 + \alpha_1 \Gamma_{ijt} + \alpha_2 \Omega_{ijt} + \epsilon_{ijt}$$

In which Ω denotes "riskiness". Absent the riskiness, the regression would change to

$$\textit{Short regression:} \quad \ln Y_{ijt} = \tilde{\alpha}_0 + \tilde{\alpha}_1 \Gamma_{ijt} + \eta_{ijt}$$

Then the omitted variable bias formula Angrist and Pischke (2009, p. 59) would specify the relationship between α_1, $\tilde{\alpha}_1$ and α_2 as

$$\tilde{\alpha}_1 = \alpha_1 + \alpha_2 \, \delta_{\Gamma,\text{riskiness}} = \alpha_1 + \alpha_2 \underbrace{Cov(\Gamma_{ijt}, \Omega_{ijt})}_{\textit{omitted variable bias}}$$

α_2 is the effect of the omitted control (i.e. riskiness) on the outcome, and $\delta_{\Gamma,\Omega}$ is the correlation of those omitted controls with the variables of interest. Having $Cov(\Gamma_{ijt}, \Omega_{ijt}) > 0$ and $\alpha_2 < 0$, would yield

$$\text{Omitted variable bias} = \alpha_2 \, Cov(\Gamma_{ijt}, \Omega_{ijt}) < 0$$

Therefore, the omitted variable bias is downward. This completes the proof. [0.05in]

A second source of endogeneity is the dependence of Γ_ijt on the choice of coverage level. Although the sign of relationship is not determined. For the sake of brevity, let's focus on the sublia as a measure of intensity of insurance enrollment. Following Jisang Yu, A. Smith, and Sumner (2018), we write the premium subsidy per dollar of insured liability as $\Gamma_j(\theta_j) = s(\theta_j)p_j(\theta_j)$, where $s(\theta_j)$ is the subsidy rate and $p_j(\theta_j)$ is the premium. Then we will have:

- An increase in the chosen coverage level for a crop may cause the subsidy rate to decrease because the subsidy rate is lower for the higher coverage levels ($s'(\theta_j) < 0$),
- The premium rate increases with the coverage level ($p'_j(\theta_j) > 0$).

So, the sign of $\Gamma'_j(\theta_j)$ is unknown. To determine is unknown. To determine the sign of omitted variable bias from absence of coverage level in the model, we set up the long and short regressions as follows:

$$\textit{Long regression:} \quad \ln Y_{ijt} = \alpha_0 + \alpha_1 \Gamma_{ijt} + \beta_\theta \theta_j + \epsilon$$

Absent the riskiness, the regression would change to

$$\textit{Short regression:} \quad \ln Y_{ijt} = \tilde{\alpha}_0 + \tilde{\alpha}_1 \Gamma_{ijt} + u$$

Then the omitted variable bias formula would specify the relationship among α_1, $\tilde{\alpha}_1$ and β_θ as

$$\tilde{\alpha}_1 = \alpha_1 + \beta_2 \underbrace{Cov(\Gamma_{ijt}, \theta_j)}_{omitted\ variable\ bias}$$

We know $\beta_\theta > 0$. Then we would have

$$if \begin{cases} Cov(\Gamma_{ijt}, \theta_j) > 0 & \rightarrow \quad OVB > 0, \\ Cov(\Gamma_{ijt}, \theta_j) < 0 & \rightarrow \quad OVB < 0 \end{cases}$$

So, if Γ and θ are positively (negatively) correlated, and θ has a positive impact on crop acreage, the OLS would underestimate (overestimate) the impact of Γ on acreage.

D Marginal Effects Tables

Tables D1-D8 are the bases for Figures 6-13, respectively.

Table D1 – Marginal change in sublia at different values of number of heating degree days, taking all other variables in Equation 15 at their average levels.

Percentiles are as follows: 1: 1%, 2: 5%, 3: 10%, 4: 25%, 5: mean, 6: 75%, 7: 90%, 8: 95% and 9: 99%.

	Margin	std. err.	z	P>|z|	[95% conf. interval]	
_at					lower	upper
1	−6.350 519	3.495 379	−1.82	0.069	−13.201 34	0.500 298 7
2	−6.369 202	3.496 005	−1.82	0.068	−13.221 25	0.482 842 8
3	−6.392 519	3.496 811	−1.83	0.068	−13.246 14	0.461 105 1
4	−6.478 521	3.500 023	−1.85	0.064	−13.338 44	0.381 398 2
5	−6.829 677	3.517 018	−1.94	0.052	−13.722 91	0.063 551 7
6	−7.011 729	3.528 262	−1.99	0.047	−13.926 99	−0.096 462 6
7	−7.484 423	3.565 075	−2.10	0.036	−14.471 84	−0.497 004 8
8	−7.847 696	3.600 672	−2.18	0.029	−14.904 88	−0.790 508 5
9	−8.636 591	3.698 866	−2.33	0.020	−15.886 24	−1.386 947

Table D2 – Marginal change in sublia at different values of number of moderate degree days, taking all other variables in Equation 15 at their average levels.

Percentiles are as follows: 1: 1%, 2: 5%, 3: 10%, 4: 25%, 5: mean, 6: 75%, 7: 90%, 8: 95% and 9: 99%.

	Margin	std. err.	z	P>|z|	[95% conf. interval]	
_at					lower	upper
1	−0.835 372	2.507 928	−0.33	0.739	−5.750 82	4.080 076
2	−0.094 631 3	2.429 246	−0.04	0.969	−4.855 867	4.666 604
3	0.374 620 7	2.388 565	0.16	0.875	−4.306 881	5.056 122
4	1.178 171	2.336 635	0.50	0.614	−3.401 55	5.757 893
5	2.181 952	2.305 083	0.95	0.344	−2.335 929	6.699 832
6	3.176 285	2.311 63	1.37	0.169	−1.354 425	7.706 996
7	4.065 262	2.349 121	1.73	0.084	−0.538 930 6	8.669 455
8	4.583 149	2.384 209	1.92	0.055	−0.089 814 4	9.256 113
9	5.369 331	2.455 002	2.19	0.029	0.557 615 9	10.181 05

Table D3 – Marginal change in sublia at different values of precipitation, taking all other variables in Equation 15 at their average levels.

Percentiles are as follows: 1: 1%, 2: 5%, 3: 10%, 4: 25%, 5: mean, 6: 75%, 7: 90%, 8: 95% and 9: 99%.

	Margin	std. err.	z	P>|z|	[95% conf. interval]	
_at					lower	upper
1	−12.346 94	2.612 372	−4.73	0.000	−17.467 10	−7.226 789
2	−13.348 20	2.545 514	−5.24	0.000	−18.337 32	−8.359 090
3	−13.936 16	2.520 123	−5.53	0.000	−18.875 51	−8.996 813
4	−15.150 67	2.501 683	−6.06	0.000	−20.053 88	−10.247 460
5	−16.891 88	2.555 890	−6.61	0.000	−21.901 34	−11.882 430
6	−18.317 22	2.667 181	−6.87	0.000	−23.544 80	−13.089 640
7	−20.038 92	2.871 306	−6.98	0.000	−25.666 58	−14.411 260
8	−21.309 33	3.062 539	−6.96	0.000	−27.311 80	−15.306 860
9	−24.912 59	3.739 803	−6.66	0.000	−32.242 47	−17.582 710

Table D4 – Marginal change in sublia at different values of precipitation squared, taking all other variables in Equation 15 at their average levels.

Percentiles are as follows: 1: 1%, 2: 5%, 3: 10%, 4: 25%, 5: mean, 6: 75%, 7: 90%, 8: 95% and 9: 99%.

	Margin	std. err.	z	P>|z|	[95% conf. interval]	
_at					lower	upper
1	−4.204 523	3.788 029	−1.11	0.267	−11.628 92	3.219 877
2	−3.433 429	3.900 081	−0.88	0.379	−11.077 45	4.210 59
3	−2.920 682	3.976 266	−0.73	0.463	−10.714 02	4.872 657
4	−1.739 376	4.156 425	−0.42	0.676	−9.885 82	6.407 068
5	0.631 242 3	4.534 779	0.14	0.889	−8.256 762	9.519 246
6	2.161 364	4.788 885	0.45	0.652	−7.224 678	11.547 41
7	4.784 996	5.238 917	0.91	0.361	−5.483 093	15.053 08
8	6.938 978	5.619 31	1.23	0.217	−4.074 666	17.952 62
9	14.142 65	6.940 934	2.04	0.042	0.538 665 8	27.746 63

Table D5 – Marginal change in fracInsured at different values of number of heating degree days, taking all other variables in Equation 15 at their average levels.

Percentiles are as follows: 1: 1%, 2: 5%, 3: 10%, 4: 25%, 5: mean, 6: 75%, 7: 90%, 8: 95% and 9: 99%.

	Margin	std. err.	z	P>|z|	[95% conf. interval]	
_at					lower	upper
1	0.081 928 4	0.062 291 3	1.32	0.188	−0.040 160 2	0.204 017
2	0.082 351	0.062 430 7	1.32	0.187	−0.040 010 8	0.204 712 8
3	0.082 878 4	0.062 605 4	1.32	0.186	−0.039 825 9	0.205 582 8
4	0.084 823 8	0.063 257 4	1.34	0.180	−0.039 158 5	0.208 806
5	0.092 766 9	0.066 033 5	1.40	0.160	−0.036 656 4	0.222 190 1
6	0.096 884 8	0.067 539 5	1.43	0.151	−0.035 490 2	0.229 259 9
7	0.107 577 1	0.071 638 1	1.50	0.133	−0.032 831	0.247 985 2
8	0.115 794 3	0.074 950 7	1.54	0.122	−0.031 106 4	0.262 694 9
9	0.133 638 9	0.082 543 3	1.62	0.105	−0.028 143	0.295 420 9

Table D6 – Marginal change in fracInsured at different values of number of moderate degree days, taking all other variables in Equation 15 at their average levels.

Percentiles are as follows: 1: 1%, 2: 5%, 3: 10%, 4: 25%, 5: mean, 6: 75%, 7: 90%, 8: 95% and 9: 99%.

	Margin	std. err.	z	P>|z|	[95% conf. interval]	
_at					lower	upper
1	−0.021 019 6	0.039 574 9	−0.53	0.595	−0.098 584 9	0.056 545 7
2	−0.034 843 2	0.041 758 0	−0.83	0.404	−0.116 687 4	0.047 001 1
3	−0.043 600 3	0.043 772 9	−1.00	0.319	−0.129 393 5	0.042 193 0
4	−0.058 596 0	0.048 155 6	−1.22	0.224	−0.152 979 3	0.035 787 3
5	−0.077 328 3	0.054 895 1	−1.41	0.159	−0.184 920 8	0.030 264 1
6	−0.095 884 4	0.062 536 8	−1.53	0.125	−0.218 454 4	0.026 685 5
7	−0.112 474 3	0.069 925 8	−1.61	0.108	−0.249 526 4	0.024 577 7
8	−0.122 139 1	0.074 408 4	−1.64	0.101	−0.267 976 8	0.023 698 7
9	−0.136 810 6	0.081 405 3	−1.68	0.093	−0.296 362 1	0.022 740 9

Table D7 – Marginal change in fracInsured at different values of precipitation, taking all other variables in Equation 15 at their average levels.

Percentiles are as follows: 1: 1%, 2: 5%, 3: 10%, 4: 25%, 5: mean, 6: 75%, 7: 90%, 8: 95% and 9: 99%.

	Margin	std. err.	z	P>|z|	[95% conf. interval]	
_at					lower	upper
1	0.133 148 4	0.067 510 5	1.97	0.049	0.000 830 3	0.265 466 5
2	0.141 715 9	0.070 562 6	2.01	0.045	0.003 415 7	0.280 016 2
3	0.146 746 9	0.072 577 7	2.02	0.043	0.004 497 4	0.288 996 5
4	0.157 139 2	0.077 196 3	2.04	0.042	0.005 837 2	0.308 441 1
5	0.172 038 3	0.084 715 1	2.03	0.042	0.005 999 8	0.338 076 8
6	0.184 234 6	0.091 493 3	2.01	0.044	0.004 910 9	0.363 558 2
7	0.198 966 7	0.100 256 1	1.98	0.047	0.002 468 4	0.395 465 1
8	0.209 837 3	0.107 040 9	1.96	0.050	0.000 041 0	0.419 633 6
9	0.240 669 5	0.127 345 0	1.89	0.059	−0.008 922 2	0.490 261 2

Table D8 – Marginal change in fracInsured at different values of precipitation squared, taking all other variables in Equation 15 at their average levels.

Percentiles are as follows: 1: 1%, 2: 5%, 3: 10%, 4: 25%, 5: mean, 6: 75%, 7: 90%, 8: 95% and 9: 99%.

	Margin	std. err.	z	P>|z|	[95% conf. interval]	
_at					lower	upper
1	0.067 222 6	0.064 941 4	1.04	0.301	−0.060 060 1	0.194 505 4
2	0.061 974 9	0.066 199 8	0.94	0.349	−0.067 774 3	0.191 724 2
3	0.058 485 4	0.067 111 4	0.87	0.384	−0.073 055 6	0.190 026 4
4	0.050 446 0	0.069 440 4	0.73	0.468	−0.085 654 8	0.186 546 7
5	0.034 312 7	0.074 931 5	0.46	0.647	−0.112 550 4	0.181 175 7
6	0.023 899 4	0.078 971 2	0.30	0.762	−0.130 881 4	0.178 680 2
7	0.006 044 2	0.086 622 3	0.07	0.944	−0.163 732 4	0.175 820 7
8	−0.008 614 8	0.093 452 2	−0.09	0.927	−0.191 777 8	0.174 548 1
9	−0.057 639 6	0.118 698 3	−0.49	0.627	−0.290 284 1	0.175 004 8

E Crop Insurance Subsidies in the World

United States is not alone in its provision of governmental crop insurance. Japan started its public crop insurance in 1939 as a rent insurance, then developed into yield insurance in 1947. In Sweden, compulsory crop insurance has been in effect since 1961 (Wright and Hewitt 1994). Brazil started its public crop insurance in 1973 (Brazil Ministry of Agriculture, Livestock and Food Supply 2016). Mexico began its all-risk crop insurance in 1964. Philippines had a credit insurance type scheme in 1970s, which evolved later into crop insurance in 1981. Sri Lanka's crop insurance has been in operation since 1958. In Mauritius, it began in 1945. India began a nation-wide crop insurance in 1985 (Wright and Hewitt 1994). Iran also started its crop insurance in 1984 (Parliament of the Islamic Republic of Iran 1983). The next three paragraphs are from Chintapalli and Tang (2022).

To sustain small farmers' operations, governments of developing countries offer various programs to support and protect their income. A common farmer-support program is input cost subsidies, wherein a government offers financial assistance to farmers by bearing a fraction of the farmers' cultivation costs. Input subsidy programs (ISPs) are adopted by many emerging economies such as Africa and India. In Africa, smallholder farmers are issued coupons that they can redeem for input packets from both government and private vendors (Jayne et al., 2018). In India, central and individual state government subsidies for seeds, irrigation, power, and fertilizers are offered to farmers (see Hemming et al., 2018, and the references therein).

Besides cost subsidy, an alternative support program is the credit-based minimum support price (MSP; also known as deficiency payment). Under a credit-based MSP, a government will not take any possession of the crop; instead, it will provide a monetary credit to farmers cultivating the crop if the prevailing market price of the crop is below the prespecified MSP. For example, in India, credit-based MSPs are currently offered for eight crops (mostly oilseeds and legumes) for which the Indian government takes no (or very low) possession of crop (Bera, 2017).

Cost subsidies can occur in different forms; for instance, the governments in Mali, Ghana, Nigeria, and other such African countries offer subsidies for seeds and fertilizers to smallholder farmers (Brooks & Wiggins, 2010; Jayne & Rashid, 2013; Jayne et al., 2018). In India, subsidies are offered not only on direct inputs such as seeds and fertilizers (Prasad, 2016) but also on other resources such as power and irrigation by both central and state governments (Bardhan & Mookherjee, 2011; Fan et al., 2008). Xu et al. (2009) find that fertilizer subsidies may reduce or surge the commercial demand for fertilizers in the area depending on whether the private sector is active or inactive. Chibwana et al. (2012) show that input subsidies can entice farmers to allocate more land to those subsidized crops in Malawi. We refer the readers to Hemming et al. (2018) for a comprehensive review of the various types of input subsidies provided to smallholder farmers.

Mahul and Stutley (2010): Agricultural (crop and livestock) insurance is currently available in more than 100 countries, either as well-developed programs or pilots (figure 1.3; see also appendix A). The majority of high-income countries (58 percent) have well-established agricultural insurance markets. In contrast, only 35 percent of low- and middle-income countries offer such products and programs. The availability of agricultural insurance is particularly low in low-income countries (8 percent). Pilot programs, which reach only a limited number of farmers and herders, are being implemented in vari-

ous forms (including named-peril crop insurance, index-based crop insurance, or livestock insurance) in eight middle-income and eight low-income countries (table 1.4). to add: map of availability of agricultural insurance in different countries with four categories: yes, no, pilot and unknown.

World Bank conducted a survey of insurance companies in 65 countries in 2008. The key objectives of the survey were to document international experience with public and private agricultural insurance in developed and developing economies and to examine the ways in which governments support agricultural insurance. The focus was on crop and livestock insurance and excluded the forestry or aquaculture insurance.