Utilizing geo-referenced and “big-ag” data to improve US agricultural policy

by

Francis Tsiboe

B.S., University of Ghana, 2011
M.S., University of Arkansas, 2015

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Agricultural Economics
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2021

Abstract

Study 1: Utilizing Topographic and Soil Features to Improve Rating for Farm-level Insurance Products

Previous studies have shown a strong correlation between topographic/soil features and agricultural production; however, linkages between these features and agricultural insurance products are scarce. Agricultural insurance is an ever-growing means of governmental support for producers globally. However, failure to set insurance premiums that accurately reflect risk exposure can lead to low participation rates and/or adverse selection. The U.S. federal crop insurance program partly guards against this at the farm-level by inducing pricing heterogeneity via a rate multiplier curve, which does not consider topographic/soil information. We develop a method for econometrically incorporating this information into existing rating procedures used by the Risk Management Agency (RMA). The empirical application leverages 149,267 farm-level observations of Kansas producers across four dryland crops (corn, soybean, sorghum, and wheat), spanning 46 years, and matched to fine-scale topographic/soil features. The results suggest that incorporating these features does improve the prediction accuracy of yield losses and can, in general, improve rating performance. However, these improvements are specific to farms with limited yield histories, as there are no improvements for farms with the commonly used yield history of ten years. This suggests substantial rating improvements for new farms or those with limited histories for a particular crop, but more general improvements for the program are not likely to occur given a large number of current participants with a full ten-year yield history.

Study 2: Tradeoffs Between Production-History-Based and Index-Based Insurance for Field Crops

Agricultural insurance products based on Actual Production History (APH) typically suffer from adverse selection, moral hazard, and high program costs associated with pricing, loss assessment, and monitoring. On the contrary, Index-based insurance offers the opportunity of reducing, and even sometimes eliminating, some of these concerns; however, by design, they cannot guarantee that indemnities will be paid when producers experience losses. This concern is commonly referred to as basis risk and is the biggest limiting factor in the potential expansion of Index insurance programs. An extensive body of literature has shown that basis risk could be reduced to an appreciable extent by improving product design. Nonetheless, a knowledge gap on farm-level tradeoffs between APH- and Index-based insurance exists because observable data is limited. The novelty of this study is that it overcomes these limitations and extends the literature by providing *ex-post* simulated evidence of the tradeoffs between Index-based and APH-based insurance at the farm level. Using a sample of 5,428 corn, soybean, sorghum, and wheat KS farms from 1973-2018 the study shows that economically significant tradeoffs do exist between APH- and Index-based insurance and that different types of index products are associated with differing levels of basis risk. Index-based insurance that protects against killing-degree-days (i.e., degree-days >30 °C) accumulation generates the most significant gains in economic rents and is associated with relatively low basis risk.

Study 3: The Potential Significance of “Big Ag Data” in Corn Futures Markets

The advent of precision agriculture technologies has left researchers to grapple with how to best-use its associated “Big Ag-Data”. While the wealth of information output from precision equipment can easily be aggregated to a higher level in real-time, this poses an interesting question of whether aggregated real-time data will be relevant vis-à-vis periodic information from public sources. To this end, this study utilized advances in event study and yield projection methodologies to test the potential market value of simulated live streamed yield monitor data vis-à-vis USDA report yields. The results shows that the market for corn exhibits only semi-strong form efficiency, as the “news” provided by the monthly Crop Production and World Agricultural Supply and Demand Estimates reports is incorporated into prices in at most two days after the release. As expected, an increase in corn yields relative to what was publicly known, elicits a futures price decrease. On the contrary, live-streamed yield information does not significantly correlate with historic market reactions. Nonetheless, this study advances the market-price event-study methodology by utilizing sources of information not previously considered. Second, the study provides policy implications centered around the ongoing debate about the economic significance of USDA reports in the presence of growing information availability in the private sector.

Utilizing geo-referenced and “big-ag” data to improve US agricultural policy

by

Francis Tsiboe

B.S., University of Ghana, 2011
M.S., University of Arkansas, 2015

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Agricultural Economics
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2021

Approved by:

Major Professor
Jesse Tack

Copyright

Abstract

Study 1: Utilizing Topographic and Soil Features to Improve Rating for Farm-level Insurance Products

Previous studies have shown a strong correlation between topographic/soil features and agricultural production; however, linkages between these features and agricultural insurance products are scarce. Agricultural insurance is an ever-growing means of governmental support for producers globally. However, failure to set insurance premiums that accurately reflect risk exposure can lead to low participation rates and/or adverse selection. The U.S. federal crop insurance program partly guards against this at the farm-level by inducing pricing heterogeneity via a rate multiplier curve, which does not consider topographic/soil information. We develop a method for econometrically incorporating this information into existing rating procedures used by the Risk Management Agency (RMA). The empirical application leverages 149,267 farm-level observations of Kansas producers across four dryland crops (corn, soybean, sorghum, and wheat), spanning 46 years, and matched to fine-scale topographic/soil features. The results suggest that incorporating these features does improve the prediction accuracy of yield losses and can, in general, improve rating performance. However, these improvements are specific to farms with limited yield histories, as there are no improvements for farms with the commonly used yield history of ten years. This suggests substantial rating improvements for new farms or those with limited histories for a particular crop, but more general improvements for the program are not likely to occur given a large number of current participants with a full ten-year yield history.

Study 2: Tradeoffs Between Production-History-Based and Index-Based Insurance for Field Crops

Agricultural insurance products based on Actual Production History (APH) typically suffer from adverse selection, moral hazard, and high program costs associated with pricing, loss assessment, and monitoring. On the contrary, Index-based insurance offers the opportunity of reducing, and even sometimes eliminating, some of these concerns; however, by design, they cannot guarantee that indemnities will be paid when producers experience losses. This concern is commonly referred to as basis risk and is the biggest limiting factor in the potential expansion of Index insurance programs. An extensive body of literature has shown that basis risk could be reduced to an appreciable extent by improving product design. Nonetheless, a knowledge gap on farm-level tradeoffs between APH- and Index-based insurance exists because observable data is limited. The novelty of this study is that it overcomes these limitations and extends the literature by providing *ex-post* simulated evidence of the tradeoffs between Index-based and APH-based insurance at the farm level. Using a sample of 5,428 corn, soybean, sorghum, and wheat KS farms from 1973-2018 the study shows that economically significant tradeoffs do exist between APH- and Index-based insurance and that different types of index products are associated with differing levels of basis risk. Index-based insurance that protects against killing-degree-days (i.e., degree-days >30 °C) accumulation generates the most significant gains in economic rents and is associated with relatively low basis risk.

Study 3: The Potential Significance of "Big Ag Data" in Corn Futures Markets

The advent of precision agriculture technologies has left researchers to grapple with how to best-use its associated "Big Ag-Data". While the wealth of information output from precision equipment can easily be aggregated to a higher level in real-time, this poses an interesting question of whether aggregated real-time data will be relevant vis-à-vis periodic information from public sources. To this end, this study utilized advances in event study and yield projection methodologies to test the potential market value of simulated live streamed yield monitor data vis-à-vis USDA report yields. The results shows that the market for corn exhibits only semi-strong form efficiency, as the "news" provided by the monthly Crop Production and World Agricultural Supply and Demand Estimates reports is incorporated into prices in at most two days after the release. As expected, an increase in corn yields relative to what was publicly known, elicits a futures price decrease. On the contrary, live-streamed yield information does not significantly correlate with historic market reactions. Nonetheless, this study advances the market-price event-study methodology by utilizing sources of information not previously considered. Second, the study provides policy implications centered around the ongoing debate about the economic significance of USDA reports in the presence of growing information availability in the private sector.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I stand tall because of God almighty and on the shoulders of a lot of people. For this reason, I first thank God almighty for the drive and grace throughout my scholarly training. I am grateful to Dr. Jesse Tack for his kindness and mentorship throughout my Ph.D. career. I appreciate the insight and help my supervisory committee offered. Thanks to Dr. Nathan Hendricks, Dr. Jisang Yu, Dr. Krishna Jagadish SV, and Dr. Stuart Heckman. I would like to thank the Department of Agricultural Economics for funding support. Thanks to Judy Duryee, Deanna Foster, and Amy Schmidt, for their welcoming smiles and kindness. To my wife, mother, and siblings, thanks for the love, laughter, well-wishes, and prayers, and understanding all this while. To my friends and colleagues, it is a pleasure knowing you. I cherish the friendship. It has been wonderful studying at Kansas State University!

Dedication

To my wife (Mary Ama Tsiboe) and daughter (Liana Owusua Tsiboe)

Chapter 1 - Introduction

The data revolution has been upon us for many years now, and public policy continues to grapple with how to best use the wealth of information currently at our disposal. It is not so much a question of whether new information can improve the efficiencies of providing public support and services for agriculture, it is more narrowly a question of what data to leverage and how to build it into existing programs. The wrong data can fail to deliver efficiencies at best, and make current programs less efficient at worst if too much noise is being built in. Consequently, this dissertation consists of three studies analyzing the potential of utilizing geo-referenced and "Big-Ag" data to improve US agricultural policy from the angle of risk management and farm support.

Globally, farmers and ranchers are currently facing increased climatic and market uncertainty in the coming decades, which suggests that support provided through agricultural insurance is likely to become more critical (Mahul and Stutley 2010; Smith and Glauber 2012). The World Bank shows that before 2008, the global agricultural insurance market generated $15 billion in premiums, which helped producers across 65 advanced and emerging countries cover losses and stabilize revenues (Mahul and Stutley 2010). While the product space of the agricultural insurance market is diverse, they can be segregated into two broad groups based on the mode of indemnity trigger and pricing, i.e., Actual Production History [APH] based and Index-based schemes. Generally, agricultural insurance products are plagued by (1) adverse selection, the case whereby the insurance pool becomes riskier due to mispricing of products, which can lead to low participation as in the case of the Federal Crop Insurance Program (FCIP) before 2002 (Smith, Glauber and Goodwin 2017; Glauber 2013); (2) moral hazard, where insureds assume riskier activities to increase the chance of indemnification (Chambers 1989; Horowitz and Lichtenberg 1993; Yu and Hendricks 2020; Park et al. 2020); (3) basis risk, where there is real or perceived disconnects between product and outcome (Jensen and Barrett 2017; Carter et al. 2017; Barnett and Mahul 2007); and (4) high program costs associated with pricing, loss assessment, and monitoring (Barnett and Mahul 2007). These problems can be reduced to an appreciable extent if the design of the suite of products is improved. The motivation of the first two studies in this dissertation is to extend the literature by proposing data-driven ways of improving the designing of crop insurance to address the stated problems.

The first study in Chapter 2 ascertains the feasibility and potential economic gains of using soil quality attributes for setting crop insurance premium rates at the farm level by pursuing two related questions: (i) does soil and topography conditioned rates lead to significant predictive and economic gains; (ii) do the gains depend on the number of yield observations available to rate-setters. To answer these questions, this study develops a method for econometrically incorporating this information into existing rating procedures used by the Risk Management Agency (RMA) and deploys an *ex-post* APH-based insurance simulation based on observed data and grounded in RMA guidelines (RMA 2018) to evaluate this new method and the significance of soil information in crop insurance rating. The empirical application leverages 149,267 farm-level observations of Kansas producers across four dryland crops (corn, soybean, sorghum, and wheat), spanning 46 years, and matched to fine-scale topographic/soil features derived using the nationwide gridded soil data from Soil Survey Geographic (SSURGO). The results suggest that features do improve the prediction accuracy of yield losses and can, in general, improve rating performance. Interestingly, these improvements are specific to farms with limited yield histories, as there are no improvements for farms with the commonly used yield history of ten years.

The second study in Chapter 3 focuses on the same crops and ascertains the tradeoffs between Production-History-based and Index-based insurance by pursuing two related objectives:

(i) ascertain the potential outcomes of a broad range of weather Index-based insurance and APH-based insurance under specified farm income goals; and given this, (ii) determine if the potential outcomes are different. The objectives are achieved by deploying an *ex-post* Index-based insurance simulation that is parallel to the APH-based insurance simulation in Chapter 2 to generate the two potential outcomes based on the same farm-level data and then assess their tradeoffs. The products in the *ex-post* Index-based insurance simulation are designed following RMA's rainfall index insurance for pasture rangeland and forage (PRF-RI). The results show that economically significant tradeoffs do exist between APH- and Index-based insurance and that different types of Index products are associated with differing levels of basis risk. Particularly, Index-based insurance that protects against excess accumulation in killing-degree-days (i.e., degree-days >30 °C) generates the most significant gains in farm income and economic rents and is associated with relatively low basis risk.

Unlike the first two that dealt with agricultural risk management from a crop insurance angle, the third study is also focused on risk management but from the futures market angle. Particularly, the final study in Chapter 4 answered a simple but important question of whether live-streamed harvest-time yields from precision technologies are potentially economically valuable. To answer this question, the study utilizes historic end-of-season farm-level corn yields that approximately represent 83% of US planted acres for 1999-2008 and Crop Progress and Condition (CPC) to construct weekly yield projection as representative of those from live yield monitors. The idea is to utilize the farm-level yield data to represent the population of farm-level US corn yields, and the weekly variation in CPC information on the proportion of annual crop harvested and under various conditions to approximate how the yield population changes throughout the harvest season. Given the simulated live-streamed harvest-time yields, the study then employs event study methodology to evaluate the potential economic significance of live-streamed yield monitor data vis-à-vis USDA reports. The results showed that corn futures market participants react to USDA reports and that live-streamed yield information does not elicit significant market reaction beyond that.

The research questions being posed in this dissertation have important implications for current and future agricultural policy. The analysis bridges data from multiple disciplines in innovative ways to leverage new insights, and the results will generate discussion among many types of stakeholders including producers, policymakers, and agribusinesses. The conclusion of this dissertation in Chapter 5 highlights some of the important implications for current and future agricultural policy.

References

Barnett, B.J., and O. Mahul. 2007. "Weather Index Insurance for Agriculture and Rural Areas in Lower-Income Countries." American Journal of Agricultural Economics 89(5):1241–1247. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8276.2007.01091.x.

Carter, M., A. de Janvry, E. Sadoulet, and A. Sarris. 2017. "Index Insurance for Developing Country Agriculture: A Reassessment." Annual Review of Resource Economics 9(1):421–438. Available at: http://www.annualreviews.org/doi/10.1146/annurev-resource-100516-053352.

Chambers, R.G. 1989. "Insurability and Moral Hazard in Agricultural Insurance Markets." American Journal of Agricultural Economics 71(3):604–616. Available at: https://onlinelibrary.wiley.com/doi/abs/10.2307/1242016.

Glauber, J.W. 2013. "The Growth Of The Federal Crop Insurance Program, 1990–2011." American Journal of Agricultural Economics 95(2):482–488. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1093/ajae/aas091.

Horowitz, J.K., and E. Lichtenberg. 1993. "Insurance, Moral Hazard, and Chemical Use in Agriculture." American Journal of Agricultural Economics 75(4):926–935. Available at: https://onlinelibrary.wiley.com/doi/abs/10.2307/1243980.

Jensen, N., and C. Barrett. 2017. "Agricultural Index Insurance for Development." Applied Economic Perspectives and Policy 39(2):199–219. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1093/aepp/ppw022.

Mahul, O., and C.J. Stutley. 2010. Government Support to Agricultural Insurance. The World Bank. Available at: http://elibrary.worldbank.org/doi/book/10.1596/978-0-8213-8217-2 [Accessed June 4, 2020].

Park, S., B.K. Goodwin, X. Zheng, and R.M. Rejesus. 2020. "Contract elements, growing conditions, and anomalous claims behaviour in U.S. crop insurance." The Geneva Papers on Risk and Insurance - Issues and Practice 45(1):157–183. Available at: http://link.springer.com/10.1057/s41288-019-00143-9.

Risk Management Agency [RMA]. 2018. "2019 Crop Insurance Handbook." Available at: https://legacy.rma.usda.gov/handbooks/18000/2019/19_18010.pdf [Accessed May 30, 2020].

Smith, V.H., and J.W. Glauber. 2012. "Agricultural Insurance in Developed Countries: Where Have We Been and Where Are We Going?" Applied Economic Perspectives and Policy 34(3):363–390. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1093/aepp/pps029 [Accessed June 18, 2019].

Smith, V.H., J.W. Glauber, and B.K. Goodwin. 2017. "Time to reform the US federal agricultural insurance program. Agricultural Policy." Agricultural Policy in Disarray Series Report. Washington, D.C.: American Enterprise Institute (AEI). Available at: http://www.aei.org/wp-content/uploads/2017/10/Time-to-Reform-the-US-Federal-Agricultural-Insurance-Program.pdf.

Yu, J., and N.P. Hendricks. 2020. "Input Use Decisions with Greater Information on Crop Conditions: Implications for Insurance Moral Hazard and the Environment." American Journal of Agricultural Economics 102(3):826–845. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1093/ajae/aaz035.

Chapter 2 - Utilizing Topographic and Soil Features to Improve Rating for Farm-level Insurance Products

2.1 Introduction

Previous research has established linkages between topographic and soil features with general agricultural production outcomes (Cox et al. 2003; Corwin et al. 2003; Juhos, Szabó and Ladányi 2016; Li et al. 2019); however specific links to large negative production shocks and/or agricultural insurance products are scarce. Initiated in Europe over two centuries ago, agricultural insurance is a large and rapidly expanding component of producer-oriented governmental support programs in both developed and developing countries (Mahul and Stutley 2010; Smith and Glauber 2012). Although a bit dated, a 2008 World Bank survey found that the global agricultural insurance market provided indemnity and index-based crop insurance products that generated $15 billion in premiums across 65 countries (Mahul and Stutley 2010).

A wide range of topographic and soil features have been linked to production, including soil texture/structure (Cox et al. 2003; Nyiraneza et al. 2012; Sene et al. 1985), available water (Campbell et al. 1993), pH (Anthony et al. 2012; Martín, Bollero and Bullock 2005), bulk density (Corwin et al. 2003), organic matter (Martín et al. 2005), nutrients (Cox et al. 2003; Di Virgilio, Monti and Venturi 2007), and slope (Kravchenko and Bullock 2000). However, these studies typically focus on the cross-sectional (spatial) effects on mean yield across locations and do not investigate implications for temporal (interannual) risk at a fixed location. Given the strong linkages to cross-sectional yield variation, one might also expect that topographic/soil features affect yield risk at a specific location as well. Furthermore, if this is the case then it suggests that components of agricultural insurance programs such as premium rates should also be affected by them. To the best of this study's knowledge, Woodard and Verteramo-Chiu (2017) is the only previous study linking a topographic/soil feature to crop insurance using observational data.

The study focused on the U.S. federal crop insurance program (FCIP) which began in the 1930s and is currently a public-private partnership where the Risk Management Agency (RMA) is mandated by the Federal Crop Insurance Corporation (FCIC) to oversee the FCIP and provides subsidized, multiple-peril individual and area-wide insurance policies covering both yield and revenue support for over 100 crops planted on a majority of U.S. cropland (RMA 2020). Major concerns for the FCIP include low participation, either caused by or in conjunction with adverse selection. The FCIP deals with these concerns by inducing pricing heterogeneity across farms based on their presumed risk in addition to subsidizing purchases. Cost savings from reducing subsidies is a perennial topic surrounding the program (United States Government Accountability Office [GAO] 2014; Congressional Research Service [CRS] 2015; Congressional Budget Office [CBO] 2017; Lusk 2017) but such a reduction would likely reduce producer participation (Congressional Budget Office [CBO] 2017). Importantly, subsidy reduction would place more importance on RMA's ability to price risk accurately across farms to guard against low participation and/or adverse selection.

There are many determinants of heterogeneous risk across farms, but few are easily observed from the rate-setter's perspective. One is to measure average yield/revenue on-farm and then presume that risk co-varies with that average, and indeed this approach is currently employed by the RMA (Coble et al. 2010). Another dimension that has received less attention is the incorporation of publicly available geo-referenced measures such as topographic and soil features as in Woodard and Verteramo-Chiu (2017), which found that conditioning yield histories on soil feature improved rating performance. The study extends the literature by pursuing two related questions: (i) does soil and topography conditioned rates lead to significant predictive and economic gains and, if so, (ii) whether these gains depend on the number of historical yield observations available to rate-setters. Intuition for the second question comes from the potential ability of repeated sampling to naturally “capture” the time-invariant linkage between location and risk.

The study also proposes a new method of incorporating soil information into crop insurance rates by recalibrating RMA’s rate multiplier curve, rather than adjusting historically reported yield series like Woodard and Verteramo-Chiu (2017). The method is applied to a sample of 149,267 observations across 5,428 farms and four dryland crops (corn, soybean, sorghum, and wheat) in Kansas (KS) spanning 46 years (1973-2018). The analysis initially focuses on soil texture features that were considered “optimal” based on machine learning algorithms, but the main results are later shown to hold across a wide range of topographical and soil features including root zone depth, available water storage, slope, exchangeable cations, soil organic carbon, and the National Commodity Crop Productivity Index (NCCPI). Measures for these variables are derived using the nationwide gridded soil data from Soil Survey Geographic (SSURGO).

The study finds that incorporating soil information does improve the predictive accuracy of losses and is associated with economically meaningful premium rate improvements for the overall sample of farms with varying yield history lengths.[1] Perhaps more interestingly, the economic gains decrease rapidly with yield-history-length as very large gains are associated with histories of less than four years and essentially zero gains associated with yield histories of ten years. This is the first documented evidence that efficiencies from incorporating soil information

[1] Revenue insurance dominates the FCIP, however it includes a yield risk component that is based on that of yield insurance.

into FCIP rate-setting procedures crucially depend on how much historical yield information is provided by producers.

2.2 Methods

2.2.1 Crop Insurance Continuous Rating Models

An extensive literature has attempted to estimate farm level rates (Carriquiry, Babcock and Hart 2008; Ramirez, Carpio and Rejesus 2011; Woodard and Verteramo-Chiu 2017), but typically rely on methods based on distributional assumptions of yields. To date, no study has utilized an identification strategy that relies on holding county base and fixed rates constant while adjusting the rate multiplier curve to better fit empirical LCRs from observed farm yield over a large temporal and spatial domain.

RMA sets base insurance rates for county/crop combinations derived from historical loss experiences adjusted for extreme losses (Coble et al. 2010) and then averages them using weather-weights as in Rejesus et al. (2015). In a sequential yet separate step, these county-level base rates are adjusted to the individual/unit level based on a presumption of risk relative to others in the same county (Coble et al. 2010). The relative risk adjustment is determined by a rate multiplier curve, which essentially embodies an assumption that risk correlates negatively with mean yield such that relatively productive insureds are lower risk and thus receive lower rates. Section 508 of the Agricultural Adjustment Act of 1938 mandates RMA to modify rating systems to be actuarially sound. Consequently, there is a large and growing literature analyzing RMA insurance rating procedures along the lines of actuarial soundness (Woodard, Sherrick and Schnitkey 2011), adverse selection (Skees and Reed 1986; Goodwin 1994), technology-induced yield trends (Adhikari, Knight and Belasco 2012; Seo et al. 2017), and heteroscedastic yields (Harri et al. 2011; Annan et al. 2014).

The main components of FCIP farm-level yield insurance contracts are the rate yield ($\bar{y}_i$), approved yield ($\ddot{y}_i$), yield guarantee ($\tilde{y}_{ig}$), coverage level (C_g), indemnity (I_{ig}), premium rate (R_{ig}), premium (P_{ig}), and subsidy (S_g). Here i denotes farm and g denotes coverage level. The rate yield is the simple average of actual production history (APH) reported by farmers subject to no adjustments. While the approved yield could in principle be the same as the rate yield, several aspects of the RMA's actuarial process can produce differences as the production history is typically adjusted higher through various mechanisms.[2] The coverage level is selected by the purchaser and is the proportion of the insured unit's approved yield used to set the yield guarantee such that $\tilde{y}_{ig} = \ddot{y}_i \cdot C_g$.[3] Assuming output price is equal to unity without loss of generality, the per-acre indemnity for a given yield outcome, y_{it}, is given by $I_{ig} = max\{0, \tilde{y}_{ig} - y_{it}\}$.[4]

Insurance policies are supposed to be priced actuarially fairly such that premiums are equal to expected indemnities: $P_{ig} = E[I_{ig}]$. Since I_{ig} are stochastic and not known at the time the policy is written, RMA sets the price as the product of a premium rate, (R_{ig}), determined using a continuous rating formula, and the yield guarantee: $P_{ig} = R_{ig}\, \tilde{y}_{ig}$.[5] The final price paid by the insured is $P_{ig}S_g$, where S_g is a subsidy factor determined by FCIC and is tied to coverage level.[6]

[2] Common adjustments made to approved yield calculations include yield exclusion, yield substitution, and trend.

[3] Federally approved coverage levels for the 2019 crop year ranged from 55-85% in 5% increments.

[4] Note that to the extent that approved yield is higher than rate yield, as is often the case, this benefits producers as the yield guarantee will be higher, and thereby, will increase indemnities for a given yield outcome and improves producer welfare (Adhikari, Knight and Belasco 2013)

[5] In practice, premiums are the product of the premium rate and liability = guarantee × price. However, in the current setup, price = 1, so liability = guarantee.

[6] For the 2019 crop insurance program, corn, soybeans, sorghum, and wheat policies with coverage levels of 0.55, 0.65, 0.75, and 0.85 had S_g respectively equal to 0.64, 0.59, 0.55, and 0.38. Between 2005-2018, the federal government subsidized on average 61.1% of farmers' premiums (RMA 2019b).

Based on RMA (2000), the specific formula that RMA uses to construct premium rates is given by[7]:

$$R_{ig} = \alpha_{cg}[\bar{y}_i/\bar{y}_{cr}]^{\beta_c} + \delta_{cg}, \qquad (2.1).$$

Here the subscript c denotes county, and α_{cg} and δ_{cg} are a base rate and fixed loading factor, both of which vary across coverage levels and are calculated from county-level aggregated loss experience data. The county base rate is scaled up or down for a particular farm based on a rate multiplier $[\bar{y}_i/\bar{y}_{cr}]^{\beta_c}$ that leverages the ratio of the producer's rate yield over the county-level reference yield $\bar{y}_{cr}$ to make this adjustment. RMA defines this reference as an average of county-level yields. For a given ratio, the base rate will be adjusted based on the value of the county-specific rating exponent β_c. Even though Equation (2.1) is a simplified version of RMA's continuous rating formula, it captures all the essential elements for this study.

The adjustment of the county base rate for an individual farm depends on two main pieces of information: (i) the farm's relative yield performance to that of its peers and (ii) the value of the rating exponent. For an average farm with a yield ratio of one, the implied rate multiplier will also take on a value of one regardless of the value of the rating exponent, and the premium rate will be $R_{ig} = \alpha_{cg} + \delta_{cg}$. In practice, farms are either going to be above or below the reference yields, and the county base rate will be adjusted accordingly based on the sign of β_c. If β_c is positive, then the rate multiplier is monotonically increasing in the yield ratio, and relatively more productive farms are considered riskier. Thus, the county base rate is adjusted upward. However, if instead, β_c is negative, then the rate multiplier is monotonically decreasing, and the opposite effect occurs with

[7] The literature is inconsistent on the difference between rate and approved yield in the FCIP. RMA actuarial documents distinguish between two yields calculated from the farmer's APH: approved yields are used in calculating farmer's guarantee and rate yields are used to calculate premium rates (RMA 2019a).

relatively more productive farms assumed to be less risky. Thus, the county base rate is adjusted downward. According to Milliman and Robertson (2000), the use of a negative β_c by RMA is based on research and is corroborated by (Botts and Boles 1958; Skees and Reed 1986).

The RMA's rating methodology approximates expected losses with rates (Coble et al. 2010). This excludes the cost associated with program delivery since they are provided for in the Administrative and Operating cost (A&O) agreements. The RMA derives expected losses – referred to as the "loss cost ratio" (LCR) – as expected indemnity divided by liability. Consequently, since LCRs measure loss per unit of exposure, an objective of RMA's method is to derive rates that reflect this. So, what can go wrong with the insurance continuous rating formula presented in Equation (2.1)? Woodard and Verteramo-Chiu (2017) postulated that biased rate-yields ($\bar{y}_i$) could lead to biased rates (R_{ig}) and showed that conditioning expected yields on soil could mitigate this.

The study deviates from Woodard and Verteramo-Chiu (2017), by not adjusting rate-yields, but rather allowing the key county-level parameter in the rate multiplier, β_c, to be re-estimated to account for topographic and soil features. Based on RMA's approximation of the expected loss component of rates with LCR, Equation (2.1) can empirically be estimated as

$$LCR_{igt} = \alpha_{cg}^{*}[\bar{y}_{it}/\bar{y}_{crt}]^{\beta_c^{*}+f(\boldsymbol{X}_i,CDR_d;\boldsymbol{\rho})} + \varepsilon_{igt} \qquad (2.2),$$

where the variable LCR_{igt} is an empirical LCR for farm i in year t under coverage level g. Details of its derivation are outlined in the data section. Note that the county level base rate and the exponent in the rate multiplier are being held fixed in this empirical model. These fixed parameters are denoted with an asterisk (*) and take on values based on their 2019 crop year values published in RMA (2019a). The loading factor is omitted from the equation by setting it equal to zero.[8]

The exponent for the rate multiplier curve has been amended to include the fixed value currently used by RMA plus an adjustment: $\beta_c^* + f(.)$, where $f(.)$ is a function of farm-level topographic and soil features ($\boldsymbol{X}_i$) and crop reporting district-level fixed effects (CDR_d). The main idea is to estimate the $\boldsymbol{\rho}$ parameters while holding both α_{cg}^* and β_c^* fixed, which can then be used to re-estimate current continuous rating exponents to be reflective of empirical LCRs.

Three types of adjustment functions are considered. The first includes only CRD level adjustments that are not based on topographic and soil features but rather ad hoc geographical boundaries: $f(.) = \sum_d^D \rho_d CRD_d$ [i.e., CRD model] where CRD_d is a dummy for crop reporting district d. The second ignores these and instead focuses on topographic and soil features: $f(.) = h(\boldsymbol{X}_i; \boldsymbol{\rho})$ [SOIL model]. The third considers both types simultaneously: $f(.) = h(\boldsymbol{X}_i; \boldsymbol{\rho}) + \sum_d^D \rho_d CRD_d)$ [CRD-SOIL model]. For comparisons to current rates, a baseline model where the adjustment function $f(.)$ is omitted entirely from the model is also included so that the four models under consideration are:

[Baseline] $$LCR_{igt} = \alpha_{cg}^*[\bar{y}_{it}/\bar{y}_{crt}]^{\beta_c^*} \qquad (2.3)$$

[CRD] $$LCR_{igt} = \alpha_{cg}^*[\bar{y}_{it}/\bar{y}_{crt}]^{\beta_c^* + \sum_d^D \rho_d CRD_d} + \varepsilon_{igt} \qquad (2.4)$$

[SOIL] $$LCR_{igt} = \alpha_{cg}^*[\bar{y}_{it}/\bar{y}_{crt}]^{\beta_c^* + h(\boldsymbol{X}_i; \boldsymbol{\rho})} + \varepsilon_{igt} \qquad (2.5)$$

---

[8] Alternatively, $LCR_{igt} - \delta_c^*$ could have been used as the dependent variable in Equation (2.2). However, fixing δ at its 2019 crop year value will alter the rate multiplier curve by making it steeper (i.e., a large absolute value for β_c). To demonstrate, suppose the true rate multiplier curve has a continuous rating exponent of -1.5 and the insureds yield ratio is 1.5 with a corresponding $LCR = 1.5^{-1.5} = 0.54$. Subtracting a fixed county rate of 0.02 from the LCR gives an effective continuous rating exponent of $\beta_c = \ln[0.54 - 0.02]/\ln[1.5] = -1.6$.

[CRD-SOIL] $LCR_{igt} = \alpha^*_{cg}[\bar{y}_{it}/\bar{y}_{crt}]^{\beta^*_c + h(\boldsymbol{X}_i;\boldsymbol{\rho}) + \sum_d^D \rho_d CRD_d)} + \varepsilon_{igt}$ (2.6).

Equations 2.4-2.6 are estimated separately for each crop using nonlinear least squares with farm-level data pooled across all counties. Given the estimates of the parameters in $\boldsymbol{\rho}$, the adjusted exponent is estimated for the three alternative models as $\beta^*_c + \sum_d^D \hat{\rho}_d CRD_d$, $\beta^*_c + h(\boldsymbol{X}_i; \hat{\boldsymbol{\rho}})$, and $\beta^*_c + h(\boldsymbol{X}_i; \hat{\boldsymbol{\rho}}) + \sum_d^D \hat{\rho}_d CRD_d)$, respectively. Note that exponents will be adjusted at the county/CRD level under the CRD model, and at the farm level for the SOIL and CRD-SOIL models.

2.2.2 Measuring Predictive Accuracy

Both in- and out-of-sample methods are used to measure the predictive accuracy of the models. In-sample accuracy is measured as the mean squared error, while out-of-sample accuracy is based on cross-validation using ten approximately equal-sized subsamples (folds). Both measures are reported relative to the baseline model (Equation [2.3]), with values below one indicating better performance. The whole process is repeated 1,000 times by bootstrap sampling the farms in the dataset to measure statistical uncertainty.

2.2.3 Measuring Economic Performance

Based on Harri et al. (2011) and Coble et al. (2007), the study assumes the role of an Agricultural Insurance Provider (AIP) to ascertain whether the models generate premium rate adjustments that are economically different. If the rate prediction from the adjustment model (Equations 2.4-2.6) is lower than the baseline (Equation 2.3), then the contract is assumed to be overpriced and thus is placed in the retain pool. However, if the rate prediction is instead higher, then the contract is assumed to be underpriced and placed in the ceded pool. By separating all policies into these two pools, one can compare the indemnities that occur based on the observed yield outcomes across pools to quantify economic differences from adopting the adjustment model.[9] A complete set of results for comparing loss ratios (indemnities over premiums) across pools is also provided in the robustness checks section below.

The cede/retain game is operationalized by utilizing an out-of-sample rating simulation approach with sixteen annual iterations from 2003-2018. For each iteration, a training sample of all prior years' data is used to estimate the models and predict premium rates for all farms in that year. For example, data before 2003 would be used to predict rates for all farms in 2003, data before 2004 would be used to predict rates for 2004, and so on. For a given iteration, rates are compared to the baseline rate and farms are separated into cede and retain pools. Indemnities are then calculated based on observed yield outcomes for the farm, summed across all years, and then divided by the total policies in each pool. The aggregate values are then used to form ceded to retained indemnity ratios with values greater than one indicating the economic significance of the predicted rates. The same bootstrap as above is used to measure statistical uncertainty.

2.3 Data

The loss experience from RMA Statplan along with accompanying Common Land Unit (CLU) data would be ideal for the current study.[10] However, RMA loss data are not publicly available. Thus, data from secondary sources are used to replicate a mini version of the loss experiences in Kansas. The four main sources of data are: (1) 46 years of farm-level Kansas corn, sorghum, soybean, and wheat yields provided by the Kansas Farm Management Association (KFMA); (2) actuarial information from RMA's 2019 Actuarial Data Master (ADM) (RMA

---

[9] The study utilized the cede/retain game because, given the limited geographical coverage of the sample, it will be erroneous to make efficient rating by the RMA the focus of the paper. Additionally, the public-private partnership associated with crop insurance delivery makes the cede/retain game an attractive robust but simple metric.

[10] The Statplan is the standardized database of all policies written by the FCIC since 1948 and is used to support sound actuarial decisions.

2019a); (3) gridded topographic and soil features from Soil Survey Geographic (SSURGO) provided by the USDA-NRCS (Soil Survey Staff 2020); and (4) gridded crop frequency layer from NASS CropScape (USDA National Agricultural Statistics Service 2019).[11]

2.3.1 Loss Experience Data

The study focuses entirely on dryland production of corn, sorghum, soybeans, and wheat in Kansas. Observations were dropped from the sample in the following order: (1) if the farm cannot be geocoded based on mailing address; (2) if the reported yield was 1.5 times the largest recorded contest yield[12]: (3) if RMA does not report insurance parameters needed to calculate county base rates and loading factors in that county: and (4) any crop/county combination with less than 4-years of data from 1999-2002[13]. All notes, tables, figures, and references in appendix A have a leading A. Figure A.1 shows the spatial representation of the 5,428 sample farms by crop. The 149,267 yield observations spanning 1973-2018 exhibit a great amount of cross-sectional and temporal variation (Figure A.2, Panel A). The mean [standard deviation] of the yields in kg/ha are 4,833 [2,137], 3,412 [1,484], 1,928 [905], and 2,470 [899] for corn, sorghum, soybeans, and wheat, respectively. Figure A.3 shows the representativeness of the KFMA data by comparing sample average yields at the crop-county-year level to yield statistics from NASS.

---

[11] All Data and models were processed on Beocat, a High-Performance Computing (HPC) cluster at Kansas State University (https://beocat.ksu.edu/)

[12] Yield Contests are annual competitions held at the state and national levels for major grains partly with the goal of; (1) recognizing and celebrating the success of high-yielding farmers; (2) promoting farming operations and best management practices to improve and sustain yields; and (3) sharing data to benchmark production and provide information to increase profitability. In this study, yields well above the highest ever recorded contest yields are deemed unrealistic, thus they are dropped from the analysis.

[13] Crop/county combinations with less than 4-years of data from 1999-2002 are drop from the analysis to ensure that all such combinations have the minimum required annual data points used in exponent estimation for the sixteen annual iterations of economic performance from 2003-2018.

A three-step algorithm is used to replicate loss experience data for each crop-year in the dataset, starting with 1983. The first step uses the yield data from ten successive years to estimate county-level transitional yields (T-yields) and county reference yields ($\bar{y}_{crt}$). For the reference yield, annual county yields were first estimated as the average yield across all the farms in that county. The reference yield was then estimated as the mean of the annual county yields over the ten previous years (Rejesus et al. 2010). Given the county reference yields, the T-yields for each county was calibrated as $\bar{y}_{crt}\vartheta_c$, where ϑ_c is the reference yield to T-yield ratio for county c, calculated from 2019 RMA values.[14]

The second step of the loss experience algorithm estimates rate yield ($\bar{y}_{it}$) and approved yield ($\ddot{y}_{it}$) for each farm/year/crop with their observed yields (y_{it}) in ten successive years serving as the basis for an APH database. In practice, rate and approved yield calculations are complex; however, the study sheds some of the complexities such as yield exclusion, yield substitution, and trend adjustment to maintain the tractability of the analysis.

Based on RMA guidelines (RMA 2018), rate yield is taken as the mean of the actual yields in the APH database; however, if there are no actual yields, the rate yield is taken as the T-yield. For approved yield, the APH database for an insured must have at least four successive yield data points (actual or assigned). If the APH database has actual yields for at least the last four successive years, the approved yield is just the simple average of the actual yields. Where the insured unit has less than four successive years of records, variable T-yields are used as replacements for the years

[14] The proportional calibration of T-yield based on reference yield was used because the calculation of T-yield is ambiguous in both the RMA and academic literature. Section 502(b) of the Agricultural Adjustment Act of 1938 defines T-yield as the maximum average production per acre or equivalent measure that is assigned to acreage for a crop year. Thus, T-yields are alternatively estimated as the upper value of the 95% confidence interval across all farms in the ten successive years, but the results largely remained the same.

with no records to meet the four-year minimum yield requirement. Particularly, missing yields for insureds with a record of zero, one, two, and three year(s) are taken as 65, 80, 90, and 100% of their counties T-yield, respectively. Finally, the approved yield for an insured is bound between 90% of its value in the previous year (i.e., yield cap) and 70–80% of the relevant T-yield (i.e., yield floor). The floor is set at 70, 75, and 80% if the four-year minimum yield requirement is short by three, two, and one yield(s).

Figure A.2, Panels B-E, shows the annual box plots of rate yields, approved yields, T-yields, and reference yields. As expected, all four types of yields are trending up with the rate yields exhibiting a larger amount of cross-sectional variability. The resulting relative yields ($\bar{y}_{it}/\bar{y}_{crt}$) are also presented in Figure A.4, and unlike the yields, they are stable over the study period as expected. The sample distribution of the loss experience by the number of actual yields used in their rate yield calculation shows that the majority meet the four-year minimum yield requirement (Figure A.5).

Empirical LCRs based on observed yields y_{it} for each farm/year/crop are calculated as follows. The approved yield ($\ddot{y}_{it}$) from above is used to construct the guaranteed yield by coverage level: $\tilde{y}_{igt} = \ddot{y}_{it} \cdot C_g$, which in turn, is used to measure actual indemnities given by $I_{igt} = max\{0, \tilde{y}_{igt} - y_{it}\}$. The ratio of these indemnities to the guaranteed yield then defines the empirical LCR: $LCR_{igt} = I_{igt}/\tilde{y}_{igt}$. C_g is assumed to be equal to 75%, the largest enrolled coverage in terms of acres of corn, soybeans, sorghum, and wheat in KS for 2002-2019 (RMA 2020). The annual averages of the LCRs are shown in Figure A.6. Overall averages [standard deviation] of the LCRs are 0.076 [0.192], 0.073 [0.185], 0.079 [0.184], and 0.061 [0.171] for corn, sorghum, soybeans, and wheat, respectively.

Preliminary descriptive analysis of the farm level data used in this study in Table A.1 shows that there is a near one for one inverse relationship between mean relative yield and risk. Thus, the RMA's actuarial methodology assumption is embodied in the raw data. Similar relationships also hold for the case of NASS county-level data (Table A.2). Table A.1 hints that the rating exponent is expected to remain negative even after they are adjusted. Consequently, the negativity restriction of the rating exponent is not directly imposed during estimation but rather evaluated ex-post.

2.3.2 Actuarial Data

The KFMA data is farm-level instead of field-level, thus, it is a closer approximation to an enterprise unit (EU), thus, actuarial information for EU dryland production of corn, sorghum, soybeans, and wheat for a coverage level of 75% retrieved from the RMA's 2019 ADM is used for the analysis. The specific actuarial parameters retrieved from the ADM are (1) county/crop continuous rating exponents; (2) county/crop reference rates (R_{cr}); (3) county/crop fixed rates (R_{cf}); (4) unit residual factors for production unit adjustments (R_p); and (5) rate differential factors for coverage level adjustments (R_g). Based on the ADM parameters, the county-level base rate (α_{cg}) and fixed loading factors (δ_{cg}) are calculated as: $\alpha_{cg} = R_p R_g R_{cr}$ and $\delta_{cg} = R_p R_g R_{cf}$. The boxplots of the retrieved and calculated actuarial parameters are shown in Figure A.7.

2.3.3 Topographic and Soil Features

The exact location of each farmer's field was unknown; however, they were best approximated using their mailing address following the procedure outlined in Note S1. This approach is not ideal with measurement error potentially leading to attenuation bias. However, the approach is somewhat representative of the real-world situation in which RMA historically did not know the exact location of insured fields nor more generally what field the farmer will eventually plant on at enrollment. Additionally, for any given year, the history on which APH is based could have come from other fields. Furthermore, the results demonstrate substantial improvements in farmers with low APH history length and no improvements for farms with a high APH history, which suggests that there is a signal in the soil measure used and that the measurement error is likely minimal.

The gSSURGO has over 500 topographic and soil features that the rating exponent could be conditioned on. A viable candidate feature must have significant within-county variation, so all features with a zero standard deviation or a coefficient of variation less than 0.01 were dropped. Next various statistical learning techniques were used to further narrow down the list of features by focusing on those with relatively high LCR predictability. Details of the specific algorithms used by these techniques are in James et al. (2013) and results are shown in Figure A.8. Based on the selection process, soil texture is chosen as the preferred feature.

The basic elements of soil texture are (1) Sand - mineral soil particles that have diameters ranging from 2 to 0.02 mm; (2) Silt - mineral soil particles that range in diameter from 0.02 to 0.002 mm; and (3) Clay - soil particles that have diameters less than 0.002 mm. The spatial distribution of these is depicted in Figure A.9 and the farm-level distribution in Figure A.10. In general, sample farms are located on soils that are abundant in soil particles classified as silt, which is the second-most occurring followed by sand. In addition to soil texture, root zone depth, available water storage, slope, exchangeable cations, soil organic carbon, and NCCPI are also considered as robustness checks.

2.4 Results

2.4.1 Continuous Rating Exponent

Equation 2.3 serves as the baseline model and the county level RMA exponents β_c^* are reported in green in Figure 2.1 Panel A. The alternative models, Equations (4)-(6), are estimated using nonlinear least squares, and the parameter estimates are reported in Table 2.1. Those parameter estimates are used to form adjusted rating exponents for each of the three alternative models, reported alongside the baseline exponents in Figure 2.1 Panel A. In general, the rating exponents remain negative after adjustment and are lower in magnitude relative to RMAs. This suggests that adjustments are inducing a flatter rate multiplier curve thereby leading to smaller adjustments above and below the reference yield and more homogeneous rates across farms within the county. The latter is a particularly interesting aspect of the results as one might expect that models that include additional farm-level information would naturally lead to more heterogeneous rates across farms. However, as it will be shown below the soil information and yield-history length can be thought of as substitutes in that information only improve rating when sample length is small. Although not empirically verified in this study, likely, the soil adjusted exponent is likely counteracting (reducing) the effects of a noisy mean yield estimate.

In Figure 2.1 Panel A, it can be observed that the cross-county variation in exponents from the CRD adjusted models (i.e., models CRD and CRD-SOIL) is relatively higher than RMA's 2019 values. On the contrary, the cross-county variation from the SOIL model is comparable to the RMA 2019 values. Figures 2.1 Panel B and Figure 2.2 show within farm and county variation of the continuous rating exponents, respectively.

Table 2.1 Regression Results

	(1) CRD Model	(2) SOIL Model	(3) CRD-SOIL Model
Soil texture			
Silt	-	0.878** (0.488)	0.030 (0.854)
Clay	-	-0.798 (0.621)	1.216 (1.367)
Sand	-	2.023** (1.006)	-0.508 (0.927)
CRD			
NW	0.768*** (0.086)	-	0.563** (0.243)
SW	-	-	-
SC	1.845*** (0.331)	-	1.647*** (0.357)
NE	-	-	-
SE	0.285** (0.105)	-	-0.072 (0.143)
NC	0.531*** (0.151)	-	0.248 (0.197)
Dryland Sorghum			
Soil texture			
Silt	-	0.382 (0.265)	0.206 (0.510)
Clay	-	0.415 (0.575)	0.565 (1.116)
Sand	-	0.149 (0.176)	-0.192 (0.205)
CRD			
NW	0.384*** (0.106)	-	0.163 (0.157)
SW	0.397*** (0.088)	-	0.154 (0.101)
SC	0.315*** (0.062)	-	0.214 (0.113)
NE	0.505*** (0.154)	-	0.224 (0.180)
SE	0.249* (0.117)	-	-0.007 (0.178)
NC	-	-	-
Dryland Soybeans			
Soil texture			
Silt	-	0.572 (0.501)	0.402 (0.694)
Clay	-	0.637 (0.848)	0.764 (1.170)
Sand	-	0.505** (0.253)	0.068 (0.340)
CRD			
NW	-	-	-
SW	-	-	-
SC	0.649*** (0.093)	-	0.275** (0.135)
NE	-	-	-
SE	0.533*** (0.079)	-	0.055 (0.118)
NC	0.623*** (0.127)	-	0.163 (0.177)
Dryland Wheat			
Soil texture			
Silt	-	0.681* (0.350)	0.826** (0.397)
Clay	-	0.659 (0.571)	0.037 (0.888)
Sand	-	0.784*** (0.239)	0.580* (0.362)
CRD			
NW	-	-	-
SW	-	-	-
SC	0.736*** (0.066)	-	0.199 (0.148)
NE	0.544*** (0.152)	-	0.017 (0.212)
SE	0.735*** (0.057)	-	0.214 (0.149)
NC	-	-	-

Significance levels: * p<0.10, ** p<0.05, ***p<0.01
Notes: The table shows the nonlinear least-squares regression results for the adjustment parameters in equations 4-6. A pooled model that included pooled data from all counties was estimated. RMA's rating parameters are county-

specific and are included in each model as fixed parameters. Three different regression models were considered, each based on a separate type of adjustment. The first includes only CRD level adjustments (CRD model) which are based on dummy variables at the CRD level, and the parameter estimates are reported in column 1. The second focuses on topographic and/or soil features (SOIL model) measured at the farm level and the parameter estimates are reported in column 2. The third includes both types of adjustments simultaneously (CRD-SOIL model). Standard errors are calculated by bootstrap sampling the farms in the dataset.

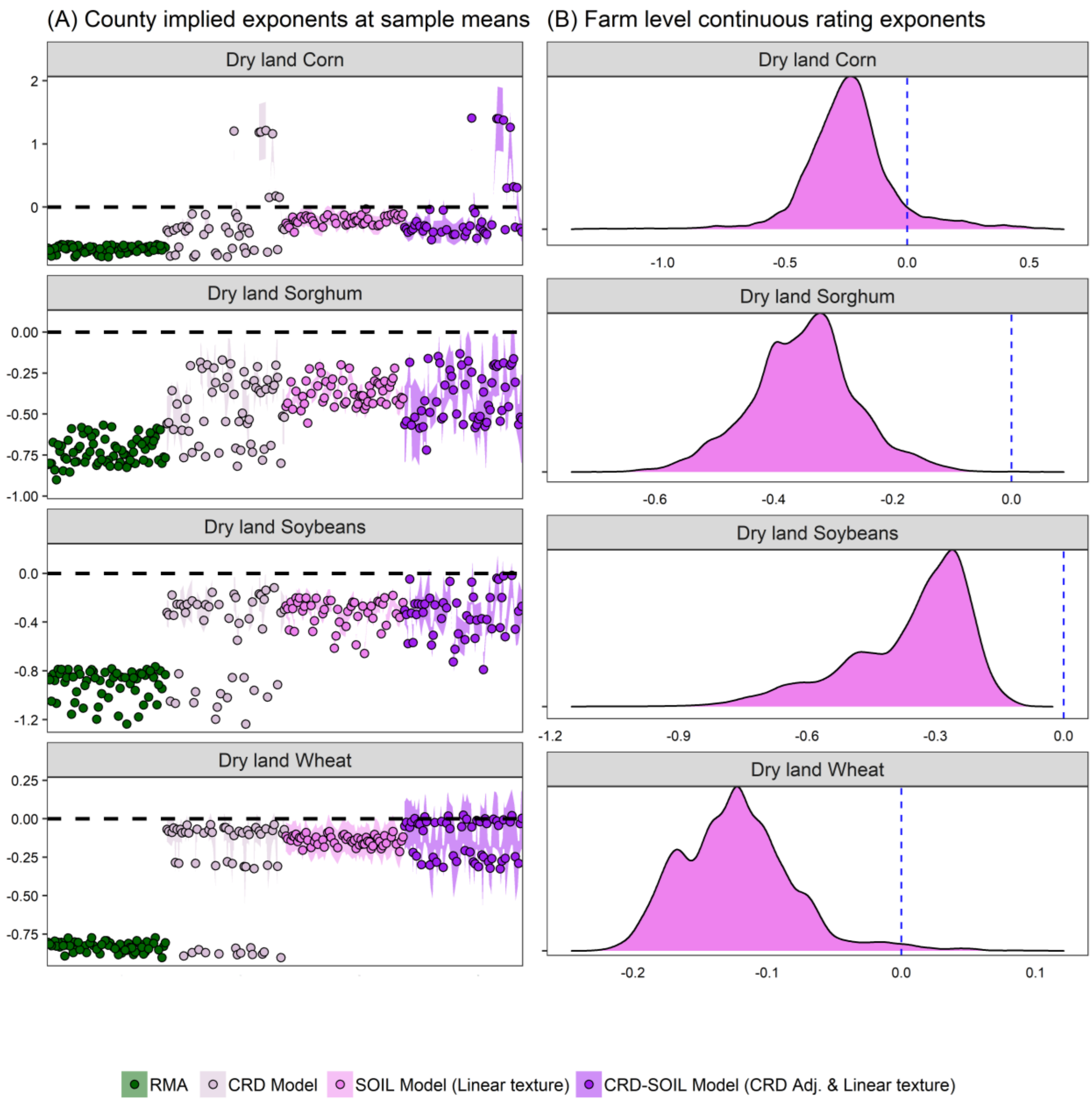


Figure 2.1: Plots of continuous rating exponents from alternative models in Kansas

Notes: Each panel of graph A shows the plots (circles) of county-level crop insurance continuous rating exponents provided by USDA Risk Management Agency (RMA) for 2019 (green) and from the models (CRD Model – Crop reporting district [CRD] conditioned exponents, SOIL Model – Linear soil texture conditioned exponents, and CRD-SOIL Model – CRD and Linear soil texture conditioned exponents). For SOIL and CRD-SOIL, the soil texture elements (clay, silt, and sand) for each county were taken as the mean of all the farms in that county. The shaded region represents the bootstrap (1,000) 95% confidence interval (CI) for the estimates, and the dashed black line marks the reference for zero. Estimates with their 95% CI overlapping with the zero lines are not statistically significant. Farm-level data was provided by the Kansas Farm Management Association (KFMA).

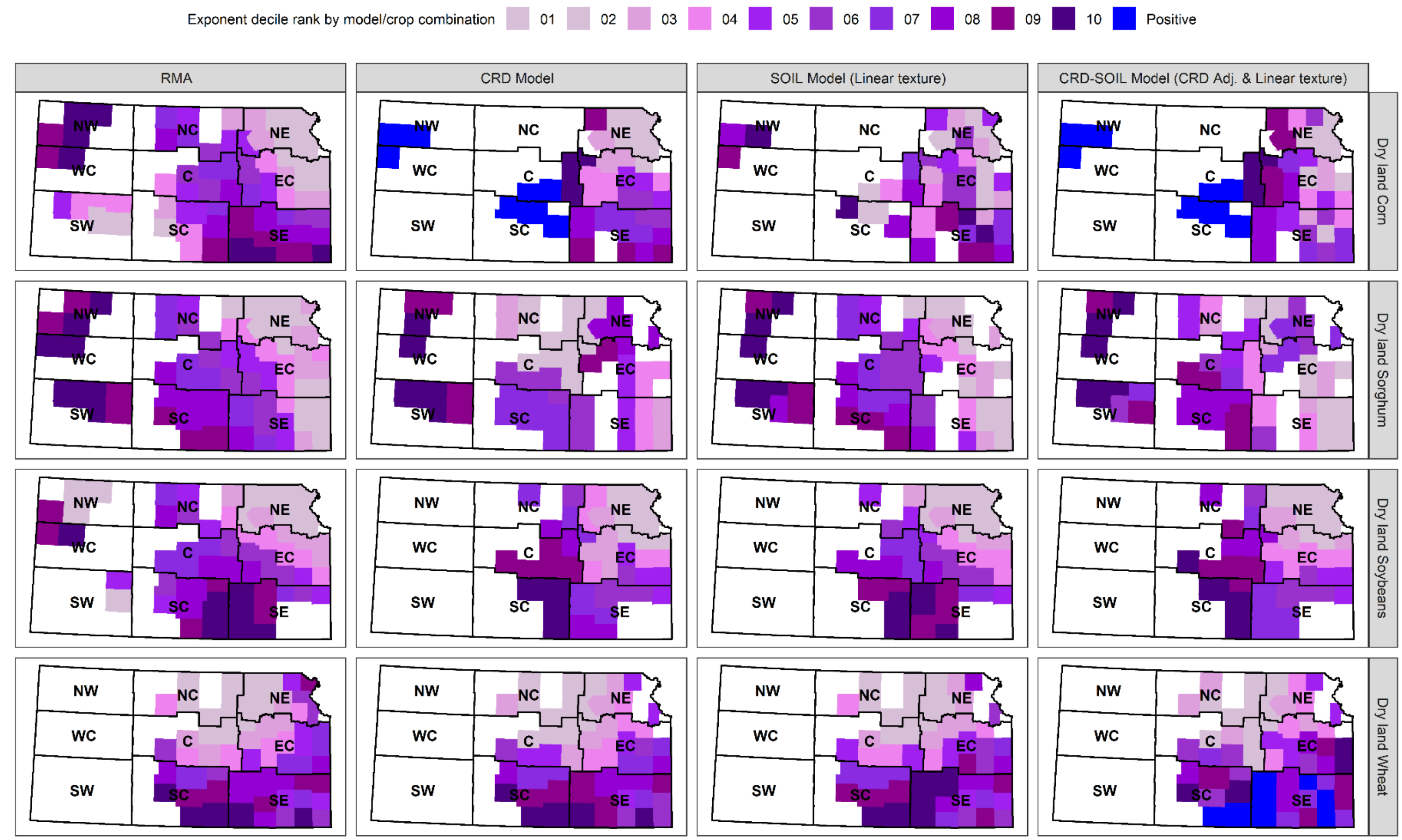


Figure 2.2: Spatial pattern in county-level continuous rating exponents for alternative models in Kansas

Notes: Each reference map shows the spatial pattern of the decile rank of crop insurance continuous rating exponents by model/crop combination. Counties with positive exponents are excluded from the ranking and displayed as blue. The model designation represents exponents provided by USDA Risk Management Agency (RMA) for 2019 and those from the models (CRD Model – Crop reporting district [CRD] conditioned exponents, SOIL Model – Linear soil texture conditioned exponents, and CRD-SOIL Model – CRD and Linear soil texture conditioned exponents). For SOIL and CRD-SOIL, the soil texture elements (clay, silt, and sand) for each county were taken as the mean of all the farms in that county. Farm-level data was provided by the Kansas Farm Management Association (KFMA).

2.4.2 Predictive Performance

Results for in- and out-of-sample prediction in Figures 2.3 Panel A and B indicate that the adjustment models reduce LCR prediction error. It is interesting to note that errors are the lowest for the models that utilize soil texture to adjust rates. Particularly, Figure 2.3 Panel A shows that out-of-sample errors from the models that include soil texture (SOIL and CRD-SOIL) are about 3% lower than that of the status quo. This suggests that there is a signal in the soil measure used and that the measurement error could be minimal.

Figure 2.4 presents the mean relative rates (to that of the RMA) from 2003-2018 for each observation using the adjustment from the SOIL model and paints a vivid picture of the resulting flatter rate multiplier curve as rates for yield ratios below one is adjusted lower while rates for ratios above one is adjusted higher. Overall rate adjustments are upward on average suggesting higher out-of-pocket insurance costs for producers, but there is substantial variation across farms from about -19% to 13%.

To provide a measure of program level differences between these rates, RMA projected price for 2019 for valuation is utilized. For the case of the SOIL model, the mean total premium, subsidy, producer paid premium, and AIPs A&O were $57.13 M, $43.99 M, $13.14 M, and $12.51 M, respectively. For the RMA 2019 exponents, similar values were $56.05 M, $43.16 M, $12.89 M, and $12.28 M, respectively. Overall, the values when soil information is included in rate-setting were only about 2% higher than those from the RMA.

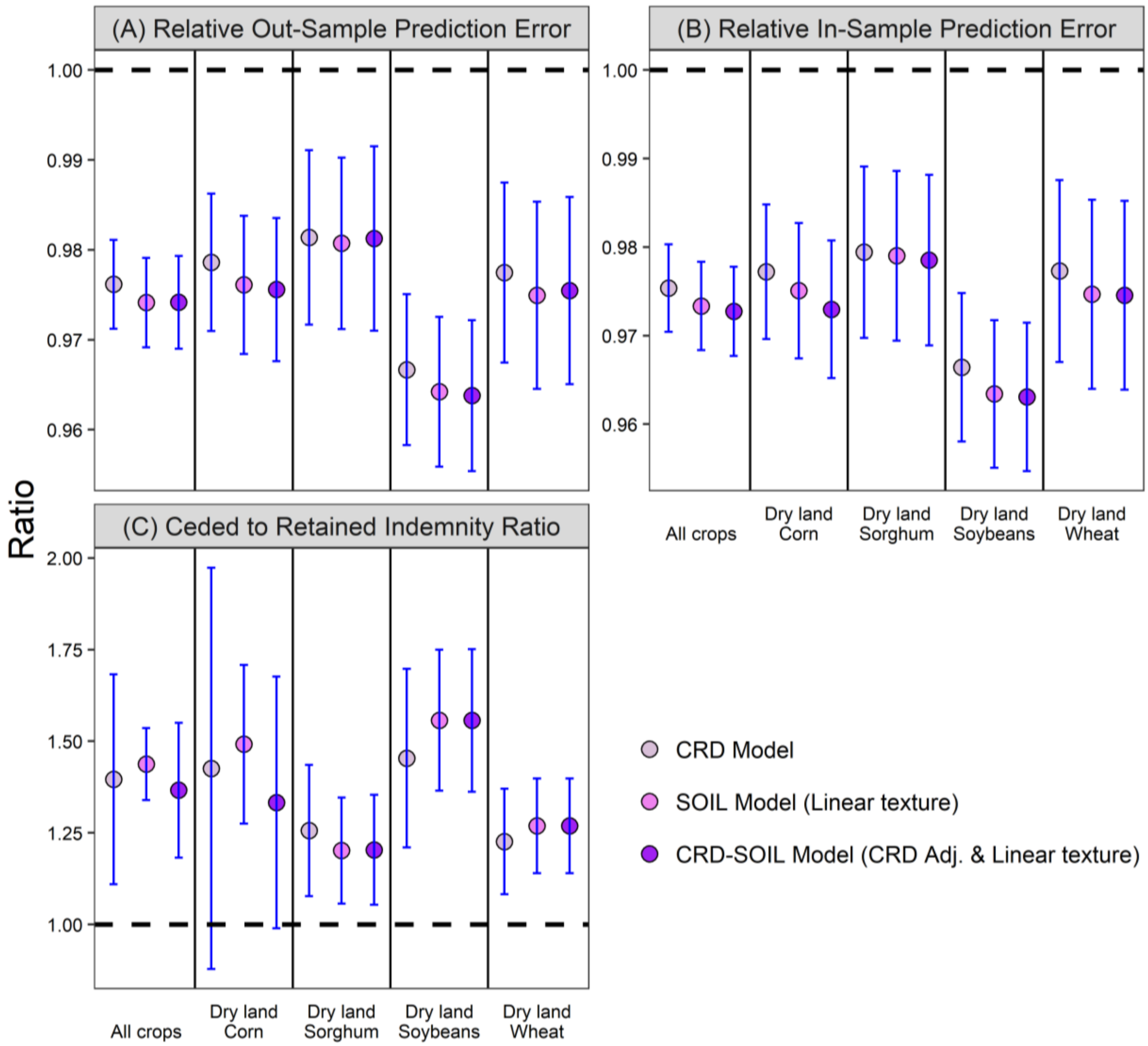


Figure 2.3: Predictive and economic performance of alternative models for estimating continuous rating exponents in Kansas

Notes: Graph shows the predictive (A and B), and economic (C) performance of crop insurance continuous rating exponents estimated from the models (CRD Model – Crop reporting district [CRD] conditioned exponents, SOIL Model – Linear soil texture conditioned exponents, and CRD-Soil Model – CRD and Linear soil texture conditioned exponents). Panels A and B are evaluated in terms of relative performance to exponents provided by USDA Risk Management Agency (RMA) for 2019. Panel C is based on Coble et al. (2007) and Harri et al. (2011) and measures the level of forgone economic rents as the ratio of indemnities from ceded to that of retained policies under a simplified Standard Reinsurance Agreement (SRA) scenario. For Panels A and B values less than one indicate how well the exponents perform better than that of the RMA, and for panel C, values greater than one indicate a relatively higher level of forgone economic rents. The error bars represent the bootstrap (1,000) 95% confidence interval (CI) for the estimates, and the dashed black line marks the reference for one. Estimates with their 95% CI overlapping with the reference line are not statistically significant. Farm-level data was provided by the Kansas Farm Management Association (KFMA).

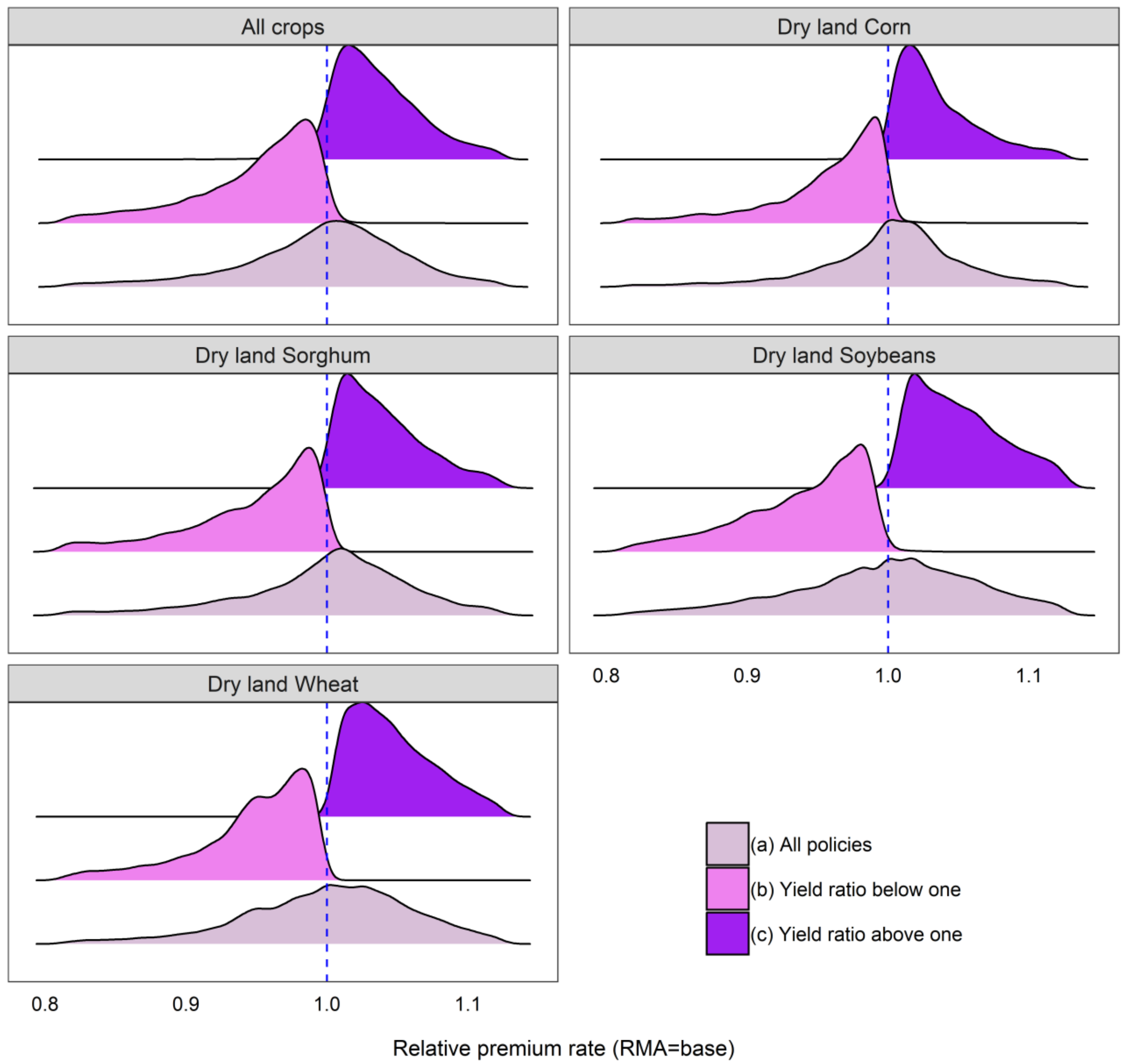


Figure 2.4: Distribution of farm-level relative premium rates

Notes: Graph show distribution of mean farm level rates from the SOIL Model (Linear soil texture conditioned exponents), relative to rates from the exponents provided by RMA for 2019. For each Panel, the top [middle] distribution is for those policies with mean relative yield ratios above [below] one, and the bottom is for the entire sample. The dashed blue reference line is the point at which rates from the SOIL Model are equal to those from the RMA. The vertical axis of the pdfs has been omitted. Farm-level data was provided by the Kansas Farm Management Association (KFMA).

2.4.3 Economic Performance

Results for the ceded to retained indemnity ratios across all policies shown in Figure 2.3 Panel C indicate that the SOIL model consistently produces rates that are economically different. For all crops together, the ratio was approximately 1.40 indicating that indemnities were 40% larger among the ceded policies relative to the retained.

All crops had ratios above one with soybeans being the largest at ~1.5 and sorghum/wheat being relatively smaller at ~1.2. The difference could be attributed to wheat being a longer season crop spanning fall, winter, and spring months; whereby a large amount of weather variation obscures the soil information in the "signal". A somewhat similar situation arises for sorghum as well since it is sown under hotter conditions than soybeans and is also harvested later under colder conditions (sometimes into December). Nonetheless, all three models produce very similar economic gains for each crop, suggesting that the adjustment based on soil covariates alone is robust to including an additional adjustment at the CRD level.

One might suspect that the economic gains from including soil information are likely to decline with the amount of historical yield information provided by the farm. Soil effects are largely time-invariant and thus can likely be captured with a long enough yield history; however, it is unclear how short a history must be for soil to provide additional information not already captured by the rate yield. To investigate this, the indemnities within the ceded and retained pools are grouped by the number of years that were used in the rate yield calculation and the cede/retain ratios for each group are reported in Figure 2.5. Results suggest that economic gains do indeed decrease as yield histories become longer. Additionally, the rate of decline is striking as gains are essentially zero across all crops for the group with rate yields based on ten years of data. Focusing on the all-crop aggregate, the percentage reduction in economic gains between rate-yields based on a 10-year history, the maximum allowable by the RMA, is approximately 63, 42, and 28%, respectively for 0–3-, 4-, and 5–9-year history.

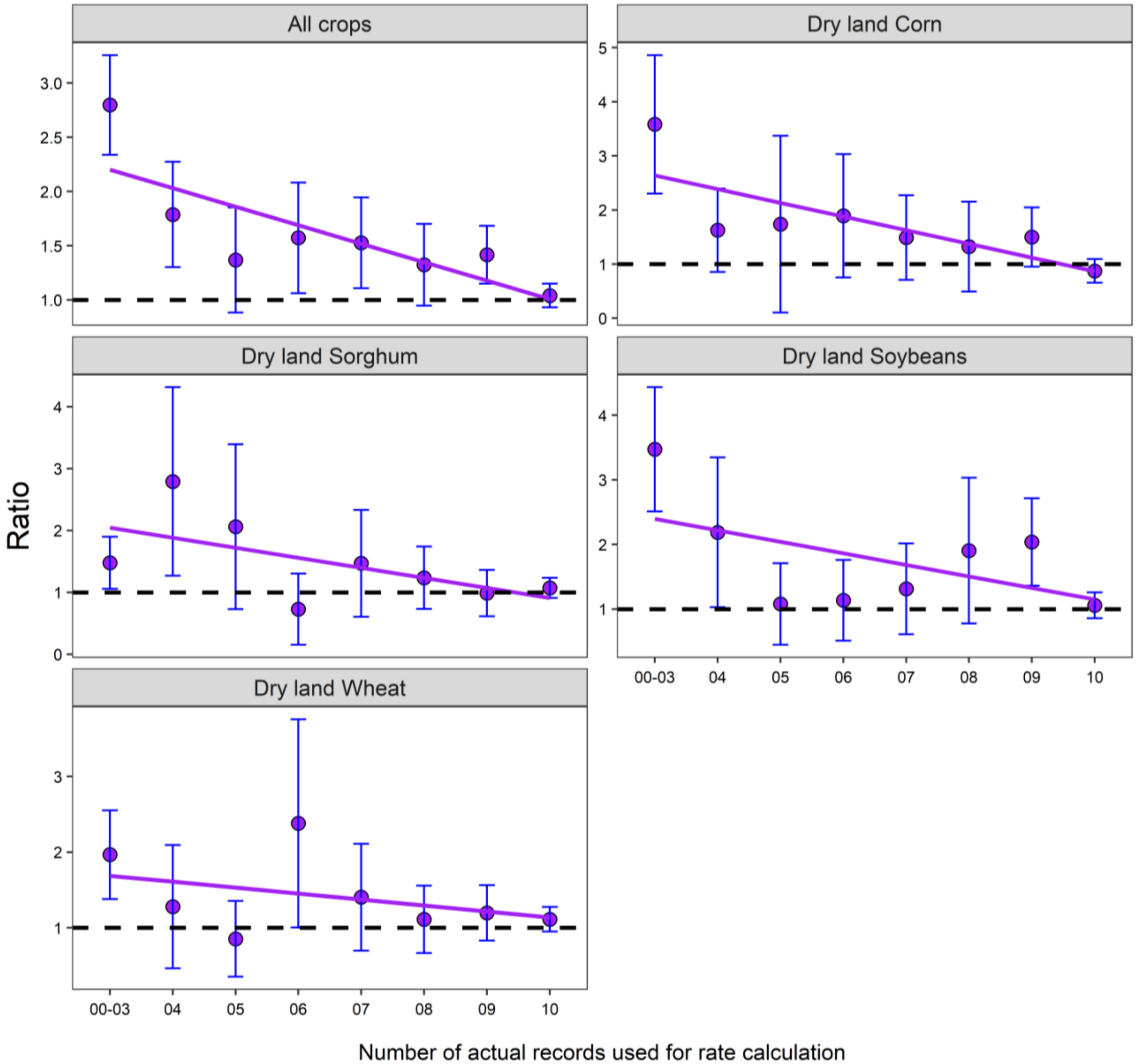


Figure 2.5: Relationship between actual production history length and economic performance of soil texture conditioned continuous rating exponents in Kansas

Notes: Graph shows the economic performance of crop insurance continuous rating exponents estimated from the SOIL Model (Linear soil texture conditioned exponents) across all crops, summarized by the length of actual production history. The performance is based on Coble et al. (2007) and Harri et al. (2011) and measures the level of forgone economic rents as the ratio of indemnities from ceded to that of retained policies under a simplified Standard Reinsurance Agreement (SRA) scenario. Values greater than one indicate a relatively higher level of forgone economic rents. The error bars represent the bootstrap (1,000) 95% confidence interval (CI) for the estimates, and the dashed black line marks the reference for one. Estimates with their 95% CI

2.4.4 Robustness of Main Finding

Figure A.11 shows that the pattern of declining economic gains as yield history length increases is robust across different dimensions of the empirical analysis. First, a wide range of alternative soil features is considered both alone and alongside the soil texture measures. These features include root zone depth, available water storage, slope, exchangeable cations, soil organic carbon, and the NCCPI. Second, while the analysis assumed a 75 percent coverage level because it was the largest enrolled across the four crops in Kansas since 2002 (RMA 2020), a full range of alternatives from 50-85 percent in 5-unit increments is considered. The third set of robustness checks extends the linear soil texture model to include polynomials of degree two to four in the adjustment function $f(.)$. The fourth set of robustness checks focuses on a key assumption of the measurement of soil information, in which the use of farm mailing addresses instead of specific geo-referenced field locations is used to match yield histories with soil data. Specifically, the buffer for aggregating soil information given the farms mailing address was varied from a 0.5- to 3-mile radius; and separably, all farms whose mailing address was in an urban area were dropped.[15] The final set of robustness checks are reported in Figure A.12 and use loss ratios (indemnities over premiums) in the cede-retain measure instead of indemnities alone since it is a more complete measure of economic rents and is commonly used in the literature. Overall, the pattern of results

[15] Depending on the radius used, 8-9.5% of the sample used fell within areas designated as urban by the Census Bureau. However, as shown in Note S1, soil information is aggregated over the area with each buffer that overlaps with the gridded crop frequency layer from NASS CropScape. For robustness checks, farms whose mailing address spatially intersected with Census Bureau's Urban Area Reference Maps were dropped regardless of their overlap with the crop frequency layer. The Census Bureau identifies two types of urban areas: (1) Urbanized Areas (UAs) of 50,000 or more people; and (2) Urban Clusters (UCs) of at least 2,500 and less than 50,000 people.

from the robustness checks in Figures A.11 and A.12 are consistent with the main findings in the manuscript.

2.5 Discussion and Conclusion

The study extends the crop insurance rating literature by incorporating topographic and soil information into rating procedures. A novel econometric approach based on RMA's procedures for pricing insurance at the farm level was developed and applied to a sample of 149,267 farm-level observations in Kansas spanning 1973-2018. The results show that including such information improves rate predictions on average and that the revised rates are economically different in the sense of a commonly used cede-retain game.

Overall, the results are largely in line with previous findings in the literature with the one key exception being that economic gains from including soil information rapidly decline with the yield history of the farm, with no gains associated with farms that provide ten years of historical yield data. This finding highlights a crucial dimension in the debate surrounding whether RMA should incorporate soil information into their rating procedures as it suggests that the proportion of policies for which ten years of data is available is an important variable in this decision.

Neither in this study nor RMA's database more generally, can it be assumed that entrance into the data is only driven by new farmers as there exist experienced farms that simply choose not to participate in various programs. So, it cannot be stated specifically from this study that the results apply directly to new farmers as a limited yield history could be driven by selection into the KFMA. However, to the extent that a combination of experienced and new farmers is driving the result, and that the benefit of including soil information for new farmers is at least as large as experienced farmers in a limited yield history context, then the results would provide a lower bound on the benefits of including soil information for new farmers. This could be further broken down into a distinction between experienced farmers growing a new crop versus young farmers with essentially no production history. In this view incorporating soil information can be beneficial for both young farmers and farmers that are switching crops, perhaps to adapt to changing environmental, climatic, and/or economic conditions. However, this is an important empirical question that warrants future research.

While the study did only focus on Kansas farms, the policy implications are likely externally valid for other major production states/regions. The results essentially show that yield history length and soil information are substitutes for inferring risk. It is common to control for farm/location “fixed effects” in production applications and the core insight is that at some point repeated sampling allows you to capture, or control for, time-invariant drivers of production variation. This insight holds for alternative moments of the yield distribution in the context of Just-Pope technology or more general “moments” models as well (Just and Pope 1979; Antle 2010). If one can repeatedly observe sample moments for two farms that are identical in every way except for their soil quality, then at some point one can disentangle the effect of soil on that moment. So, in principle, the more interesting question is how many repeated draws one requires to make this distinction, and it is likely that the amount of weather/pest/disease variation in the data matters a lot for this threshold since they would affect signal-to-noise ratios. In general, dryland crop production in Kansas is considered more variable relative to other major crop-producing regions such as the U.S. Corn Belt, so if it takes ten years of data to capture soil effects here it would probably be less in many other places. However, this is an empirical question that warrants future research.

Although no evidence of economic gains associated with ten-year yield histories was found, there are some additional considerations for utilizing soil conditioned rates that were not directly assessed here and thus might be considered by future research. First, incorporating soil information could help guard against moral hazard, as farmers can easily alter yields through various adjustments to production practices (e.g. fertilizer, pest control, seeding rates, etc.) but it is difficult in practice to adjust soil quality, especially within the growing season when moral hazard concerns may be highest (Coble et al. 1997). Second, several studies have shown that federal farm program payments impact land values (Barnard et al. 1997; Lence and Mishra 2003; Roberts, Kirwan and Hopkins 2003; Taylor and Brester 2005; Latruffe and Le Mouël 2009). Shaik, Helmers, and Atwood (2005) assert that any future efforts to reduce net agriculture subsidies could have large effects on land prices like that of the 1960s or 1970s. Thus, in addition to rates being important from an insurance perspective, getting them right or wrong could have implications for a farm's financial status through land capitalization.

In closing, several caveats to the analysis are worth mentioning. First, it focuses solely on the dryland operations of Kansas farms that produce corn, soybeans, sorghum, or wheat. Subsequent studies can overcome this by expanding the scope of this study to include a wider variety of crops, production practices (e.g., irrigation), and locations; however, it should be noted that the availability of farm-level panel data required for this type of analysis is quite limited. Second, the topographic and soil information used in this study is based on the mailing address of the farms in the KFMA database. Robustness checks in the analysis provide some evidence that this is a plausible working assumption, but future work might consider more tightly matched soil and production information if possible.

References

Adhikari, S., T.O. Knight, and E.J. Belasco. 2012. "Evaluation of Crop Insurance Yield Guarantees and Producer Welfare with Upward-Trending Yields." Agricultural and Resource Economics Review 41(3):367–376. Available at: https://www.cambridge.org/core/product/identifier/S1068280500001325/type/journal_article.

Adhikari, S., T.O. Knight, and E.J. Belasco. 2013. "Yield Guarantees and the Producer Welfare Benefits of Crop Insurance." Journal of Agricultural and Resource Economics 38(1):78–92. Available at: https://www.jstor.org/stable/23496740.

Annan, F., J. Tack, A. Harri, and K. Coble. 2014. "Spatial Pattern of Yield Distributions: Implications for Crop Insurance." American Journal of Agricultural Economics 96(1):253–268. Available at: https://academic.oup.com/ajae/article-lookup/doi/10.1093/ajae/aat085.

Anthony, P., G. Malzer, S. Sparrow, and M. Zhang. 2012. "Soybean Yield and Quality in Relation to Soil Properties." Agronomy Journal 104(5):1443–1458. Available at: http://doi.wiley.com/10.2134/agronj2012.0095.

Antle, J.M. 2010. "Asymmetry, partial moments, and production risk." American Journal of Agricultural Economics.

Barnard, C.H., G. Whittaker, D. Westenbarger, and M. Ahearn. 1997. "Evidence of Capitalization of Direct Government Payments into U.S. Cropland Values." American Journal of Agricultural Economics 79(5):1642–1650. Available at: https://onlinelibrary.wiley.com/doi/abs/10.2307/1244396.

Botts, R.R., and J.N. Boles. 1958. "Use of Normal-Curve Theory in Crop Insurance Ratemaking." Journal of Farm Economics 40(3):733. Available at: https://academic.oup.com/ajae/article-lookup/doi/10.2307/1235383.

Campbell, C.A., F. Selles, R.P. Zentner, and B.G. McConkey. 1993. "Available Water and Nitrogen Effects on Yield Components and Grain Nitrogen of Zero-Till Spring Wheat." Agronomy Journal 85(1):114–120. Available at: http://doi.wiley.com/10.2134/agronj1993.00021962008500010022x.

Carriquiry, M.A., B.A. Babcock, and C.E. Hart. 2008. "Using a farmer's beta for improved estimation of expected yields." Journal of Agricultural and Resource Economics 33(1):52–68. Available at: www.jstor.org/stable/41220613.

Coble, K.H., R. Dismukes, and J.W. Glauber. 2007. "Private crop insurers and the reinsurance fund allocation decision." American Journal of Agricultural Economics.

Coble, K.H., T.O. Knight, B.K. Goodwin, M.F. Miller, R.M. Rejesus, and G. Duffield. 2010. "A Comprehensive Review of the RMA APH and COMBO Rating Methodology Final

Report." Available at: https://legacy.rma.usda.gov/pubs/2009/comprehensivereview.pdf [Accessed May 23, 2019].

Coble, K.H., T.O. Knight, R.D. Pope, and J.R. Williams. 1997. "An Expected-Indemnity Approach to the Measurement of Moral Hazard in Crop Insurance." American Journal of Agricultural Economics 79(1):216–226. Available at: https://onlinelibrary.wiley.com/doi/abs/10.2307/1243955.

Congressional Budget Office [CBO]. 2017. "Options to Reduce the Budgetary Costs of the Federal Crop Insurance Program." Available at: https://www.cbo.gov/system/files/115th-congress-2017-2018/reports/53375-federalcropinsuranceprogram.pdf [Accessed September 13, 2020].

Congressional Research Service [CRS]. 2015. "Proposals to Reduce Premium Subsidies for Federal Crop Insurance." Available at: https://www.everycrsreport.com/files/20150320_R43951_5c31d4456d377edd9eae660d5cda968ea4299874.pdf [Accessed June 4, 2020].

Corwin, D.L., S.M. Lesch, P.J. Shouse, R. Soppe, and J.E. Ayars. 2003. "Identifying Soil Properties that Influence Cotton Yield Using Soil Sampling Directed by Apparent Soil Electrical Conductivity." Agronomy Journal 95(2):352–364. Available at: http://doi.wiley.com/10.2134/agronj2003.3520.

Cox, M.S., P.D. Gerard, M.C. Wardlaw, and M.J. Abshire. 2003. "Variability of Selected Soil Properties and Their Relationships with Soybean Yield." Soil Science Society of America Journal 67(4):1296–1302. Available at: http://doi.wiley.com/10.2136/sssaj2003.1296.

Goodwin, B.K. 1994. "Premium Rate Determination In The Federal Crop Insurance Program: What Do Averages Have To Say About Risk?" Journal of Agricultural and Resource Economics 19(2):382–395. Available at: https://ideas.repec.org/a/ags/jlaare/30747.html [Accessed May 19, 2019].

Harri, A., K.H. Coble, A.P. Ker, and B.J. Goodwin. 2011. "Relaxing Heteroscedasticity Assumptions in Area-Yield Crop Insurance Rating." American Journal of Agricultural Economics 93(3):707–717. Available at: https://academic.oup.com/ajae/article-lookup/doi/10.1093/ajae/aar009.

James, G., D. Witten, T. Hastie, and R. Tibshirani. 2013. An Introduction to Statistical Learning. New York, NY: Springer New York. Available at: http://link.springer.com/10.1007/978-1-4614-7138-7.

Juhos, K., S. Szabó, and M. Ladányi. 2016. "Explore the influence of soil quality on crop yield using statistically-derived pedological indicators." Ecological Indicators 63:366–373. Available at: https://linkinghub.elsevier.com/retrieve/pii/S1470160X15007347.

Just, R.E., and R.D. Pope. 1979. "Production Function Estimation and Related Risk Considerations." American Journal of Agricultural Economics 61(2):276. Available at: http://ajae.oxfordjournals.org/content/61/2/276.abstract [Accessed October 21, 2015].

Kravchenko, A.N., and D.G. Bullock. 2000. "Correlation of Corn and Soybean Grain Yield with Topography and Soil Properties." Agronomy Journal 92(1):75–83. Available at: http://doi.wiley.com/10.2134/agronj2000.92175x.

Latruffe, L., and C. Le Mouël. 2009. "CAPITALIZATION OF GOVERNMENT SUPPORT IN AGRICULTURAL LAND PRICES: WHAT DO WE KNOW?" Journal of Economic Surveys 23(4):659–691. Available at: http://doi.wiley.com/10.1111/j.1467-6419.2009.00575.x.

Lence, S.H., and A.K. Mishra. 2003. "The Impacts of Different Farm Programs on Cash Rents." American Journal of Agricultural Economics 85(3):753–761. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-8276.00480.

Li, D., T. Nanseki, Y. Chomei, T. Sasaki, T. Butta, and A. Numata. 2019. "Determining the Effects of Nitrogen Fertilizer and Soil Chemical Properties on Rice Yield: Evidence from a Large-Scale Farm in Hokuriku Region, Japan." International Journal of Plant Production 13(4):329–338. Available at: http://link.springer.com/10.1007/s42106-019-00058-4.

Lusk, J.L. 2017. "Distributional Effects of Crop Insurance Subsidies." Applied Economic Perspectives and Policy 39(1):1–15. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1093/aepp/ppw002.

Mahul, O., and C.J. Stutley. 2010. Government Support to Agricultural Insurance. The World Bank. Available at: http://elibrary.worldbank.org/doi/book/10.1596/978-0-8213-8217-2 [Accessed June 4, 2020].

Martín, N.F., G. Bollero, and D.G. Bullock. 2005. "Associations between field characteristics and soybean plant performance using canonical correlation analysis." Plant and Soil 273(1–2):39–55. Available at: http://link.springer.com/10.1007/s11104-004-6639-1.

Milliman & Robertson. 2000. "Actuarial Documentation of Multiple Peril Crop Insurance Ratemaking Procedures." RMA Actuarial Methodology Publications. Available at: https://legacy.rma.usda.gov/pubs/2000/mpci_ratemaking.pdf [Accessed May 9, 2020].

Nyiraneza, J., A.N. Cambouris, N. Ziadi, N. Tremblay, and M.C. Nolin. 2012. "Spring Wheat Yield and Quality Related to Soil Texture and Nitrogen Fertilization." Agronomy Journal 104(3):589–599. Available at: http://doi.wiley.com/10.2134/agronj2011.0342.

Ramirez, O.A., C.E. Carpio, and R.M. Rejesus. 2011. "Can Crop Insurance Premiums Be Reliably Estimated?" Agricultural and Resource Economics Review 40(1):81–94. Available at:

https://www.cambridge.org/core/product/identifier/S1068280500004536/type/journal_article.

Rejesus, R.M., K.H. Coble, M.F. Miller, R. Boyles, B.K. Goodwin, and T.O. Knight. 2015. “Accounting for weather probabilities in crop insurance rating.” Journal of Agricultural and Resource Economics 40(2):306–324.

Rejesus, R.M., B.K. Goodwin, K.H. Coble, and T.O. Knight. 2010. “Evaluation of the reference yield calculation method in crop insurance.” Agricultural Finance Review 70(3):427–445. Available at: https://www.emerald.com/insight/content/doi/10.1108/00021461011088530/full/html.

Risk Management Agency [RMA]. 2018. “2019 Crop Insurance Handbook.” Available at: https://legacy.rma.usda.gov/handbooks/18000/2019/19_18010.pdf [Accessed May 30, 2020].

Risk Management Agency [RMA]. 2019a. “ACTUARIAL RELEASE: Actuarial Release–19-044.” Available at: https://www.rma.usda.gov/Policy-and-Procedure/Bulletins-and-Memos/2019/Actuarial-19-044 [Accessed May 30, 2020].

Risk Management Agency [RMA]. 2019b. “Crop Year Government Cost of Federal Crop Insurance Program.” Available at: https://www.rma.usda.gov/-/media/RMAweb/AboutRMA/Program-Budget/18cygovcost.ashx?la=en [Accessed May 19, 2019].

Risk Management Agency [RMA]. 2000. “Premium Rate Calculations for the Continuous Rating Model.” Available at: http://www.rma.usda.gov/ftp/Continuous_Rating/cr_instr.pdf [Accessed May 30, 2020].

Risk Management Agency [RMA]. 2020. “Summary of Business Data.” Available at: https://prodwebnlb.rma.usda.gov/apps/SummaryOfBusiness/ [Accessed June 12, 2020].

Roberts, M.J., B. Kirwan, and J. Hopkins. 2003. “The Incidence of Government Program Payments on Agricultural Land Rents: The Challenges of Identification.” American Journal of Agricultural Economics 85(3):762–769. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-8276.00481.

Sene, M., M.J. Vepraskas, G.C. Naderman, and H.P. Denton. 1985. “Relationships of Soil Texture and Structure to Corn Yield Response to Subsoiling.” Soil Science Society of America Journal 49(2):422–427. Available at: http://doi.wiley.com/10.2136/sssaj1985.03615995004900020030x.

Seo, H.S., T. Kim, M.K. Kim, and B.A. McCarl. 2017. “Introduction of the trend-adjusted crop insurance program: Participation- and coverage-level responses.” Agricultural Finance Review.

Shaik, S., G.A. Helmers, and J.A. Atwood. 2005. “The Evolution of Farm Programs and Their Contribution to Agricultural Land Values.” American Journal of Agricultural Economics

87(5):1190–1197. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8276.2005.00806.x.

Skees, J.R., and M.R. Reed. 1986. "Rate Making for Farm-Level Crop Insurance: Implications for Adverse Selection." American Journal of Agricultural Economics 68(3):653. Available at: https://academic.oup.com/ajae/article-lookup/doi/10.2307/1241549.

Smith, V.H., and J.W. Glauber. 2012. "Agricultural Insurance in Developed Countries: Where Have We Been and Where Are We Going?" Applied Economic Perspectives and Policy 34(3):363–390. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1093/aepp/pps029 [Accessed June 18, 2019].

Soil Survey Staff. 2020. "Gridded Soil Survey Geographic (gSSURGO) Database for the Conterminous United States (FY2020 official release)." United States Department of Agriculture, Natural Resources Conservation Service. Available at: https://nrcs.app.box.com/v/soils/folder/94124173798 [Accessed May 30, 2020].

Taylor, M.R., and G.W. Brester. 2005. "Noncash Income Transfers and Agricultural Land Values." Review of Agricultural Economics 27(4):526–541. Available at: https://academic.oup.com/aepp/article-lookup/doi/10.1111/j.1467-9353.2005.00262.x.

United States Government Accountability Office [GAO]. 2014. "Considerations in Reducing Federal Premium Subsidies." Available at: https://www.gao.gov/assets/670/665267.pdf [Accessed June 4, 2020].

USDA National Agricultural Statistics Service. 2019. "CropScape - Cropland Data Layer (2019 release)." Available at: https://nassgeodata.gmu.edu/CropScape/.

Di Virgilio, N., A. Monti, and G. Venturi. 2007. "Spatial variability of switchgrass (Panicum virgatum L.) yield as related to soil parameters in a small field." Field Crops Research 101(2):232–239. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0378429006002346.

Woodard, J.D., B.J. Sherrick, and G.D. Schnitkey. 2011. "Actuarial impacts of loss cost ratio ratemaking in U.S. crop insurance programs." Journal of Agricultural and Resource Economics.

Woodard, J.D., and L.J. Verteramo-Chiu. 2017. "Efficiency Impacts of Utilizing Soil Data in the Pricing of the Federal Crop Insurance Program." American Journal of Agricultural Economics 99(3):757–772. Available at: https://academic.oup.com/ajae/article-lookup/doi/10.1093/ajae/aaw099.

Chapter 3 - Tradeoffs Between Production-History-Based and Index-Based Insurance for Field Crops

3.1. Introduction

Initiated in Europe over two centuries ago, the agricultural insurance sector is a large and rapidly expanding component of support programs for farmers and ranchers in both developed and developing countries (Mahul and Stutley 2010; Smith and Glauber 2012a). The most recent global survey in 2008 by the World Bank shows that the global agricultural insurance market across 65 advanced and emerging countries generated $15 billion in premiums, which helped producers cover losses and stabilize revenues (Mahul and Stutley 2010). The product space of the agricultural insurance market includes individual/area-wide policies covering yield and revenue support for both crops and livestock. These policies are further segregated into two broad groups based on the mode of indemnity trigger and pricing, i.e., Actual Production History [APH] based and Index-based schemes.[16] While APH-based policies abound, the global importance and the range of Index insurance products in recent times have also expanded in both developing and developed countries which have generated extensive literature (Carter et al. 2017; Barnett and Mahul 2007; Miranda and Farrin 2012; Jensen and Barrett 2017; Vroege, Dalhaus and Finger 2019).

APH-based policy pricing and indemnifications are based on the actual farming experience of the purchaser (Risk Management Agency [RMA] 2020), while Index-based policies are based

[16] APH based products include both revenue and yield insurance but the analysis will focus only on APH based yield insurance products.

on the realizations of an Index that is external to the purchaser (Collier, Barnett and Skees 2010; Coble et al. 2020). In addition to the design difference, relatively speaking the former suffers from adverse selection (Smith, Glauber and Goodwin 2017; Glauber 2013), moral hazard (Chambers 1989; Horowitz and Lichtenberg 1993; Yu and Hendricks 2020; Park et al. 2020), and high program delivery cost (Barnett and Mahul 2007). Thus, if properly designed to address moral hazard and adverse selection, Index-based insurance offers the opportunity of reducing the cost associated with program delivery (Barnett and Mahul 2007). For example, Belasco, Cooper and Smith (2019) showed that replacing the current suite of policies in the US Federal Crop Insurance Program (FCIP) with an Index-based crop disaster program could lead to savings of $3-4 billion annually realized through focusing agricultural support on systemic weather risk rather than idiosyncratic risk; and reducing program delivery cost. In the US, rainfall index insurance for pasture rangeland and forage (PRF-RI), vegetation Index insurance, and area-based yield/revenue insurance for row crops are already available (Risk Management Agency [RMA] 2020). However, in general, these products suffer from weak demand due to basis risk (Jensen and Barrett 2017; Carter et al. 2017; Barnett and Mahul 2007).

An extensive body of literature has shown that basis risk could be reduced to an appreciable extent by improving product design (Conradt, Finger and Spörri 2015; Dalhaus and Finger 2016; Dalhaus, Musshoff and Finger 2018; Vroege et al. 2019; Bucheli, Dalhaus and Finger 2020; Vroege et al. 2021). However, assuming the best Index is operationalized, the tradeoffs between APH- and Index-based insurance at the farm level remain largely unknown for (at least) two reasons. First, Index-based insurance covering a broad range of weather indices for traditional field crops is missing, thus observable data is lacking as evidenced by the limited datasets used in previous studies (Carriker et al. 1991; Miranda 1991; Smith, Chouinard and Baquet 1994; Deng, Barnett and Vedenov 2007; Jensen, Barrett and Mude 2016, Barnett et al. 2005) or simulated data based on strong assumptions (Stigler and Lobell 2021). Second, for any meaningful insights to be drawn from these tradeoffs, one should observe separately the potential outcomes of both APH- and Index-based insurance for the same farmer under identical conditions so that direct comparative advantages can be made.

The novelty of this study is that it overcomes the data limitations and fills the knowledge gap of the missing tradeoffs. Two related objectives are pursued: (i) ascertain the potential outcomes of a broad range of weather Index-based insurance and APH-based insurance under specified farm income goals; and given this, (ii) determine if the potential outcomes are different. The objectives are achieved by deploying two parallel *ex-post* simulations to generate the two outcomes at the farm-level and then assesses their tradeoffs. The empirical strategy is applied to farm-level yields for corn, soybean, sorghum, and wheat under known conditions in Kansas and spanning 46 years (1973-2018). The results show that economically significant tradeoffs do exist between APH- and Index-based insurance and that different types of Index products are associated with differing levels of basis risk. Particularly, Index-based insurance that protects against excess accumulation in killing-degree-days (i.e., degree-days >30 °C) generates the most significant gains in economic rents and is associated with relatively low basis risk. [17]

The remainder of Chapter 3 is organized as follows. Section 3.2 provides background information of both APH- and Index-based insurance as well as their core differences and touches on basis risk as it relates to this study. The methods section follows next in section 3.3 and it outlines the product designs for the insurance products simulated, how purchasers choose among

---

[17] The study focusses on yield protection even though most crop insurance in the U.S. today is revenue protection. Revenue insurance includes a yield risk component that is based on that of yield insurance.

alternative Index-based insurance enrolment parameters, and how the tradeoffs between APH- and Index-based insurance are evaluated. The data section is in Section 3.4, and Sections 3.5 and 3.6 present the core results and conclusions, respectively.

3.2. Background Information

Agricultural insurance costs and payouts are normally quoted in equivalent monetary terms that are tied to the price per unit of the underlying products and their extensive margin (i.e., land size). In what follows, the study abstracts from that reality by assuming that the underlying product is priced at unity with an extensive margin of one. Thus, without loss of generality, all cost and payouts are per unit bases on output terms (i.e., kg/ha).

3.2.1 Basic elements of APH-based contracts

The building blocks for an APH-based contract are the rate yield ($\bar{y}_i^A$), approved yield ($\ddot{y}_i^A$), yield guarantee ($\tilde{y}_{ig}^A$), coverage level (C_{ig}^A), indemnity (I_{ig}^A), premium rate (R_{ig}^A), premium (P_{ig}^A), and subsidy (S_g^A), where i denotes farm, g denotes coverage level, and A denotes APH-based contract type. The rate and approved yield are both derivatives of the insureds reported APH such that the former is the simple average of APH and the latter is the same but with upward adjustments including yield exclusion, yield substitution, and trend. [18] The purchaser elects the coverage level to indicate the proportion of the approved yield to be insured such that $\tilde{y}_{ig} = \ddot{y}_i \cdot C_g$.[19] The per-

[18] Note that to the extent that approved yield is higher than rate yield, as is often the case, this benefits producers as the yield guarantee will be higher, and thereby, will increase indemnities for a given yield outcome and improves producer welfare (Adhikari, Knight and Belasco 2013)

[19] Federally approved coverage levels for the 2019 crop year ranged from 55-85% in 5% increments.

acre indemnity for a given yield outcome, y_{it}, is given by $I_{ig} = max\{0, \tilde{y}_{ig} - y_{it}\}$.[20] The final price paid by the insured is $P_{ig} = R_{ig}\, \tilde{y}_{ig} S_g$, where the premium rate, (R_{ig}), is determined using a continuous rating formula (Risk Management Agency [RMA] 2000) and S_g is a subsidy factor determined by FCIC and is tied to coverage level.[21]

3.2.2 Basic elements of Index-based contracts

The building blocks for an Index-based contract are the Index variable, a grided surface, Index interval, Index interval weight (W^I_{irv}), expected grid Index (E^I_{rv}), trigger grid Index (T^I_{ig}) (same as coverage level), final grid Index (F^I_{rv}), base premium rate (R^I_{rvg}), policy protection per unit, county base value ($\bar{y}^I_c$), productivity factor ($\overline{PF}^I_i$), premium (P^I_{irvg}), premium subsidy (S^I_g), indemnity (I^I_{irvg}), and payment calculation factor, where *i* denotes farm, *r* denotes grid ID, *v* denotes interval ID, g denotes coverage level, and the exponent *I* denotes Index-based contract type.

As indicated in the introduction, Index-based insurance based on a broad range of indices (e.g., weather) for traditional field crops in the US is missing, so the study draws from the PRF-RI. The Index variable (i.e., the variable used in constructing the Index) for the PRF-RI is grided precipitation but we considered other variables as well. The Index interval is the specified period (combinations of at most two successive months without overlap) for which data on the Index variable is collected. Given the two-month interval rule, insureds then choose among 11 intervals

---

[20] In practice, indemnities are the product of the yield shortfall and price = $max\{0, \tilde{y}_{ig} - y_{it}\}\times$ price. However, in the current setup, price = 1, so $I_{ig} = max\{0, \tilde{y}_{ig} - y_{it}\}$.

[21] For the 2019 crop insurance program, corn, soybeans, sorghum, and wheat policies with coverage levels of 0.55, 0.65, 0.75, and 0.85 had S_g respectively equal to 0.64, 0.59, 0.55, and 0.38. Between 2005-2018, the federal government subsidized on average 61.1% of farmers' premiums (Risk Management Agency [RMA] 2019b).

during the year: Jan/Feb, Feb/Mar, Mar/Apr, Apr/May, May/Jun, Jun/Jul, Jul/Aug, Aug/Sep, Sep/Oct, Oct/Nov, and Nov/Dec. The no overlap rule (e.g., cannot choose Aug/Sep and Sep/Oct jointly for the same grid) essentially limits the maximum number of intervals to six. Finally, purchasers must assign weights ranging from 10-60% to each interval selected such that the sum of weights across all selected intervals within a given grid is 100%.

The Index variable and interval are used to calculate two key pieces of information for each grid and interval: expected grid Index (E_{rv}^{I}) (i.e., mean accumulated value of the Index variable over a base period) and final grid Index (F_{rv}^{I}) (current accumulated value of the Index variable), both expressed as a percentage of E_{rv}^{I}. The other piece of information is the trigger grid Index (T_{ig}^{I}) (analogous to coverage level for the case of APH). Historic values of the E_{rv}^{I} and F_{rv}^{I} are used to estimate base premium rates for each grid, Index interval, and T_{ig}^{I} combination which helps to minimize adverse selection and moral hazard.

Like that of the APH-based products, the size of the premium and indemnity depends on the policy protection per unit (i.e., liability) which is the product of the insured acres, share of acres insured, and dollar amount of protection (DAP). The DAP is the product of county base value ($\bar{y}_{c}^{I}$), productivity factor ($\overline{PF}_{i}^{I}$), coverage level (same as T_{ig}^{I}), and the Index interval weight (W_{irv}^{I}). Here $\bar{y}_{c}^{I}$ reflects the mean level of output per unit at the county level and $\overline{PF}_{i}^{I}$ is selected by the insured to individualize coverage based on their perceived relative productivity to that of their peers in the same county. The producer paid premium (P_{irvg}^{I}) for a given combination of grid/interval/coverage is given as the product of the liability, base premium rate (R_{rvg}^{I}) for the grid/interval/coverage, and a subsidy rate (S_{g}^{I}) tied to coverage and determined by the FCIC. Depending on the nature of the insurance, when there is a loss/gain in the Index value the indemnity due for a given grid/interval/coverage is the product of a payment calculation factor (PCF) and the liability.

Conventionally, $PCF = max\left\{0, \frac{T_{ig}^{I} - F_{rv}^{I}}{T_{ig}^{I}}\right\}$ for products that insure against a shortfall in the accumulation of the underlying Index variable. However, while the study is unaware of any such products, an alternative calculation is $PCF = max\left\{0, \frac{F_{rv}^{I} - T_{ig}^{I}}{T_{ig}^{I}}\right\}$ for products that insure against excess in accumulation of the underlying Index variable. Here, the study adopts the shortfall case to adequately represent some basic ideas. Putting all these together, the premium (P_{ig}^{I}) and indemnity (I_{ig}^{I}) for a given farm and coverage level across all grids and intervals are given by

$$P_{ig}^{I} = \sum_r \sum_v P_{irvg}^{I} = \bar{y}_c^{I} \cdot T_{ig}^{I} \cdot \overline{PF}_i^{I} \cdot \sum_r \sum_v \left[W_{irv}^{I} \times R_{rvg}^{I}\right] \quad (3.1)$$

$$I_{ig}^{I} = \sum_r \sum_v I_{irvg}^{I} = \bar{y}_c^{I} \cdot T_{ig}^{I} \cdot \overline{PF}_i^{I} \cdot \sum_r \sum_v \left[W_{irv}^{I} \times max\left\{0, \frac{T_{ig}^{I} - F_{rv}^{I}}{T_{ig}^{I}}\right\}\right] \quad (3.2).$$

3.2.3 Similarities and differences between APH- and Index-based contracts[22]

Below is a simple description of the insurance cycle intended to highlight key differences between the administration of APH- and Index-based contracts. The insurance cycle for both contracts broadly follows a similar path that starts with the purchaser providing the needed information to the insurance provider for a contract to be drafted and priced. For the case of the APH scheme, the purchaser provides an APH database if available and then selects the desired coverage level. For the case of the Index scheme, the purchaser need not provide an APH database but rather the grid IDs associated with the fields they want to insure. Along with the grid IDs, the purchaser will also elect the intervals for each grid, the coverage level, and their perceived

[22] It is worth noting that the insurance variables like premiums and indemnities may not directly comparable across APH- and Index- based insurance products as presented so far. A normalization routine in this study is linear as such it is likely to be overly simple. A more robust procedure would be to use a Schlenker and Roberts (2009) type model in the same spirit as Belasco, Cooper, and Smith (2019).

productivity factor. The insurance provider will use the information together with predetermined parameters from RMA to price the contract for the purchaser for a specified insurance period.

After the purchase of the contract, the next stage is what happens after a loss is recorded. For the case of the APH scheme, the policyholder must first write a notice of damage/loss for each unit insured within 72 hours of initial discovery, but not later than 15 days after the end of the insurance period unless otherwise stated. This will allow the insurance provider to send loss adjusters to inspect and gather information on the damage/loss to assist the policyholder in filing the claim for indemnity. At the end of the insurance period, a check is sent to the policyholder if indemnities exceed the premium due. If the premium due exceeds the indemnity or no loss/damage was recorded, the policyholder instead receives the balance of the premium due as a bill. It is worth noting that the policyholder is responsible for establishing the time, location, cause, and amount of any loss. For the case of the Index-insurance, the policyholder need not contact the insurance provider for indemnification. The insurance provider calculates the final grid Index for the period for which the contract covers and makes indemnity payments to the policyholder given any shortfall in the final grid Index and insured elected parameters in the contract. Thus, indemnification for the case of Index insurance is made before the season ends, usually no more than 60 days following the determination of the final grid Index.

Major drawbacks of the APH-system are its large administrative costs, and the possibility of moral hazard on the part of policyholders at enrollment and during production to increase the chance of indemnification. One advantage of the APH-system over the Index-system is that it makes use of the policyholder's experience to rate the policy. Additionally, indemnification is based on actual loss/damage, so policyholders receive a payout when they experience an actual loss. This cannot be said about the Index insurance because of the basis risk.

3.2.4 Basis risk

In simple terms, basis risk is the disconnect between the Index variable and output. In practice, this can translate into three situations: (1) indemnification without actual loss; (2) indemnification that is less or higher than actual loss; and (3) actual loss without indemnification (Coble et al. 2020) thereby reducing the usefulness of Index insurance as a risk management tool. Several studies confirm basis risk as a supply-side issue (related to product design) that reduces demand for Index insurance (Vedenov and Barnett 2004; Barnett and Mahul 2007; Giné, Townsend and Vickery 2008; Binswanger-Mkhize 2012; Smith and Glauber 2012b; Elabed et al. 2013; Jensen et al. 2016; Clarke 2016; Jensen and Barrett 2017; Carter et al. 2017).

According to Coble et al. (2020), basis risk is the variance of the conditional distribution of actual losses given a specific Index value and they note that it is often difficult to model because of limited data. Thus, basis risk tends to be measured as the linear correlation/covariance between the Index and losses, but this may be misleading because of potential nonlinearities that are likely to exist (Collier et al. 2010). A simple but intuitive measure widely used is a false negative probability (FNP), i.e., the probability of actual loss without indemnification (Elabed et al. 2013; Yu et al. 2019). According to Elabed et al. (2013), if the FNP is large then farmers will place less value on Index insurance and may choose not to buy it; thereby directly reducing its feasibility as an effective risk management instrument. This study also uses FNP as the indicator to measure the level of basis risk associated with a broad range of Index insurance products for traditional row crops.

3.3. Methods

The study simulates stylized schemes for APH- and Index-based insurance like Chapter 2. Premiums and indemnities are simulated *ex-post* such that all farmers in a historic farm-level data spanning 46 years (1973-2018) participate in both schemes for every year data is available for them. This ensured that adverse selection (since all farmers participate in both schemes) and moral hazard (yield remain the same under both schemes) are ruled out. While this is an abstraction from reality since it is unlikely for a single farmer to purchase both APH- and Index-based insurance, it makes it possible for one to observe separately the potential outcomes of both schemes for the same farmer under identical conditions.

The APH-insurance simulation is the same as that of Chapter 2 but for the case where their adjustment function is omitted entirely (i.e., RMA 2019 parameters are taken as given). For the APH-based simulations, the analysis assumed a 75 percent coverage level since it has the largest enrollment in Kansas since 2002 for the crops considered (RMA 2020). For the Index-based insurance simulation, the study simulates outcomes for indices based on precipitation, soil moisture, and 19 indices for degree-days for the 11 Index intervals discussed above. The 19 indices for degree-days are based on thresholds of 15 to 29°C (i.e., beneficial growing-degree-days) and 30 to 33°C (i.e., harmful killing-degree-days) all in 1°C increment. Finally, across all the simulations, premium subsidies are not considered because since they are tied to coverage rather than product, they will cancel out when comparing two products of the same coverage level as in this study.

3.3.1 Index insurance designs

The study draws insights from the literature by assuming that output for the i^{th} farmer for season t (y_{it}) is not only random but also stochastically dependent on some random process $X(t)$ (e.g., weather) with a realized seasonal value of X_t. Elabed et al. (2013) represents this relationship as

$$y_{it} = g(X_t) + \vartheta_t + \eta_{it} \qquad (1).$$

Here, the function $g(\cdot)$ approximates the impact of X_t on y_{it}; ϑ_t accounts for the residual impact of the random process $X(t)$ not captured by X_t; and η_{it} captures the impact of other random factors uncorrelated with X_t. In practice, the parameters of the approximation function $g(\cdot)$ determines the design parameters of an Index insurance product based on the random process $X(t)$, and ϑ_t and η_{it} captures any inherent basis risk of the Index insurance. According to literature, the basis risk associated with ϑ_t can be reduced by improving the design parameters of the Index insurance such that $g(\cdot)$ predicts y_{it} precisely (Conradt, Finger and Bokusheva 2015; Bucheli et al. 2020). In this study, attention is given to the choice of the random process $X(t)$, i.e., the Index variable, rather than the parameters in $g(\cdot)$ as Bucheli et al. (2020) does.

For a given insurance period t, Index variable, interval, the expected grid Index (E_{rvt}^{I}) is the mean accumulated value of the Index variable over all periods $t-j$ such that $j \geq 2$. Likewise, the final grid Index (F_{rvt}^{I}) is just the current accumulated value of the Index variable. Depending on the Index variable, the Index insurance product was designed as a policy that protects against a shortfall or an excess in the accumulation of the underlying Index variable. Particularly, policies based on precipitation, soil moisture, and growing-degree-days (i.e., degree-days with thresholds of 15 to 29°C) insured against a shortfall in accumulation of the underlying Index variable. Policies based on killing-degree-days (i.e., degree-days with thresholds of 30 to 33°C) insured against excess in accumulation. The rationale for this distinction is that naturally, relatively more precipitation and high soil moisture improves yields. Also, previous studies show that crop yield growth increases linearly up to a temperature threshold (i.e., the optimal temperature) and then sdecreases linearly for every 1°C increase above that threshold : the optimal temperature for various crops tend to be between 29-33°C (Schlenker and Roberts 2009; Tack, Barkley and Nalley 2015).

In practice, the insurance provider will use historic values of the E_{rvt}^{I} and F_{rvt}^{I} to estimate base premium rates for each grid, Index interval, and T_{ig}^{I} combination, annually. However, because of data limitation, the study used the entire sample space for each grid to estimate the rates for the respective grid so that for each farm/year policy simulated, the grid level rates do not change but the E_{rvt}^{I} and F_{rvt}^{I} do.

The current RMA procedure used to rate the PRF-RI relies on nonparametric empirical burn rates which are bounded based on rates derived from parametric distributions (Lognormal, Truncated Normal, and Gram Charlier (GC) expansion) (Coble et al. , 2020). As suggested by Coble et al. (2020), this study uses a similar approach but expands the range of the bounding parametric rates to include those from Weibull and Gamma distributions. Furthermore, the study excludes the rates from the Gram Charlier (GC) expansion distribution. The specific bounding undertaken in the rating process involves the following steps:

1. For each Index variable/grid/interval/coverage combination, calculate the empirical burn rates as the mean of the PCF; i.e., $max\left\{0, \frac{T_{ig}^{I}-F_{rv}^{I}}{T_{ig}^{I}}\right\}$ for the case of insuring against a shortfall and $max\left\{0, \frac{F_{rv}^{I}-T_{ig}^{I}}{T_{ig}^{I}}\right\}$ for the case of insuring against an excess.
2. Use the burn rates to estimate parametric rates implied by the Lognormal, Truncated Normal, Weibull, and Gamma distributions
3. If the burn rate is less than the maximum of the parametric rates an initial raw rate is set to the minimum of the parametric rates.
4. If the burn rate is higher than the maximum of the parametric rates an initial raw rate is set to the maximum of the parametric rates.

5. If the burn rate is between the minimum and maximum of the parametric rates, the raw rate is set to the burn rate.
6. The raw rate is then loaded by dividing by 0.88 to get the base premium rate.

For the PRF-RI, the final base rate for a given grid is determined via a spatial smoothing algorithm that takes a weighted average of the grid's raw rate and those from contiguous grids. In this study, the spatial smoothing algorithm was not employed because of technical constraints.

The study utilizes the E^I_{rvt}, F^I_{rvt}, and the base premium rate as the design parameters in the simulation that follows.

3.3.2 Index insurance interval allocation

In this study, the decision to sign up for insurance is automatic, but for each farm/year combination, the study must choose the coverage level (i.e., trigger grid Index [T^I_{ig}]), productivity factor ($\overline{PF}^I_i$), and Index intervals and assign Index interval weight (W^I_{irv}) to them. For ease of comparison to corresponding APH-based insurance simulation, the analysis assumed a 75 and 125% coverage level for products that protect against a shortfall and excess in accumulation of the respective Index, respectively. The $\overline{PF}^I_i$ was taken as $\bar{y}_i/\bar{y}_{cr}$, which is the producer's rate yield ($\bar{y}_i$) over a county-level reference yield ($\bar{y}_{cr}$). Also, it will be shown in the data section, the empirical implementation that aggregates E^I_{rv}, F^I_{rv}, and R^I_{rvg} which essentially removes the need for grid ID selections. So, the only variable left to choose is W^I_{iv} for all farm/year combinations in the dataset and each of the 21 Index-insurance products deployed in the simulation.

Given the assumptions made so far, and following similar empirical applications elsewhere (Popp and Rudstrom 2000; Nalley et al. 2009; Barkley, Peterson and Shroyer 2010), the study relies on Markowitz portfolio theory (Markowitz 1952) to simulate how farmers purchase insurance by allocating W^I_{irv}. Here the intuition is that direct production and Index intervals are investment assets. The distribution of proceeds from direct production (i.e., $v = 0$) and each of the 11 Index intervals (i.e., $\forall v \in [1,11]$) are given by $y_{it0} \sim iid\ N(\bar{y}_{i0}, \sigma_{i0}^2)$ and $y_{itv} \sim iid\ N(\bar{y}_{iv}, \sigma_{iv}^2)$; and the covariance of any two assets is $\text{cov}(\bar{y}_{ik}, \bar{y}_{iv}) = \sigma_{ivk}$. Thus, the expected proceeds ($E(\pi_{it})$) and variance of proceeds ($V(\pi_{it})$) for the i^{th} farm at time t is given by

$$E(\pi_{it}) = W_{i0}^I E(y_{it0}) + \sum_v W_{itv}^I E(y_{itv}) \tag{3.3}$$

$$E(y_{itv}) = \bar{y}_c^I \cdot T_{ig}^I \cdot \overline{PF}_{it}^I \cdot max\left\{0, \frac{T_{ig}^I - E(F_{itv}^I)}{T_{ig}^I}\right\} - R_{itgv}^I \tag{3.4}$$

$$V(\pi_{it}) = \sum_v \sum_k W_{iv}^I W_{ik}^I \sigma_{ivk} \tag{3.5}$$

All variables are defined in sections 3.2.1 and 3.2.2, but the most important thing to note is that farm proceeds are the summation of the direct expected proceeds from production ($E(y_{it0})$) plus the expected proceeds from insurance for each Index interval (y_{itv}). Also note that the direct cost of production and premium subsidies are eliminated, because they are the same for both APH- and Index-based schemes.

Given the above set up, the farmer can allocate the weights to each investment by pursuing one of five objectives: (1) finding the global minimum variance portfolio; (2) finding the global maximum proceeds portfolio; (3) finding the efficient portfolio for a target level of risk; (4) finding the efficient portfolio for a target level of proceeds; or (5) equally allocated portfolio. In this study, only objective (1) is considered since it reflects the reason for insurance purchase (i.e., to deal with risk), and the study assumes that farmers are risk-averse.[23] Thus, the constrained minimization problem for each farm/year combination is

$$\min_{W_{iv}^I} V(\pi_{it}) = \sum_{v=0}^{11} \sum_{k=0}^{11} W_{iv}^I W_{ik}^I \sigma_{ivk} \tag{3.6}$$

[23] Alternatively, farmers may be buying insurance to maximum expected payments since the premiums are subsidized. However, since subsidies are not considered, the study does not consider this objective.

[constraint 1] $W_{i0}^{I} = 1$

[constraint 2] $\sum_{v=1}^{12} W_{iv}^{I} = 1$

[constraint 3] $W_{i,v}^{I} \cdot W_{i,v+1}^{I} = 0 \; \forall v \in [1,11]$

[constraint 4] $W_{iv}^{I} = \begin{cases} 0.1 \leq W_{iv}^{I} \leq 0.6 & W_{iv}^{I} > 0 \\ 0 & W_{iv}^{I} = 0 \end{cases}$.

Since insurance purchase is an afterthought of a production decision, W_{i0}^{I} is always set to one [constraint 1], but the sum of weight for the Index intervals must equal one [constraint 2]. Additionally, as in the case of the PRF-RI, purchasers must assign weights to the non-overlapping interval [constraint 3] that range from 10-60% [constraint 4].

The solution for the global minimum variance portfolio from the constrained minimization problem is used to calculate *ex-post* the premiums and indemnities for each farm/year combination and then compared to those from the APH-based schemes along four dimensions: (1) indemnification pattern (i.e., basis risk), (2) farm-level outcome, and (3) economic significance.

3.3.3 Comparisons of potential outcomes

As noted above, the study used alternative Index variables to gauge which is associated with low basis risk. To do this, the study used two measures. The first is false-negative probability (FNP) (Elabed et al. 2013; Yu et al. 2019); for a given Index variable this is defined as $FNP = Pr\left[I_{ig}^{I} = 0 | I_{ig}^{A} > 0\right] = Pr\left[I_{ig}^{I} = 0, I_{ig}^{A} > 0\right] / Pr\left[I_{ig}^{A} > 0\right]$. This is operationalized by first computing the size of the indemnities from both schemes across all farm/year combinations, and then empirically compute the joint and marginal probabilities to obtain the FNP. This was done for each type of Index variable considered. The FNP does not tell us about the level of basis risk at the farm level so the study considers another measure defined as $basis = \frac{max\left\{0, I_{ig}^{A} - I_{ig}^{I}\right\}}{\bar{y}_{i}^{A}} \times$

100%. Here a no-zero value indicates the extent of the short fall in the indemnity due the farmer with relatively higher values indicating relatively high short falls.

For the economic significance, this study uses the same method as Chapter 2 with the only difference being that the soil-adjusted outcomes are replaced with outcomes from Index-based insurance. However, unlike Chapter 2, this study reports the cede to retain LR ratios instead of indemnity ratios. Finally, to get a measure of statistical uncertainty, the various outcome measures are repeatedly calculated 1,000 times by bootstrap sampling the farms in the data.

3.4. Data

3.4.1 Sources

Six main sources of data are utilized; (1) 46 years of farm-level Kansas corn, sorghum, soybean, and wheat yields provided by the Kansas Farm Management Association (KFMA); (2) actuarial information from RMA's 2019 Actuarial Data Master (ADM) (Risk Management Agency [RMA] 2019a); (3) 4-by-4-kilometer gridded daily temperature and precipitation from the PRISM Climate Group at Oregon State University (http://prism.oregonstate.edu) and Professor Wolfram Schlenker at Columbia University (http://www.columbia.edu/~ws2162/links.html); (4) 2008-2019 30-by-30-meter gridded national cropland data layers (CDLs) from NASS CropScape (USDA National Agricultural Statistics Service 2019); (5) soil moisture content from state-of-the-art land surface model (LSM) data from the North American Land Data Assimilation System Phase 2 (NLDAS-2); and (6) 1979-2019 weekly crop progress reports from NASS Quick Stats (United States Department of Agriculture [USDA] 2019). [24,25]

---

[24] The NLDAS-2 used in this study were acquired as part of the mission of NASA's Earth Science Division and archived and distributed by the Goddard Earth Sciences (GES) Data and Information Services Center (DISC).

[25] All data and models were processed on Beocat, a High-Performance Computing cluster at Kansas State University (https://beocat.ksu.edu/)

3.4.2 Yield and APH-based loss experience data

The yield data from KFMA and APH-based insurance actuarial information from RMAs ADM are the same as Chapter 2, thus, the reader is referred to the relevant sections in Chapter 2 for extensive details. To summarize, The KFMA data has 5,428 sample farms with 149,267 yield observations spanning 1973-2018 that exhibits a great amount of cross-sectional and temporal variation. The mean of the yields for the four crops considered in kg/ha is 4,833, 3,412, 1,928, and 2,470 for corn, sorghum, soybeans, and wheat, respectively. The respective standard deviation was 2,137, 3,412, 1,928, and 2,470 kg/ha. Chapter 2 showed the representativeness of the KFMA data by comparing sample average yields at the crop-county-year level to yield statistics from NASS. Given the KFMA data, Chapter 2 used a three-step algorithm grounded in RMA guidelines (RMA 2018) to replicate 1,000 simulations of loss experience data for an APH-based insurance scheme for each crop-year in the dataset, starting with 1983. The study utilizes these simulations for the APH-based insurance scheme in this study.

3.4.3 Index-based insurance data

The basic idea for the Index-based insurance simulation is to match each of the 1,000 simulations from Chapter 2 with a corresponding Index-insurance policy. As indicated in the methods section, the Index-based insurance simulation included indices based on precipitation, soil moisture, and 19 indices for degree-days. The precipitation and temperature for degree-days from the sources cited above were available in a raster map with four-kilometer cell sizes. While the precipitation data was taken as given, the degree-days variables were derived following similar research that used similar data (Schlenker and Roberts 2009; Tack et al. 2015; Shew et al. 2020). The soil moisture from NLDAS-2 were hourly estimates for different soil layers based on three different LSMs in raster maps with 14-kilometer cell sizes. Particularly, this study relies on soil moisture stored in the 0–10 cm soil layer as it has been shown to best predict crop yields when compared to alternatives in the same database (Ortiz-Bobea et al. 2019). Given this choice, the soil moisture stored in the 0–10 cm soil layer was taken as the mean across the three sources (NOAH, SAC, and MOSAIC) available in NLDAS-2.

The final data needed to operationalize the analysis was the relevant months within the growing season to base the Index intervals on. To get these data, the study used the weekly crop progress reports from NASS Quick Stats. For each crop, the study selected only the months and adjoining ones (i.e., previous, and subsequent) for which the reports indicated at least 1% progress for planting and at most 50% for harvesting for the respective crop. Using this method, the relevant Index intervals from 1979-2019 for the crops are corn (Feb/Mar, Mar/Apr, Apr/May, May/Jun, Jun/Jul, Jul/Aug, Aug/Sep, Sep/Oct, Oct/Nov), sorghum (Feb/Mar, Mar/Apr, Apr/May, May/Jun, Jun/Jul, Jul/Aug, Aug/Sep, Sep/Oct, Oct/Nov), soybeans (Apr/May, May/Jun, Jun/Jul, Jul/Aug, Aug/Sep, Sep/Oct, Oct/Nov, Nov/Dec, Dec/Jan), and wheat (Feb/Mar, Mar/Apr, Apr/May, May/Jun, Jun/Jul, Jul/Aug, Aug/Sep, Sep/Oct, Oct/Nov).

Given the above data, the study constructs E^{I}_{rv}, F^{I}_{rv}, and R^{I}_{rvg} for each of the Index variables and interval at the grid level and their respective raster. The study then aggregated the E^{I}_{rv}, F^{I}_{rv}, and R^{I}_{rvg} to the farm level. The aggregation was necessary because the exact location of each farmer's field was unknown; however, they were best approximated using their mailing address following the procedure outlined in Note S1 in appendix A. As noted in Chapter 2, while not ideal because of potential measurement error and attenuation bias, the approach is somewhat representative of the real-world situation in which RMA operates. The specific aggregation involves the following steps:

1. Geocoded farm mailing address to get an approximated spatial coordinate of the farm

2. Created a 1.5-mile radius circular spatial polygon using the spatial coordinate of the farm as its centroid
3. Cropped and masked the grid surface for each Index that overlaps with the circular polygon
4. Cropped and masked the 3m crop frequency layer that overlaps with the circular polygon
5. Extract grid IDs from portions of each masked grid surface that overlap with the masked 3m crop frequency layer.
6. Calculated the weight for each grid ID by counting the number of grids from (5) and then divided that by the total number of grids from (5)
7. The information (E_{rv}^{I}, F_{rv}^{I}, and R_{rvg}^{I}) for each farm is taken as the weighted average of the information in their assigned grid IDs.

3.5. Results

3.5.1 Basis risk

Index-based insurance that protects against excess accumulation in killing-degree-days is associated with relatively low basis risk. Figure 3.1 reports the estimated probabilities based on the mismatch between APH- and Index-based insurance indemnifications. Since the APH-based product is directly tied to actual production, it is taken as the truth. Thus, any deviation from its indemnification pattern can be used as a gauge for basis risk. Given a null hypothesis that a producer is indemnified based on the APH-product, the study focuses on the incorrect rejection of a true null hypothesis, i.e., where the producer is not indemnified by the Index product despite experiencing a loss. The study finds that generally, the Index products simulated have FNP less than 0.5. However, there are some interesting dynamics. First, Index products that protect against reduced accumulation in precipitation or soil moisture are generally associated with relatively low FNP when compared with those that protect against reduced accumulation in growing-degree-days. On the contrary, Index products that protect against reduced accumulation in precipitation or soil moisture are associated with relatively high FNP when compared with those that protect against excess accumulation in killing-degree-days.

Figure 3.2 shows the spatial distribution of the second measure for basis risk (i.e., the extent of the short fall in the indemnity due the farmer). The darker areas in Figure 3.2 shows those areas where the typical farmer received relatively less indemnity from the index product relative to what they should have received from the APH product. Figure 3.2 corroborates the conclusion from the FNP on Figure 3.1 as it can be observed that the index product that protects against excess accumulation in killing-degree-days is associated with a relatively less short fall (0.2 to 1.79%) in the indemnity due the farmer when compared to the other products.

3.5.3 Economic performance

Compared to APH-based products, Index-based insurance that protects against excess accumulation in killing-degree-days generates significant gains in economic rents. Results for the ceded to retained Loss-Ratio ratios across all policies and by crops are shown in Figure 3.3. Focusing on all crops aggregate measure, Index insurance protecting against excess accumulation in killing-degree-days with thresholds of 33 °C produces outcomes that are economically significant when compared to policies based on APH. For all crops together, the ratio was approximately 1.1 indicating that LR was 10% larger among the ceded policies relative to the retained. There is also significant heterogeneity across the four crops with corn being the largest (approximately 1.25) and wheat the smallest (approximately 1). Extreme temperatures have been shown to significantly reduce crop yields, thus it is not surprising that Index-based insurance that protects against their excess is economically significant.

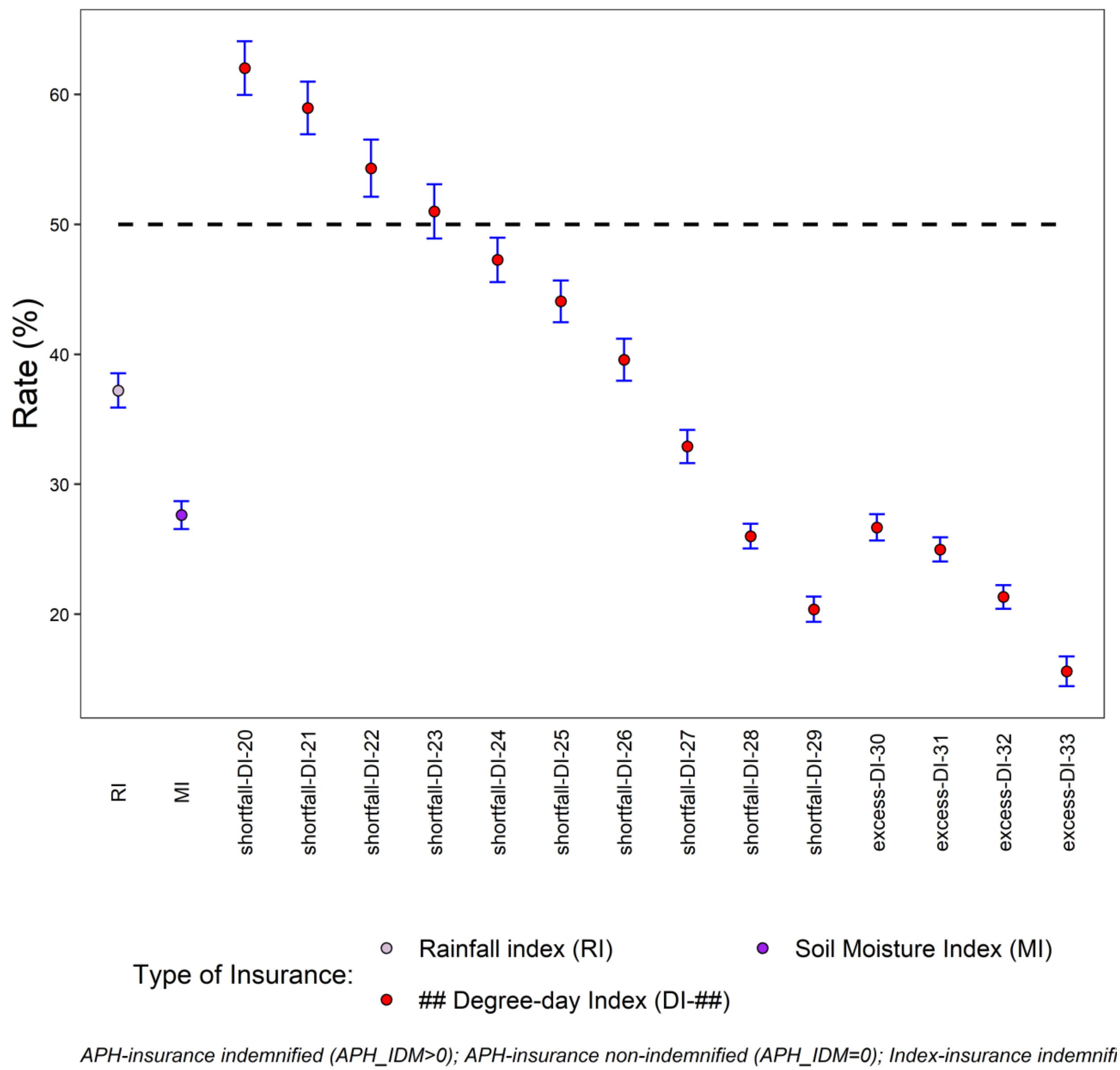


Figure 3.1: False-Negative-Probabilities for Various Index-Based Insurance Products

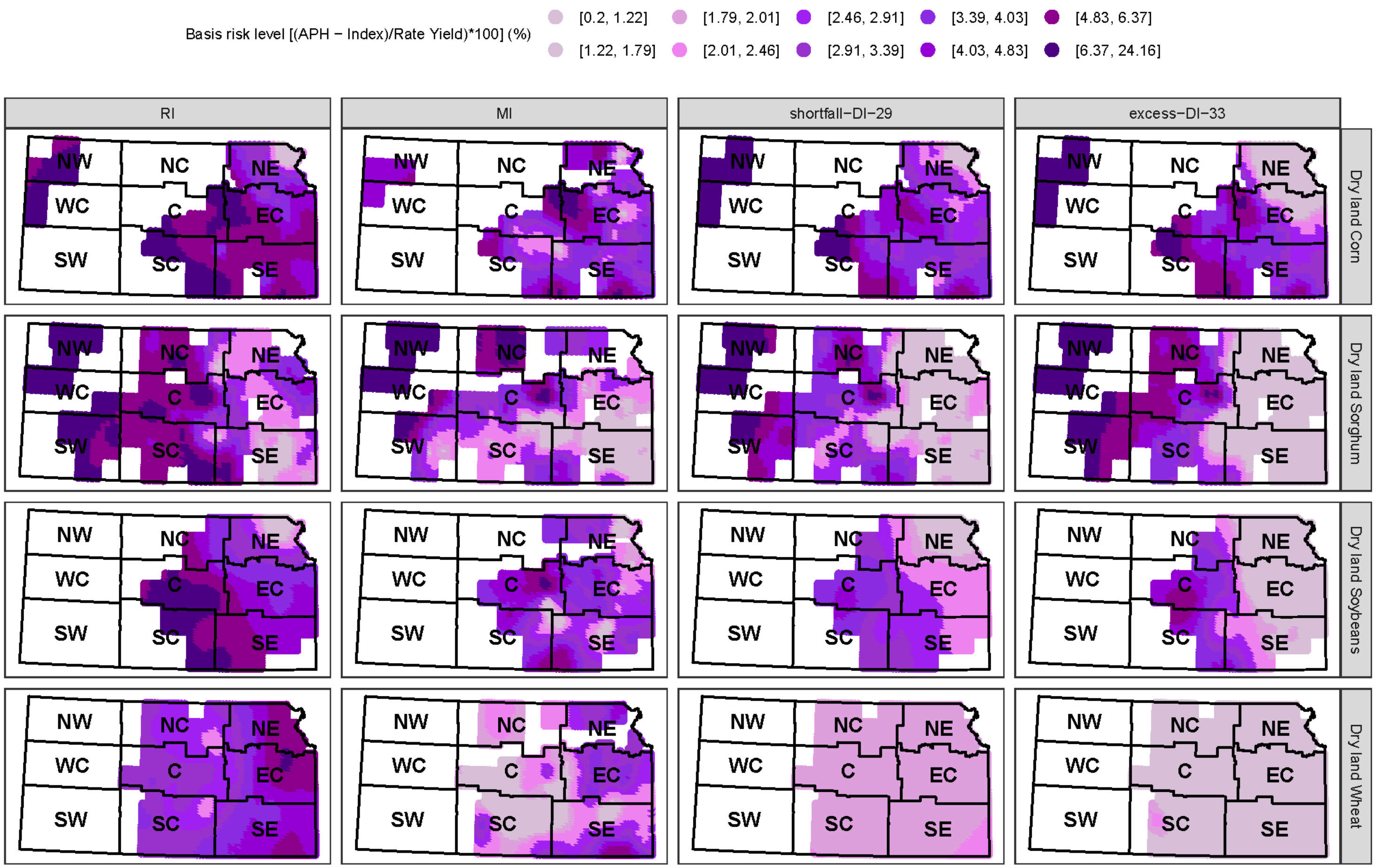


Figure 3.2: Spatial pattern in basis risk for alternative index insurance products in Kansas

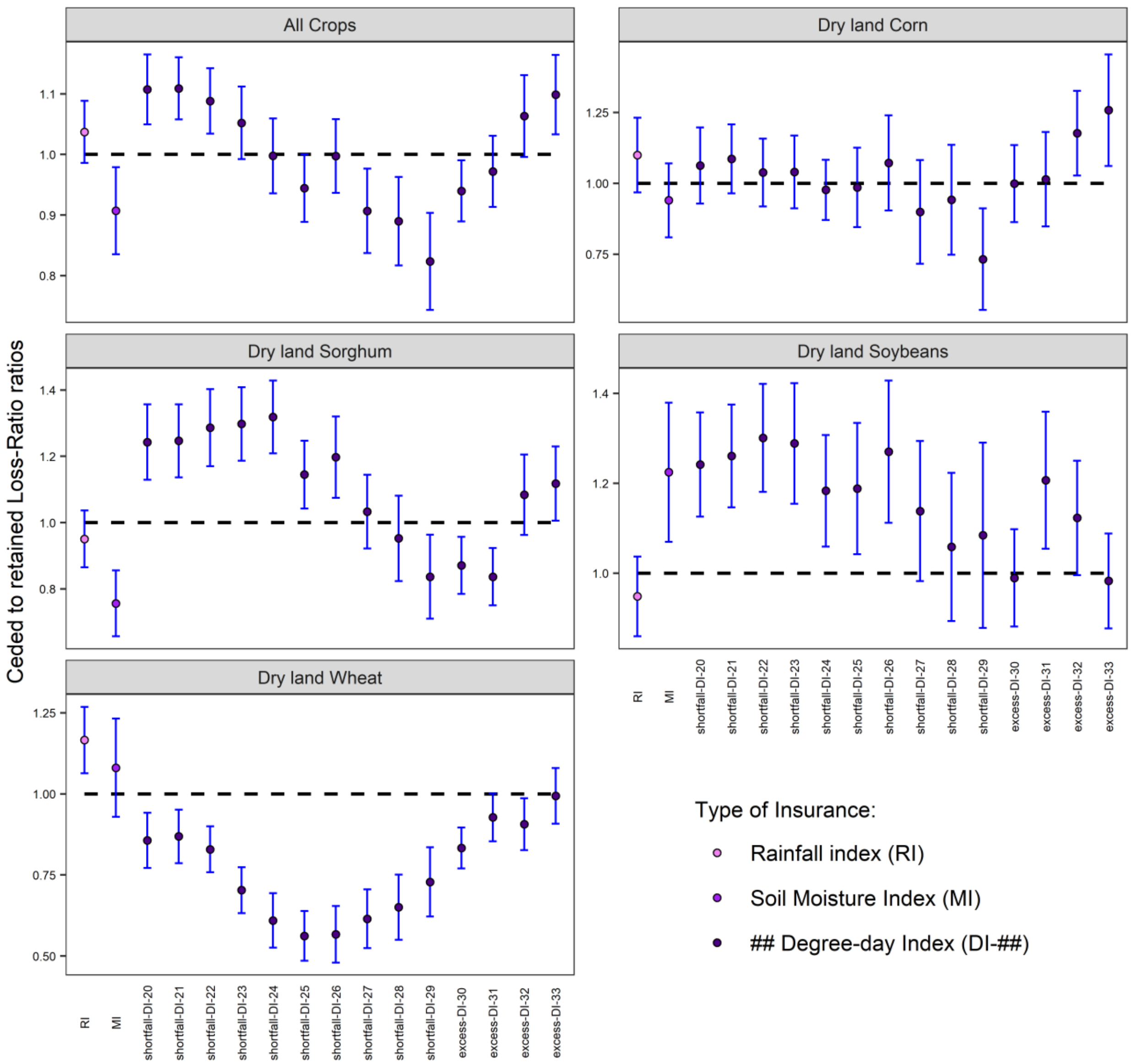


Figure 3.3: Economic Performance of Index-Based Crop Insurance

3.6. Conclusion

This study extends the crop insurance literature by providing simulated evidence of the tradeoffs between APH- and Index-based insurance at the farm level which remains largely unknown. Using a sample of 5,428 corn, soybean, sorghum, and wheat KS farms from 1973-2018, and implementing an empirical strategy that formulates two parallel simulations, the study showed that economically significant tradeoffs do exist between APH- and Index-based insurance and that different types of Index products are associated with differing levels of basis risk. Particularly, Index-based insurance that protects against excess accumulation in killing-degree-days generates the most significant gains in economic rents and is associated with relatively low basis risk. The findings are important given the dual role of a government-led insurance scheme of providing a risk coping mechanism and transferring funds to farmers. The results suggest that where farm-level production data is limited, exploring Index-based insurance that protects against excess accumulation in killing-degree-days can achieve this dual objective.

There are several caveats to the analysis. First, the focus is solely given to dryland operations of Kansas farms that produce corn, soybeans, sorghum, and/or wheat. Subsequent studies can broaden the scope by including more locations, crops, and production practices (e.g., irrigation). Second, the weather information used in this study is based on the mailing address of the farms in the KFMA database. Thus, for those producers who do not live on or close to their farmland, the information could include significant measurement error. Finally, the results rely on premiums and indemnities that are simulated *ex-post* based on historic data. To make this possible a strong assumption made was that all farmers in the historic data participate in both schemes, which had implications for adverse selection and moral hazard. Ongoing work includes efforts to measure the relevance of these concerns.

Reference

Adhikari, S., T.O. Knight, and E.J. Belasco. 2013. "Yield Guarantees and the Producer Welfare Benefits of Crop Insurance." Journal of Agricultural and Resource Economics 38(1):78–92. Available at: https://www.jstor.org/stable/23496740.

Barkley, A., H.H. Peterson, and J. Shroyer. 2010. "Wheat Variety Selection to Maximize Returns and Minimize Risk: An Application of Portfolio Theory." Journal of Agricultural and Applied Economics 42(1):39–55. Available at: https://www.cambridge.org/core/product/identifier/S107407080000328X/type/journal_article.

Barnett, B.J., J.R. Black, Y. Hu, and J.R. Skees. 2005. "Is Area Yield Insurance Competitive with Farm Yield Insurance?" Journal of Agricultural and Resource Economics 30(2):285–301. Available at: https://ageconsearch.umn.edu/record/31216/files/30020285.pdf.

Barnett, B.J., and O. Mahul. 2007. "Weather Index Insurance for Agriculture and Rural Areas in Lower-Income Countries." American Journal of Agricultural Economics 89(5):1241–1247. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8276.2007.01091.x.

Belasco, E.J., J. Cooper, and V. Smith. 2019. "The Development of a Weather-based Crop Disaster Program." American Journal of Agricultural Economics forthcoming paper.

Binswanger-Mkhize, H.P. 2012. "Is There Too Much Hype about Index-based Agricultural Insurance?" Journal of Development Studies 48(2):187–200. Available at: http://www.tandfonline.com/doi/abs/10.1080/00220388.2011.625411.

Bucheli, J., T. Dalhaus, and R. Finger. 2020. "The optimal drought index for designing weather index insurance." European Review of Agricultural Economics. Available at: https://academic.oup.com/erae/advance-article/doi/10.1093/erae/jbaa014/5904966.

Carriker, G.L., J.R. Williams, J. G. Art Barnaby, and J.R. Black. 1991. "Yield and Income Risk Reduction under Alternative Crop Insurance and Disaster Assistance Designs." Western Journal of Agricultural Economics 16(2):238–250. Available at: https://www.jstor.org/stable/40982749.

Carter, M., A. de Janvry, E. Sadoulet, and A. Sarris. 2017. "Index Insurance for Developing Country Agriculture: A Reassessment." Annual Review of Resource Economics 9(1):421–438. Available at: http://www.annualreviews.org/doi/10.1146/annurev-resource-100516-053352.

Chambers, R.G. 1989. "Insurability and Moral Hazard in Agricultural Insurance Markets." American Journal of Agricultural Economics 71(3):604–616. Available at: https://onlinelibrary.wiley.com/doi/abs/10.2307/1242016.

Clarke, D.J. 2016. "A Theory of Rational Demand for Index Insurance." American Economic Journal: Microeconomics 8(1):283–306. Available at: http://pubs.aeaweb.org/doi/10.1257/mic.20140103.

Coble, K., B. Goodwin, M.F. Miller, R. Rejesus, A. Harri, and D. Linton. 2020. "Review of the Pasture, Rangeland, Forage Rainfall Index Crop Insurance Program Indexing and Rating Methodology Final Report." Report to RMA by Sigma Agricultural Risk and Actuarial Services, LLC. Available at: https://www.rma.usda.gov/-/media/RMA/Publications/Risk-Management-Publications/prf_contractor_report.ashx?la=en.

Collier, B., B. Barnett, and J. Skees. 2010. "State of Knowledge Report —Data Requirements for the Design of Weather Index Insurance." Available at: http://globalagrisk.com/Pubs/2010_GlobalAgRisk_State_of_Knowledge_Data_sept.pdf.

Conradt, S., R. Finger, and R. Bokusheva. 2015. "Tailored to the extremes: Quantile regression for index-based insurance contract design." Agricultural Economics 46(4):537–547. Available at: http://doi.wiley.com/10.1111/agec.12180.

Conradt, S., R. Finger, and M. Spörri. 2015. "Flexible weather index-based insurance design." Climate Risk Management 10:106–117. Available at: https://linkinghub.elsevier.com/retrieve/pii/S2212096315000212.

Dalhaus, T., and R. Finger. 2016. "Can Gridded Precipitation Data and Phenological Observations Reduce Basis Risk of Weather Index–Based Insurance?" Weather, Climate, and Society 8(4):409–419. Available at: https://journals.ametsoc.org/doi/10.1175/WCAS-D-16-0020.1.

Dalhaus, T., O. Musshoff, and R. Finger. 2018. "Phenology Information Contributes to Reduce Temporal Basis Risk in Agricultural Weather Index Insurance." Scientific Reports 8(1):46. Available at: http://www.nature.com/articles/s41598-017-18656-5.

Deng, X., B.J. Barnett, and D. V. Vedenov. 2007. "Is There a Viable Market for Area-Based Crop Insurance?" American Journal of Agricultural Economics 89(2):508–519. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8276.2007.00975.x.

Elabed, G., M.F. Bellemare, M.R. Carter, and C. Guirkinger. 2013. "Managing basis risk with multiscale index insurance." Agricultural Economics 44(4–5):419–431. Available at: http://doi.wiley.com/10.1111/agec.12025.

Giné, X., R. Townsend, and J. Vickery. 2008. "Patterns of Rainfall Insurance Participation in Rural India." The World Bank Economic Review 22(3):539–566. Available at: https://academic.oup.com/wber/article/22/3/539/1662754.

Glauber, J.W. 2013. "The Growth Of The Federal Crop Insurance Program, 1990–2011." American Journal of Agricultural Economics 95(2):482–488. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1093/ajae/aas091.

Horowitz, J.K., and E. Lichtenberg. 1993. "Insurance, Moral Hazard, and Chemical Use in Agriculture." American Journal of Agricultural Economics 75(4):926–935. Available at: https://onlinelibrary.wiley.com/doi/abs/10.2307/1243980.

Jensen, N., and C. Barrett. 2017. "Agricultural Index Insurance for Development." Applied Economic Perspectives and Policy 39(2):199–219. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1093/aepp/ppw022.

Jensen, N.D., C.B. Barrett, and A.G. Mude. 2016. "Index Insurance Quality and Basis Risk: Evidence from Northern Kenya." American Journal of Agricultural Economics 98(5):1450–1469. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1093/ajae/aaw046.

Mahul, O., and C.J. Stutley. 2010. Government Support to Agricultural Insurance. The World Bank. Available at: http://elibrary.worldbank.org/doi/book/10.1596/978-0-8213-8217-2 [Accessed June 4, 2020].

Markowitz, H. 1952. "Portfolio Selection." The Journal of Finance 7(1):77. Available at: https://www.jstor.org/stable/2975974?origin=crossref.

Miranda, M.J. 1991. "Area-Yield Crop Insurance Reconsidered." American Journal of Agricultural Economics 73(2):233–242. Available at: https://onlinelibrary.wiley.com/doi/abs/10.2307/1242708.

Miranda, M.J., and K. Farrin. 2012. "Index Insurance for Developing Countries." Applied Economic Perspectives and Policy 34(3):391–427. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1093/aepp/pps031.

Nalley, L.L., A. Barkley, B. Watkins, and J. Hignight. 2009. "Enhancing Farm Profitability through Portfolio Analysis: The Case of Spatial Rice Variety Selection." Journal of Agricultural and Applied Economics 41(3):641–652. Available at: https://www.cambridge.org/core/product/identifier/S1074070800003126/type/journal_article.

Ortiz-Bobea, A., H. Wang, C.M. Carrillo, and T.R. Ault. 2019. "Unpacking the climatic drivers of US agricultural yields." Environmental Research Letters 14(6):064003. Available at: https://iopscience.iop.org/article/10.1088/1748-9326/ab1e75.

Park, S., B.K. Goodwin, X. Zheng, and R.M. Rejesus. 2020. "Contract elements, growing conditions, and anomalous claims behaviour in U.S. crop insurance." The Geneva Papers on Risk and Insurance - Issues and Practice 45(1):157–183. Available at: http://link.springer.com/10.1057/s41288-019-00143-9.

Popp, M., and M. Rudstrom. 2000. "Crop enterprise diversification and specialty crops." Agricultural Finance Review 60(1):85–98. Available at: https://www.emerald.com/insight/content/doi/10.1108/00214710080001112/full/html.

Risk Management Agency [RMA]. 2018. "2019 Crop Insurance Handbook." Available at: https://legacy.rma.usda.gov/handbooks/18000/2019/19_18010.pdf [Accessed May 30, 2020].

Risk Management Agency [RMA]. 2019a. "ACTUARIAL RELEASE: Actuarial Release–19-044." Available at: https://www.rma.usda.gov/Policy-and-Procedure/Bulletins-and-Memos/2019/Actuarial-19-044 [Accessed May 30, 2020].

Risk Management Agency [RMA]. 2019b. "Crop Year Government Cost of Federal Crop Insurance Program." Available at: https://www.rma.usda.gov/-/media/RMAweb/AboutRMA/Program-Budget/18cygovcost.ashx?la=en [Accessed May 19, 2019].

Risk Management Agency [RMA]. 2000. "Premium Rate Calculations for the Continuous Rating Model." Available at: http://www.rma.usda.gov/ftp/Continuous_Rating/cr_instr.pdf [Accessed May 30, 2020].

Risk Management Agency [RMA]. 2020. "Summary of Business Data." Available at: https://prodwebnlb.rma.usda.gov/apps/SummaryOfBusiness/ [Accessed June 12, 2020].

Schlenker, W., and M.J. Roberts. 2009. "Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change." Proceedings of the National Academy of Sciences of the United States of America.

Shew, A.M., J.B. Tack, L.L. Nalley, and P. Chaminuka. 2020. "Yield reduction under climate warming varies among wheat cultivars in South Africa." Nature Communications 11(1):4408. Available at: http://www.nature.com/articles/s41467-020-18317-8.

Smith, V.H., H.H. Chouinard, and A.E. Baquet. 1994. "Almost Ideal Area Yield Crop Insurance Contracts." Agricultural and Resource Economics Review 23(1):75–83. Available at: https://www.cambridge.org/core/product/identifier/S1068280500000435/type/journal_article.

Smith, V.H., and J.W. Glauber. 2012a. "Agricultural Insurance in Developed Countries: Where Have We Been and Where Are We Going?" Applied Economic Perspectives and Policy 34(3):363–390. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1093/aepp/pps029 [Accessed June 18, 2019].

Smith, V.H., and J.W. Glauber. 2012b. "Agricultural Insurance in Developed Countries: Where Have We Been and Where Are We Going?" Applied Economic Perspectives and Policy 34(3):363–390. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1093/aepp/pps029.

Smith, V.H., J.W. Glauber, and B.K. Goodwin. 2017. "Time to reform the US federal agricultural insurance program. Agricultural Policy." Agricultural Policy in Disarray Series Report. Washington, D.C.: American Enterprise Institute (AEI). Available at:

http://www.aei.org/wp-content/uploads/2017/10/Time-to-Reform-the-US-Federal-Agricultural-Insurance-Program.pdf.

Stigler, M., and D. Lobell. 2021. "On the benefits of index insurance in US agriculture: a large-scale analysis using satellite data." Stanford University, Center for Food Security and the Environment. Available at: https://raw.githubusercontent.com/MatthieuStigler/MatthieuStigler.github.io/master/docs/cropIns_paper_latest.pdf.

Tack, J., A. Barkley, and L.L. Nalley. 2015. "Effect of warming temperatures on US wheat yields." Proceedings of the National Academy of Sciences 112(22):6931–6936. Available at: http://www.pnas.org/lookup/doi/10.1073/pnas.1415181112.

United States Department of Agriculture [USDA]. 2019. "USDA - National Agricultural Statistics Service, Quick Stats 2.0." Available at: http://quickstats.nass.usda.gov/ [Accessed May 5, 2019].

USDA National Agricultural Statistics Service. 2019. "CropScape - Cropland Data Layer (2019 release)." Available at: https://nassgeodata.gmu.edu/CropScape/.

Vedenov, D. V., and B.J. Barnett. 2004. "Efficiency of weather derivatives as primary crop insurance instruments." Journal of Agricultural and Resource Economics 29(3):387–403. Available at: https://jareonline.org/articles/efficiency-of-weather-derivatives-as-primary-crop-insurance-instruments/ [Accessed March 16, 2021].

Vroege, W., J. Bucheli, T. Dalhaus, M. Hirschi, and R. Finger. 2021. "Insuring crops from space: the potential of satellite-retrieved soil moisture to reduce farmers' drought risk exposure." European Review of Agricultural Economics 48(2):266–314. Available at: https://academic.oup.com/erae/article/48/2/266/6132278.

Vroege, W., T. Dalhaus, and R. Finger. 2019. "Index insurances for grasslands – A review for Europe and North-America." Agricultural Systems 168:101–111. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0308521X18307200.

Yu, J., and N.P. Hendricks. 2020. "Input Use Decisions with Greater Information on Crop Conditions: Implications for Insurance Moral Hazard and the Environment." American Journal of Agricultural Economics 102(3):826–845. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1093/ajae/aaz035.

Yu, J., M. Vandeveer, J.D. Volesky, and K. Harmoney. 2019. "Estimating the Basis Risk of Rainfall Index Insurance for Pasture, Rangeland, and Forage." Journal of Agricultural and Resource Economics 44(1):179–193 ISSN.

Chapter 4 - The Potential Significance of "Big Ag Data" in Corn Futures Markets

4.1 Introduction

Since the 1970s the United States Department of Agriculture (USDA) has published multiple report series that provide agricultural stakeholders with current and expected market conditions, thereby reducing uncertainties about prices and quantities. A few of these reports that are of interest in the context of this study include the: Annual Acreage; Annual Prospective Plantings; Weekly Crop Progress and Condition (CPC); Monthly Grain Stocks; Monthly World Agricultural Supply and Demand Estimates (WASDE); and Monthly Crop Production (CP). These USDA reports rely on statistical survey approaches to collect production and usage data, as such, they are not available in real-time but rather on well-established release dates throughout the year. Due to the evolving priorities of the USDA, and the growth of the private sector in providing relatively low-cost market information and analysis, the debate about the economic significance of the USDA reports has generated extensive research in the academic literature (Gorham 1978; Ying, Chen and Dorfman 2019; McKenzie and Darby 2017; Schaefer, Myers and Koontz 2004; Isengildina, Irwin and Good 2006; Sumner and Mueller 1989). Most of these articles employed some variant of event study methodology to show that amid private-sector information, the USDA reports significantly impacted markets, suggesting that the reports have economic significance.[26]

In recent years, the adoption and use of precision agriculture technologies have increased, generating increased attention to the output data of precision equipment. An annual survey of stakeholders in the precision agriculture technology supply chain in 2015 showed that two of the

[26] Section 4.2 elaborates on the event study methodology.

most adopted technologies between 2013 and 2015 were unmanned aerial vehicles (UAVs) and "Big Data" (Erickson and Widmar 2015). A unique aspect of precision agriculture is the potential of capturing near real-time information such as planted area, and time and level of input application or harvest (yield). According to Sykuta (2016), the near real-time information captured by precision equipment can easily be aggregated to a higher level. This poses an interesting contrast to the USDA report methodology which relies on periodic surveys to estimate production and usage information.

Previous studies have examined the possibility of generating production information – particularly yield – that is equally if not more accurate than those published in the USDA reports. Several of these studies feed weather data into stylized yield models to estimate near real-time end-of-season yield forecasts. However, one that is of interest to this study utilizes a unique dataset of end-of-season farm-level corn yields akin to that generated by precision technologies to simulate aggregated end-of-season yields. Tack et al. (2019a) utilized various strategies that reflect conditions that private-sector aggregators are likely to face when estimating national end-of-season yields from precision technologies. Vis-à-vis USDA final end-of-season yields, Tack et al. (2019a) showed that non-random sampling schemes are associated with biases that can be effectively removed by benchmarking procedures for removing systematic prediction error.

Building on Tack et al. (2019a), this study seeks to answer a simple but important question of whether live-streamed harvest-time yields from precision technologies are potentially economically significant. To answer this question, the study utilizes historic end-of-season farm-level corn yields that approximately represent 83% of US planted acres for 1999-2008 and CPC to construct weekly yield projection as representative of those from live yield monitors. The idea is to utilize the farm-level yield data to represent the population of farm-level US corn yields, and the weekly variation in CPC information on the proportion of annual crop harvested and under various conditions to approximate how the yield population changes throughout the harvest season. Given the simulated live-streamed harvest-time yields, the study then employs event study methodology to test the potential economic significance of live-streamed yield monitor data vis-à-vis USDA reports.

The remainder of Chapter 4 is organized as follows. Section 4.2 provides a review of event study methodologies as it pertains to the importance of USDA reports to participants in futures markets. The methods section follows next in section 4.3 and it outlines the specific event study methodologies utilized and key hypothesis tested. Section 4.4 presents the data used and how the live-streamed harvest-time yields were simulated from them. Finally, Section 4.5 and 4.6 presents the core results and conclusions, respectively.

4.2. Event Study Literature

The event study methodology, introduced by Fama et al. (1969) is used in the accounting and finance disciplines as the standard methodology for testing the null hypothesis of market efficiency, and to examine the impact of some announcement or event on the wealth of the firm's security holders (Binder 1998)[27]. In agricultural economics, researchers have used the event study methodology to ascertain the warning signs of looming food crises (World Food Programming (WFP) and Centre of Research and Studies on Economic Development (CERDI) 2012) and to analyze the impact of food contamination (Li et al. 2010) or market situation (Gorham 1978; Ying et al. 2019; Isengildina et al. 2006) information release on commodity prices and quantities. In the context of this research, the main idea is that, if markets are efficient, the conditional expectation

[27] See Corrado (2011) and Binder (1998) for an extensive review of the event study methodology since Fama et al. (1969).

of the final prices of contracts at maturity should be well represented by futures prices. Thus, spikes in the variability of futures return reflect changes in market participants' expectations of the maturity prices due to news in the USDA reports. Conditional on the contents of the news, and importantly if the news is valuable to market participants, the changes in the futures return can either be positive or negative. Furthermore, if the market is efficient, the reaction to any news in the USDA reports should be instantaneous. Following this idea, the event study methodologies used in analyzing the announcement effects of USDA reports are of three strands.

Regardless of the nature of the analysis, all studies use some measure of future price return. The main measures used are

$$\Delta P_i = \frac{P_{i,d}}{P_{i,d-j}} - 1 \quad (4.1)$$

$$|\Delta P_i| = \left|\frac{P_{i,d}}{P_{i,d-j}} - 1\right| \quad (4.2)$$

$$r_i = 100 \times \left(\frac{\ln P_{i,d}}{\ln P_{i,d-j}}\right) \quad (4.3),$$

where the subscript P_d is the settlement price of commodity i's nearby futures contract on day d. While *j* can take on any value greater than zero, it is naturally set to one. Typically, close-to-open, close-to-close, or open-to-close methods are used in the determination of $P_{i,d}$ and $P_{i,d-1}$. For close-to-open, $P_{i,d-1}$ and $P_{i,d}$ are the closing and opening futures price for day $d-1$, and d, respectively, and for close-to-close, they both represent closing and opening futures price for the respective days. Open-to-close follow a similar nomenclature.

As indicated earlier, if the markets are efficient, the impact of any new information should be reflected instantaneously in futures prices. Thus, for USDA reports released at the end or beginning of the day's trading session, any new information should be incorporated into the price at the beginning of the preceding day's session. Thus, the close-to-open method will be appropriate. However, if reports are released during the trading session, the close-to-close method is appropriate. Some of USDA's reports before 2008 were released during the trading hours, however, subsequent releases after 2008 are typically done at the end of trading hours. Depending on the period of study and type of USDA report analyzed, several studies have used either the close-to-open or close-to-close method.

The first strand of event study methodologies, utilized by the early literature and as a preliminary test for the recent, relies on simple parametric (e.g., t-tests and F-tests) and nonparametric chi-square (e.g., Savage test, Kruskal-Wallis test, and Van der Waerden test) test of difference in measures of future price variability following a report release and that of non-release days. The second strand of event study methodologies used in analyzing the announcement effects of USDA reports utilizes time series regression frameworks. These studies regressed measures of future price variability on a dummy for the release of several types of USDA reports and other control variables. Consequently, the second strand only provides a yes/no answer to whether the USDA reports influence the actions of market participants.

The third strand methodologies also utilize a regression framework. However, unlike the second strand that utilizes a dummy to represent an announcement effect of the reports, the third utilizes a measure of the extent of the surprise in the reports. Furthermore, unlike the second strand that utilizes all data points over their study period, the third makes use of only the data points around the announcement dates. The prototypical framework for the third strand is represented as

$$r_{i,t} = \alpha + x_{i,t}^{e}\gamma + x_{i,t}^{u}\beta + \sum_{j}^{m} x_{i,t-j}^{u}\delta_j + \mu_{i,t} \,,$$

$$t = -k, \ldots, 0, \ldots, +k, \qquad i = 1, \ldots, I, \qquad \text{and} \qquad j < k \qquad (4.4).$$

First, the time index is $t = -k, \ldots, 0, \ldots, +k$, where zero indicates the daytime trading session immediately following the release of an issue (i) of a given report (e.g., for this study CP and/or

WASDE). The release of one issue is taken as one event, hence the event index is i, and it takes on values from 1 to I (I is the total number of issues from the inception of the given report to date). The variable $r_{i,t}$ has the same definition as before, and it could be calculated based on a close-to-open, close-to-close, or open-to-close basis. The variable $x_{i,t}^{e}$, is a vector of expected information known to the market participant at the close of trading day $t-1$; $x_{i,t}^{u}$ is a vector of unanticipated information (the surprise), derived as $x_{i,t}^{a} - x_{i,t}^{e}$, where $x_{i,t}^{a}$ is a vector of announced information in report issue i. Finally, $\mu_{i,t}$ is a stochastic term; and α, γ, β, and δ_j parameters to be estimated.

It follows from rational market expectations that $x_{i,t}^{e} = E\left[x_{i,t}^{a} | \Omega_{i.t-1}\right]$, where $\Omega_{i.t-1}$, is a vector of the information set at the close of trading day $t-1$, such that $x_{i,t}^{u}$ is uncorrelated with $\Omega_{i.t-1}$. Furthermore, it also follows from the efficient market hypothesis that $\gamma = 0$, because $\Omega_{i.t-1}$ will be reflected in prices at the close of trading day $t-1$. Additionally, $\delta_j \neq 0$ will violate the notion that the reaction to any news in the USDA reports should be instantaneous. Consequently, if markets are efficient, the relevant equation for the analysis reduces to

$$r_{i,t} = \alpha + x_{i,t}^{u}\beta + \mu_{i,t}\ , \qquad t = -k, \dots, 0, \dots, +k, \qquad \text{and} \qquad i = 1, \dots, I, \qquad (4.5).$$

Based on the presented framework, the third strand will usually estimate Equation (4.4) and then test the null hypothesis (jointly or individually); $\gamma = 0$ and $\delta_j = 0$. If both are not rejected, they then proceed to estimate Equation (4.5) to ascertain the effect on futures price return after the surprise ($x_{i,t}^{u}$) is realized. If the null hypothesis fails, then Equation (4.5) is used.

4.3 Methods

This study employed all three strands of event study methodologies used in analyzing the announcement effects of USDA reports. However, the preferred model is the methodology that falls under the regressions with a degree of surprise measure (Equations 4.4 and 4.6). The main idea is that if markets are efficient, the conditional expectation of the maturity price of contracts should be well represented by futures prices. Thus, spikes in the variability of futures return reflect changes in market participants' expectations of the final prices due to news. Conditional on this news, and importantly if it is valuable to market participants, the changes in the futures return can either be positive or negative. Furthermore, if the market is efficient, the reaction to the news will be instantaneous. More importantly for this study, if the market reaction to the news from live-streamed and USDA information is non-zero, then this implies that the former can provide useful information beyond what is available in the latter.

Previous studies have taken expected information in Equation (4.4) as $x_{i,t-1}^{a}$ (a naïve assumption) (Lehecka 2014; McKenzie and Darby 2017; Gorham 1978), the average of market analyst expectations (Frank, Garcia and Irwin 2008; Garcia et al. 1997; Colling and Irwin 1990), or the average of proprietary information (Schaefer et al. 2004). In this study, two sources of surprise are used; (1) "public surprise" calculated based on $x_{i,t}^{e} = x_{i,t-1}^{a}$; and (2) "live surprise" calculated based on $x_{i,t}^{e} = x_{i,t}^{l}$, where $x_{i,t}^{l}$ is the weekly simulated live-streamed harvest-time yields akin to live-streamed yield monitor data. The study included up to three days of historic surprise and controlled month of year and the autocorrelation in return variability by including two lags.

Based on the specified in Equation 4.4, this study tests the following hypotheses. If $\beta = 0$ fails with $\delta_j = 0$ and $\gamma = 0$, then futures prices reflect both public and private information and will not react to USDA reports or live-streamed yield data since they do not provide "news" to the market. In this situation, the market exhibits strong form efficiency (Fama 1970). On the other hand, the rejection of $\beta = 0$ can be taken as importance of the respective sources of "news" to the market. More importantly if both β for "public surprise" and "live surprise" is non-zero, then this implies that live-streamed data can provide useful information beyond what is available in USDA reports. Furthermore, if $\beta \neq 0$ is coupled with $\delta_j = 0$ and $\gamma = 0$, then the underlying assumption is that markets exhibit only semi-strong form efficiency, as the "news" provided by the report is instantaneously incorporated into prices.

4.4. Data

Data for USDA reports on corn yields were retrieved from various issues of WASDE and CP, and the daily corn futures prices from the Chicago Board of Trade (CBOT). Rather than simply dropping potentially useful data in low-trading months or leaving the results susceptible to confounding via sparse trading, the study utilizes the Adjemian and Irwin (2018) method of generating a composite contract series for corn that chooses each day's trading data from the nearby and harvest contract series, based on highest trading volume. The average for the composite contract open and close price and the daily returns are shown in Figure 4.1. The average composite contract price over the entire sample was \$2.25/kg and the average for 2019 (the last year in the data) was \$3.03/kg. The average close-to-open daily returns over the entire sample were 0.005% and the average for 2019 was -0.003%.

In addition to the price data, the study used corn planted acres and production level (bu) from the Risk Management Agency (RMA) Actual Production History (APH) database spanning from 1999 to 2008 to construct live-streamed data akin to that of yield monitors. The total number of APH observations is about 1.5 million from 156,906 farms in 1,919 counties and 47 states. On average, farm size and yields were estimated at 82.15 hectares and 7,865 kg/ha, respectively.

To construct the live-streamed data, the study assumed that the seasonal productivity for each farm (i) is equal to their end of season (t) yield ($\bar{Y}_{it}$). Secondly, for harvest week w, the study further assumes that, for each farm, the proportion of planted acres (A_{it}^{p}) that are available for harvest is equal to that of the state (s) level statistic (θ_{stw}) published in the weekly CPC. Thus, farm i's harvested acres (A_{istw}^{h}) and quantity (Q_{istw}^{h}) for harvest week w are given by; $A_{istw}^{h} = A_{it}^{p} \times \theta_{stw}$ and $Q_{istw}^{h} = A_{istw}^{h} \times \bar{Y}_{it}$. The variables A_{istw}^{h} and Q_{istw}^{h} are then taken as the weekly live-streamed data from each farm during harvest. Given the live-streamed data, the study utilizes four different non-random aggregation methods similar to those in Tack et al. (2019b) to estimate weekly harvest time live-streamed yields. The methods used, which are extensively discussed in Tack et al. (2019b) are; (1) all simple average; (2) all acreage weighted average; (3) I-state acreage weighted average; and (4) C-belt acreage weighted average. The acreage weights are calculated as $\tau_{istw} = \frac{A_{istw}^{h}}{\sum_{i}^{N} A_{istw}^{h}}$.

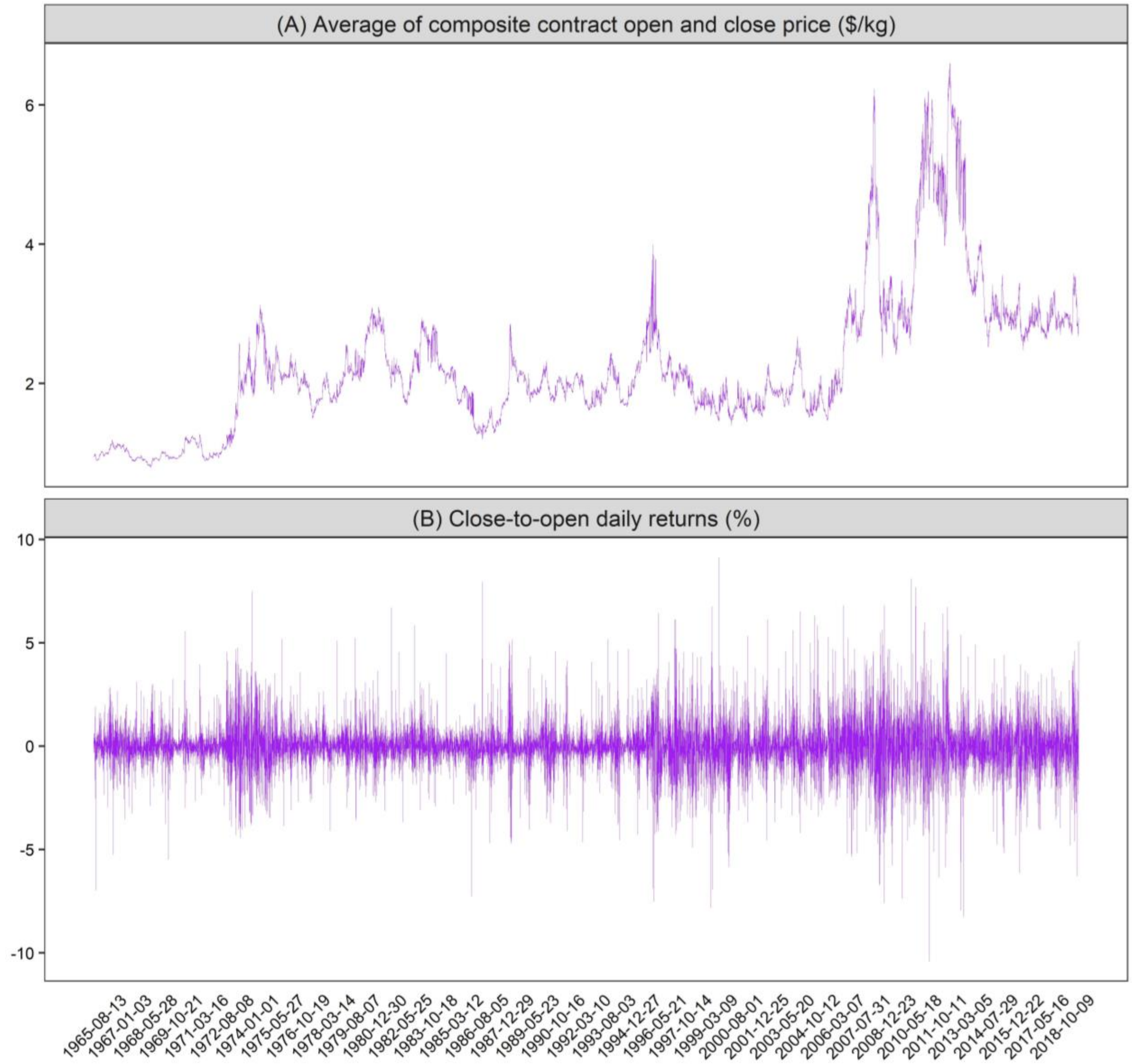


Figure 4.1: Time Series of Daily Returns for Corn Futures Price

Based on initial work by Tack et al. (2019b), non-random sampling schemes are associated with biases which can be effectively removed by benchmarking procedures for removing systematic prediction error. In this spirit, this study utilized two adjustments. The first is based on the long-run relationship

$$\bar{y}_t = \sigma_0 \hat{y}_{tw} + \varepsilon_t \tag{4.7}$$

Where $\bar{y}_t$ is the final yield for season t published by USDA several seasons later, and $\hat{y}_{tw}$ is this study's weekly live-streamed yield estimate. Thus, given the estimate of the long run correction term ($\hat{\sigma}_0$), the benchmarked live-streamed yield estimate is given by $\hat{y}_{tw}^{*} = \hat{\sigma}_0 \hat{y}_{tw}$. For the second benchmarking procedure, the study assumes the correction term is a function of harvest time information available during harvest week w. This was modeled as

$$\bar{y}_t = [\sigma_0 + \sigma_h(1 - \theta_{tw})]\hat{y}_{tw} + \varepsilon_t \tag{4.8}$$

Where θ_{tw} is the proportion of planted acres harvested and is calculated as $\theta_{tw} = \frac{\sum_i^N A_{istw}^h}{\sum_i^N A_{it}^p}$. Given the parameter estimates of Equation 4.8, the benchmarked live-streamed yield estimate is given by $\hat{y}_{tw}^{\#} = [\hat{\sigma}_0 + \hat{\sigma}_h(1 - \theta_{tw})]\hat{y}_{tw}$.

The constructed live-streamed data are shown in Figure 4.2. In most cases, the live-streamed data gives a low forecast as the season starts that increases as more acreage is harvested. Without any adjustments, the live-streamed forecast is higher than the equivalent forecast from USDA. However, after adjustments via benchmarking, the live-streamed forecast converges to the USDA forecast. Furthermore, the similarity between Figure 4.2 Panels B and C indicates that simple benchmarking based on a single long run correction term is robust to complex benchmarking that relies on harvest time information.

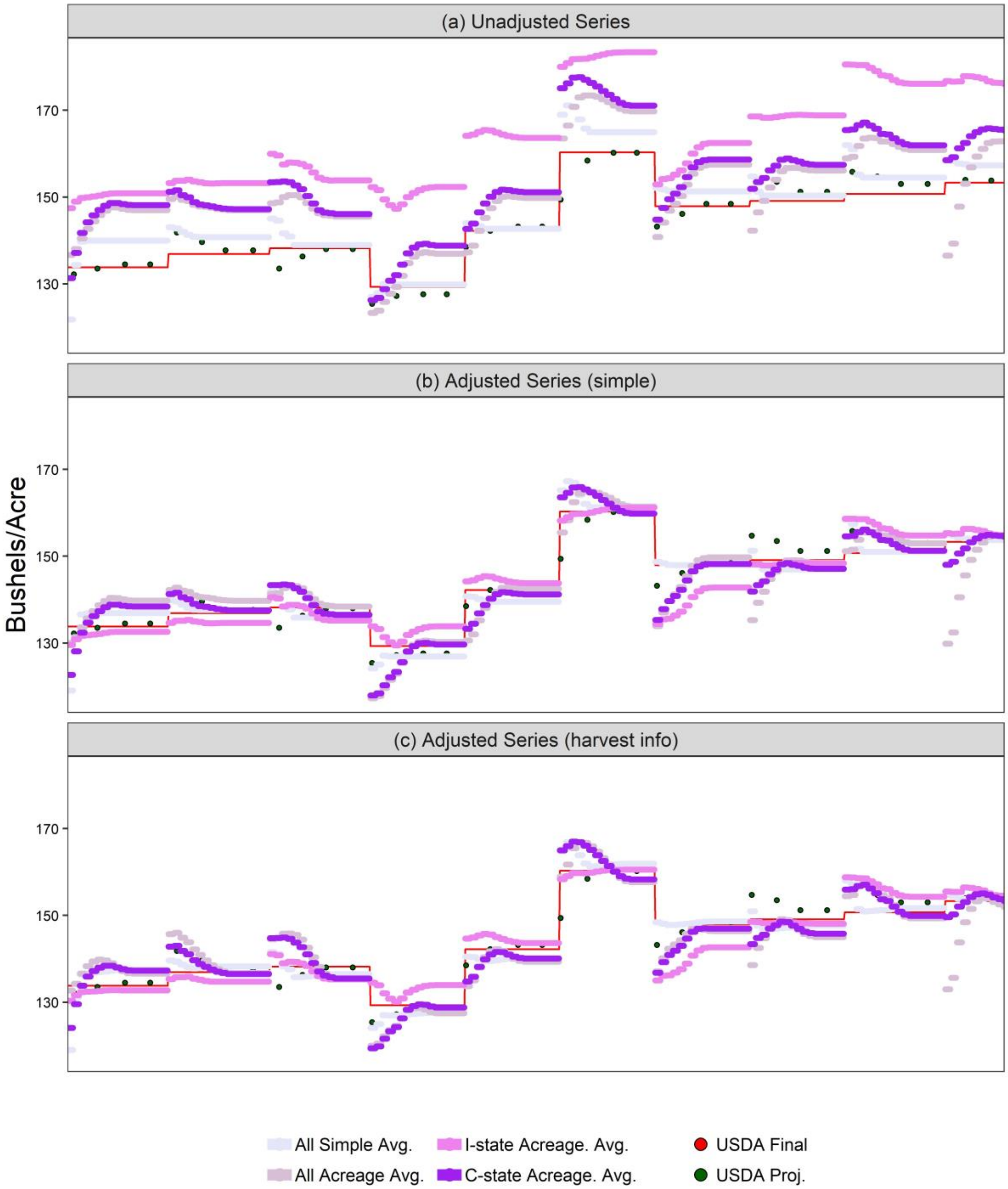


Figure 4.2: Actual and Public/Live Projected Corn Yields

The study used the live-streamed and USDA forecast to construct the level of surprises shown in Figure 4.3. The mean level of public surprise was 0.13% indicating that the yield of information from public sources (i.e., USDA report) was 0.13% higher than expected. Along benchmarking lines, the live-streamed data provided yield information that was 6.51% lower than expected when the forecast was not benchmarked. When benchmarked, the surprise was 0.24 and 0.27% higher when benchmarked with a simple long-run multiplier and harvest time information, respectively. Figure 4.3 shows that generally, the level of surprise from the benchmarked live surprise is the same as that from the public surprise. The unadjusted live surprise is mostly lower than the public surprise.

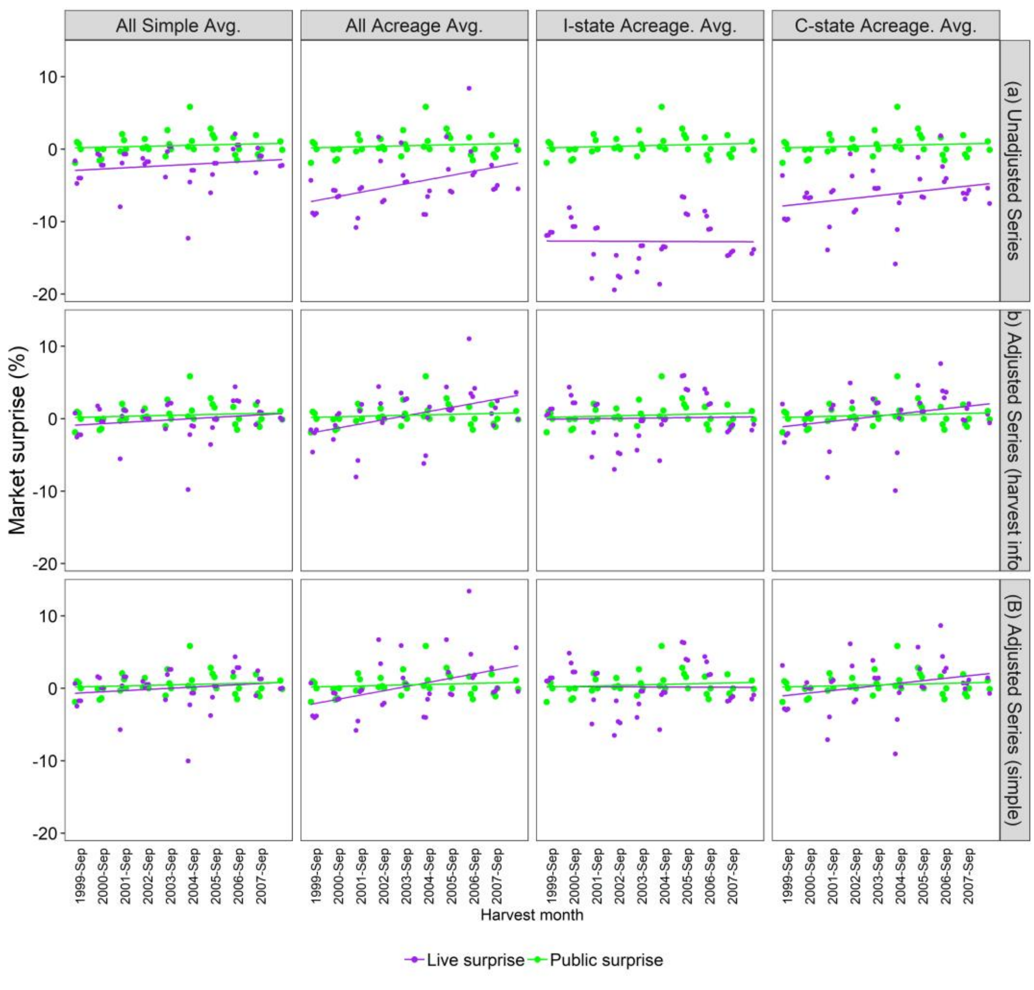


Figure 4.3: Level of market surprise about yield and usage information in USDA Crop Production Reports

4.5. Results

Corn futures market participants react to USDA reports. Results from parametric and nonparametric tests based on the first strand of event study methodology (Table 4.1) and regressions based on the second strand (Table 4.2) suggest that return variability for USDA report release days is significantly ($p<0.05$) different from non-release days. Particularly, the CP and WASDE reports command the greatest return variability. Differences in variability across months also show that the reaction to reports could be influenced by the production cycles. The regression result for the conditional impact of reports on return variability on Table 4.2 also implied that release day variability is 0.16% higher than non-release days.

Under the EMH approach, Table 4.1 shows that the market for corn exhibits only semi-strong form efficiency, as the “news” provided by CP and WASDE is incorporated into prices in at most two days after the release. As expected, an increase in corn yield (a supply-side factor) relative to what was publicly known, elicits a futures price decrease. Given that the mean price was $3.03/kg in 2019 and a 1% unanticipated increase in yield would elicit a 0.45% decrease in futures prices, a decrease of 1.36 cents/kg. This is reasonable as the law of supply dictates price to fall with quantity. In this case, market participants expected a low yield but the yield from a certified public source (USDA) was higher which translates to an increase in supply than they anticipated so they update their willingness to pay (price) downwards. On the other hand, live-streamed yield information does not significantly correlate with historic market reaction.

Table 4.1: Diagnostic Test ON Corn Futures Price Return Reaction to USDA Reports, 1965-2019

Report	N	Variance		Variance homogeneity tests			
		Report day	Non-report day	F test	Levene	Brown and Forsythe	Kruskal-Wallis
Report type							
Any report	2,021	2.22	1.38	0.62***	70.24***	70.66***	45.06***
Prospective Planting (PP)	32	9.51	1.64	0.17***	37.80***	36.61***	13.56***
Acreage Report (AR)	33	9.21	1.80	0.20***	17.19***	16.89***	10.07***
Crop Progress/Condition (CPC)	1,362	1.67	1.53	0.92**	2.20	2.19	1.59
Crop Production (CP)	70	2.19	0.56	0.26***	42.89***	42.34***	28.13***
WASDE	325	2.32	1.19	0.51***	24.85***	24.86***	15.83***
WASDE & CPC	42	1.97	1.43	0.72	1.38	0.75	0.08
WASDE & CP	132	3.92	1.02	0.26***	47.94***	47.85***	29.36***
WASDE, CP & CPC	25	4.58	1.31	0.29***	15.31***	11.14***	19.55***
Release month [reports]							
January [WASDE]	45	7.06	1.49	0.21***	36.02***	34.00***	24.51***
February [WASDE]	40	1.28	0.67	0.52***	4.72**	3.72*	2.41
March [PP, WASDE]	74	4.28	1.64	0.38***	10.37***	10.53***	3.33*
April [CPC, WASDE]	178	1.46	1.34	0.92	0.44	0.43	1.91
May [CPC, WASDE]	236	2.18	1.61	0.74***	6.96***	6.94***	6.11**
June [AR, CPC, WASDE]	235	3.14	2.21	0.70***	2.72*	2.66	1.35
July [CPC, WASDE]	215	2.94	2.07	0.71***	5.61**	5.62**	4.49**
August [CP, CPC, WASDE]	219	2.94	1.43	0.48***	19.21***	19.14***	9.54***
September [CP, CPC, WASDE]	233	1.23	0.90	0.73***	8.11***	7.94***	9.92***
October [CP, CPC, WASDE]	246	1.85	1.05	0.57***	9.89***	9.64***	3.15*
November [CP, CPC, WASDE]	210	0.98	0.74	0.75***	2.14	1.92	0.20
December [CPC, WASDE]	90	0.67	0.81	1.22	0.02	0.01	1.43
Regime							
1981/85	334	2.45	1.61	0.66***	14.80***	14.54***	10.70***
1986/89	377	1.08	0.68	0.63***	11.73***	11.73***	9.78***
1990/95	173	1.35	0.89	0.66***	4.37**	3.64*	1.44
1995/01	282	1.04	0.66	0.63***	6.92***	6.98***	2.86*
2001/12	304	3.39	2.11	0.62***	15.15***	15.14***	10.50***
2013/18	551	3.07	1.97	0.64***	15.19***	15.31***	7.94***

Significance levels: * p<0.10, ** p<0.05, ***p<0.01

Table 4.2: Corn Futures Price Return Reaction to USDA Reports, 1965-2019

	Return type				
	Close-to-Open	**|Close-to-Open|**	**Close-to-Close**	**Open-to-Close**	**Open-to-Open**
Model parameters					
Constant	0.16**(0.08)	0.01***(0.00)	0.26**(0.10)	0.15(0.12)	0.07(0.11)
Event time	0.00(0.02)	0.00(0.00)	-0.02(0.02)	-0.01(0.03)	0.01(0.02)
Time maturity (ln[weeks])	0.00(0.01)	0.00***(0.00)	0.00(0.01)	0.00(0.01)	0.00(0.01)
Lagged return					
One day	0.05***(0.02)	0.02*(0.01)	0.02*(0.01)	0.28***(0.02)	-0.02(0.01)
Two day	-0.02(0.02)	0.03***(0.01)	-0.02(0.01)	-0.12***(0.02)	-0.03**(0.01)
Three day	-0.01(0.01)	0.00(0.01)	0.00(0.01)	0.04**(0.02)	-0.01(0.01)
Report (base=WASDE)					
Prospective Planting (PP)	-0.01(0.15)	0.00**(0.00)	-0.15(0.16)	-0.32*(0.16)	-0.21(0.16)
Acreage Report (AR)	-0.19(0.14)	0.00(0.00)	-0.24(0.16)	-0.30(0.19)	-0.38**(0.18)
Progress/Condition (CPC)	-0.04(0.05)	0.00(0.00)	-0.11*(0.06)	-0.17**(0.07)	-0.13**(0.06)
Crop Production (CP)	-0.01(0.07)	0.00*(0.00)	0.00(0.09)	0.02(0.11)	0.01(0.10)
WASDE & CPC	0.18*(0.11)	0.00(0.00)	0.24*(0.13)	0.25*(0.14)	0.29**(0.13)
WASDE & CP	-0.02(0.08)	0.00(0.00)	-0.05(0.10)	-0.06(0.11)	-0.02(0.10)
WASDE, CP & CPC	-0.04(0.13)	0.00(0.00)	0.10(0.16)	0.20(0.19)	0.11(0.14)
Effect by Report (base=WASDE)					
Prospective Planting (PP)	0.05(0.07)	0.00***(0.00)	0.05(0.08)	-0.02(0.08)	0.00(0.08)
Acreage Report (AR)	-0.11(0.08)	0.00(0.00)	0.16(0.11)	0.31**(0.13)	0.01(0.11)
Progress/Condition (CPC)	0.02(0.02)	0.00(0.00)	0.03(0.03)	0.04(0.03)	0.03(0.03)
Crop Production (CP)	-0.03(0.03)	0.00(0.00)	0.00(0.04)	0.01(0.05)	-0.01(0.04)
WASDE & CPC	0.06(0.06)	0.00(0.00)	0.08(0.08)	0.07(0.09)	0.08(0.08)
WASDE & CP	-0.02(0.03)	0.00***(0.00)	-0.03(0.05)	-0.01(0.05)	-0.01(0.04)
WASDE, CP & CPC	0.12*(0.07)	0.00(0.00)	0.17(0.12)	0.14(0.16)	0.15(0.14)
Release month (base=Dec)					
Jan	0.01(0.16)	0.00**(0.00)	-0.07(0.17)	0.06(0.17)	0.09(0.18)
Feb	-0.11(0.12)	0.00*(0.00)	-0.19(0.14)	-0.12(0.15)	-0.04(0.15)
Mar	-0.08(0.12)	0.00(0.00)	-0.11(0.15)	0.01(0.16)	0.11(0.15)
Apr	-0.05(0.09)	0.00(0.00)	-0.10(0.11)	0.02(0.12)	0.07(0.12)
May	-0.12(0.09)	0.00**(0.00)	-0.14(0.11)	-0.01(0.12)	0.04(0.11)
Jun	-0.08(0.10)	0.00***(0.00)	-0.11(0.12)	-0.02(0.13)	0.04(0.12)
Jul	-0.06(0.09)	0.00***(0.00)	-0.14(0.11)	-0.04(0.12)	0.03(0.12)
Aug	-0.10(0.07)	0.00***(0.00)	-0.15(0.09)	-0.02(0.11)	0.06(0.10)
Sep	-0.16**(0.07)	0.00(0.00)	-0.18*(0.10)	0.00(0.11)	0.02(0.10)
Oct	-0.04(0.07)	0.00(0.00)	-0.04(0.08)	0.13(0.10)	0.21**(0.10)
Nov	-0.09(0.06)	0.00(0.00)	-0.17**(0.08)	-0.09(0.10)	0.01(0.09)
Regime (base=2013/19)					
1981/85	-0.09*(0.05)	0.00(0.00)	-0.15**(0.06)	-0.10(0.07)	-0.09(0.06)
1986/89	-0.01(0.04)	0.00***(0.00)	-0.11*(0.06)	-0.13*(0.07)	-0.08(0.06)
1990/95	-0.01(0.05)	0.00***(0.00)	-0.01(0.07)	0.05(0.09)	0.03(0.07)
1995/01	-0.06(0.04)	0.00***(0.00)	-0.06(0.06)	-0.03(0.06)	-0.05(0.05)
2001/12	-0.06(0.06)	0.00(0.00)	-0.03(0.06)	0.01(0.07)	-0.05(0.07)
Model diagnostics					
Sample size	8,606	8,606	8,606	8,606	8,606
R-squared (%)	0.74	10.64	0.63	8.79	0.72
Log-likelihood	-14,143.69	28,431.49	-15,895.85	-16,916.44	-16,274.95
Model significance	1.52**	15.51***	1.53**	12.88***	1.57**
AIC	28,359.38	-56,790.99	31,863.71	33,904.88	32,621.90
BIC	28,613.55	-56,536.82	32,117.87	34,159.04	32,876.07

Significance levels: * p<0.10, ** p<0.05, ***p<0.01

Table 4.3: Corn Futures Price Return Reaction to Yield "News" Announced in World Agricultural Supply and Demand Estimates (WASDE) and Crop Production Reports, 1965/19

	All Avg. surprise	All Acr. Avg. surprise	I-state surprise	C-belt surprise
		Unadjusted data		
Public yield surprise reaction				
Day 1	-0.445** (0.171)	-0.712*** (0.180)	-0.707*** (0.165)	-0.616*** (0.161)
Day 2	-	-0.242* (0.139)	-0.256** (0.125)	-0.175 (0.125)
Day 3	-	-0.250* (0.134)	-0.242* (0.140)	-0.167 (0.138)
Public usage surprise reaction				
Day 1	-0.114 (0.369)	-0.013 (0.329)	-0.055 (0.356)	-0.076 (0.348)
Day 2	-	0.102 (0.263)	0.071 (0.250)	0.048 (0.253)
Day 3	-	0.283 (0.208)	0.241 (0.209)	0.229 (0.202)
Live surprise reaction				
Day 1	-	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Day 2	-	0.022 (0.036)	0.009 (0.023)	0.010 (0.033)
Day 3	-	-0.014 (0.037)	0.001 (0.023)	-0.004 (0.035)
Sample size	115	115	115	115
R-squared (%)	46.529	41.088	42.573	43.527
Model significance	4.406***	4.442***	4.838***	4.504***
AIC	375.985	387.130	384.193	382.267
BIC	430.883	442.028	439.092	437.166
EMH test 1	2.848**	2.108*	2.031*	2.520**
		Adjusted data [JBT]		
Public yield surprise reaction				
Day 1	-	-0.697*** (0.190)	-0.693*** (0.149)	-0.588*** (0.161)
Day 2	-	-0.253 (0.157)	-0.263** (0.128)	-0.168 (0.131)
Day 3	-	-0.227 (0.138)	-0.233* (0.138)	-0.141 (0.133)
Public usage surprise reaction	-			
Day 1	-	-0.024 (0.340)	-0.057 (0.382)	-0.089 (0.359)
Day 2	-	0.085 (0.259)	0.059 (0.253)	0.049 (0.252)
Day 3	-	0.293 (0.204)	0.256 (0.212)	0.238 (0.203)
Live surprise reaction	-			
Day 1	-	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Day 2	-	0.021 (0.096)	0.010 (0.083)	-0.017 (0.078)
Day 3	-	-0.037 (0.096)	-0.015 (0.085)	-0.036 (0.076)
Sample size	-	115	115	115
R-squared (%)	-	41.175	42.512	43.561
Model significance	-	4.473***	5.427***	4.484***
AIC	-	386.959	384.316	382.197
BIC	-	441.858	439.215	437.095
EMH test 1	-	2.204*	1.972*	2.482**
		Adjusted data [JFT0]		
Public yield surprise reaction				
Day 1	-	-0.677*** (0.189)	-0.692*** (0.150)	-0.578*** (0.159)
Day 2	-	-0.234 (0.151)	-0.262** (0.128)	-0.160 (0.129)
Day 3	-	-0.214 (0.139)	-0.233* (0.139)	-0.135 (0.135)
Public usage surprise reaction	-			
Day 1	-	-0.042 (0.339)	-0.059 (0.379)	-0.097 (0.358)
Day 2	-	0.084 (0.260)	0.055 (0.251)	0.041 (0.251)
Day 3	-	0.267 (0.205)	0.256 (0.212)	0.224 (0.204)
Live surprise reaction	-			
Day 1	-	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Day 2	-	0.024 (0.085)	0.014 (0.081)	-0.015 (0.068)
Day 3	-	-0.045 (0.086)	-0.018 (0.083)	-0.037 (0.066)
Sample size	-	115	115	115
R-squared (%)	-	41.485	42.593	43.773
Model significance	-	4.604***	5.426***	4.452***
AIC	-	386.352	384.154	381.765
BIC	-	441.251	439.053	436.664
EMH test 1	-	2.224*	1.938*	2.474**

Significance levels: * p<0.10, ** p<0.05, ***p<0.01

4.6. Discussion and Conclusion

The results are interesting as the advent of precision agriculture technologies and its associated revolution of “Big Ag-Data” has left researchers to grapple with how to best use the wealth of information available. Since this information can be aggregated to a higher level in real-time, it poses an interesting question of whether equivalent but periodic information from public sources will remain relevant. To this end, this study utilized advances in event study and yield projection methodologies to test the potential economic significance of simulated live-streamed yield monitor data vis-à-vis USDA yields.

The results support the narrative that corn markets in the US exhibit only semi-strong form efficiency. This implies that “news” accompanying the arrival of a report is incorporated into prices immediately. This conclusion corroborates that of Gorham (1978), where the study showed that corn and, to a lesser extent, wheat reports had significant announcement effects on close-to-close price returns from 1950-77. Colling and Irwin (1990) also showed that the hog futures market exhibited semi-strong form efficiency. Particularly they showed that close-to-close price returns from 1981-88 (a) do not react to anticipated changes in reported information, (b) reacts rationally to unanticipated changes in reported information, and (c) adjusts within a day to unanticipated information following the release of reports. Using a similar framework as Colling and Irwin (1990), Lehecka (2014) drew similar conclusions for corn and soybean market efficiency and reaction to USDA CPC reports from 1986-2012. McKenzie and Darby (2017) also showed that USDA provides the futures market with important information, which is vital to the price discovery process.

Given the marked market movement elicited by USDA provided information, the results showed that real-time yields akin to that of yield monitors do not correlate with historic market reactions. Perhaps, since USDA report release days are known well in advance, the market moves in anticipation of their release. Thus, one caveat to this study is that since the real-time yields were not available on the days analyzed, these may not necessarily reflect the reactions of market participants. It is reasonable to have that all the reactions in the price data are captured by the USDA reports since there were observed by a large cross-section of the market. An extension and verification of this result will be to use actual data from yield monitors that were available to a subset if not all market participants.

Despite the caveat, this study advances the market-price event-study methodology by utilizing sources of information not previously considered. Second, the study provides policy implications centered around the ongoing debate about the economic significance of USDA reports in the presence of growing information availability in the private sector.

References

Adjemian, M.K., and S.H. Irwin. 2018. "USDA Announcement Effects in Real-Time." American Journal of Agricultural Economics 100(4):1151–1171. Available at: https://academic.oup.com/ajae/article/100/4/1151/4985883.

Binder, J. 1998. "The Event Study Methodology Since 1969." Review of Quantitative Finance and Accounting 11(2):111–137. Available at: http://link.springer.com/article/10.1023/A%3A1008295500105%5Cnhttp://link.springer.com/article/10.1023/A:1008295500105#%5Cnhttp://link.springer.com/content/pdf/10.1023%2FA%3A1008295500105.pdf.

Colling, P.L., and S.H. Irwin. 1990. "The Reaction of Live Hog Futures Prices to USDA 'Hogs and Pigs Reports.'" American Journal of Agricultural Economics 72(1):84. Available at: https://academic.oup.com/ajae/article-lookup/doi/10.2307/1243147.

Corrado, C.J. 2011. "Event studies: A methodology review." Accounting and Finance 51(1):207–234.

Erickson, B., and D.A. Widmar. 2015. "2015 Precision Agricultural Services : Dealership Survey Results, Draft." Purdue University. Available at: https://agribusiness.purdue.edu/wp-content/uploads/2019/08/2015-crop-life-purdue-precision-dealer-survey.pdf.

Fama, E.F. 1970. “Efficient Capital Markets: A Review of Theory and Empirical Work.” The Journal of Finance 25(2):383. Available at: https://www.jstor.org/stable/2325486?origin=crossref.

Fama, E.F., L. Fisher, M.C. Jensen, and R. Roll. 1969. “The Adjustment of Stock Prices to New Information.” International Economic Review 10:1. Available at: http://content.ebscohost.com/ContentServer.asp?T=P&P=AN&K=5711094&S=R&D=bth&EbscoContent=dGJyMNHr7ESeqa84y9fwOLCmr0uepq5Ss6u4Sa+WxWXS&ContentCustomer=dGJyMOzpsEqyqbdOuePfgeyx44Dt6fIA%5Cnhttp://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=5711094&si.

Frank, J., P. Garcia, and S.H. Irwin. 2008. “To What Surprises Do Hog Futures Markets Respond?” Journal of Agricultural and Applied Economics 40(01):73–87. Available at: https://www.cambridge.org/core/product/identifier/S107407080002798X/type/journal_article.

Garcia, P., S.H. Irwin, R.M. Leuthold, and L. Yang. 1997. “The value of public information in commodity futures markets.” Journal of Economic Behavior & Organization 32(4):559–570. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0167268197000139.

Gorham, M. 1978. “Public and Private Sector Information in Agricultural Commodities.” Economic review-Federal reserve bank of San Francisco.:30–38.

Isengildina, O., S.H. Irwin, and D.L. Good. 2006. “The value of USDA situation and outlook information in hog and cattle markets.” Journal of Agricultural and Resource Economics.

Lehecka, G. V. 2014. “The Value of USDA crop progress and condition information: Reactions of corn and soybean futures markets.” Journal of Agricultural and Resource Economics.

Li, Y., J.E. Wailes, A. McKenzie, and M. Thomsen. 2010. “LL601 Contamination and Its Impact on U.S. Rice Prices.” Journal of Agricultural and Applied Economics 42(01):31–38. Available at: http://journals.cambridge.org/abstract_S1074070800003278 [Accessed October 21, 2015].

McKenzie, A.M., and J.L. Darby. 2017. “Information Content of USDA Rice Reports and Price Reactions of Rice Futures.” Agribusiness 33(4):552–568. Available at: http://doi.wiley.com/10.1002/agr.21489.

Schaefer, M.P., R.J. Myers, and S.R. Koontz. 2004. “Rational expectations and market efficiency in the U.S. live cattle futures market: The role of proprietary information.” Journal of Futures Markets 24(5):429–451. Available at: http://doi.wiley.com/10.1002/fut.10124.

Sumner, D.A., and R.A.E. Mueller. 1989. “Are Harvest Forecasts News? USDA Announcements and Futures Market Reactions.” American Journal of Agricultural Economics 71(1):1. Available at: https://academic.oup.com/ajae/article-lookup/doi/10.2307/1241769.

Sykuta, M.E. 2016. “Big Data in Agriculture: Property Rights, Privacy and Competition in Ag Data Services.” International Food and Agribusiness Management Review Special Issue 19(A):58–74.

Tack, J., K.H. Coble, R. Johansson, A. Harri, and B.J. Barnett. 2019a. “The Potential Implications of ‘Big Ag Data’ for USDA Forecasts.” Applied Economic Perspectives and Policy. Available at: https://academic.oup.com/aepp/advance-article/doi/10.1093/aepp/ppy028/5272601.

Tack, J., K.H. Coble, R. Johansson, A. Harri, and B.J. Barnett. 2019b. “The Potential Implications of ‘Big Ag Data’ for USDA Forecasts.” Applied Economic Perspectives and Policy.

World Food Programming (WFP), and Centre of Research and Studies on Economic Development (CERDI). 2012. “The ALert for Price Spike (ALPS).” Available at: http://foodprices.vam.wfp.org/alps.aspx [Accessed October 10, 2016].

Ying, J., Y. Chen, and J.H. Dorfman. 2019. “Flexible Tests for USDA Report Announcement Effects in Futures Markets.” American Journal of Agricultural Economics 101(4):1228–1246. Available at: https://academic.oup.com/ajae/article/101/4/1228/5485173.

Chapter 5 - Overall Conclusion

This dissertation consists of three studies analyzing the potential of utilizing geo-referenced and "Big-Ag" data to improve US agricultural policy from the angle of risk management and farm support. The motivation was that the data revolution has been upon us for many years now, but public policy continues to grapple with how to best use the wealth of information currently at our disposal. The first study in Chapter 2 extends the crop insurance rating literature by incorporating fine-scale topographic and soil information into rating procedures. A novel econometric approach based on RMA's procedures for pricing insurance at the farm level was developed and applied to a sample of 149,267 farm-level observations in Kansas spanning 1973-2018. The results suggest that features do improve the prediction accuracy of yield losses and can, in general, improve rating performance. Interestingly, these improvements are specific to farms with limited yield histories, as there are no improvements for farms with the commonly used yield history of ten years. As the first to document this, Chapter 2 highlights a crucial dimension in the debate surrounding whether RMA should incorporate soil information into their rating procedures as it suggests that the proportion of policies for which ten years of data is available is an important variable in this decision. The methodological contribution of this study can easily be adopted by RMA to annually parameterize the rate multiplier curve along several dimensions (e.g., soil, coverage level, or production practice). Furthermore, while Chapter 2 does not directly assess non-economic gains, soil conditioned rates could help guard against moral hazard, as farmers can easily alter yields and less so their soil. Additionally, tying rates directly to land could have implications for a farm's financial status through land capitalization.

The second study in Chapter 3 focuses on the same crops and ascertains the tradeoffs between Production-History-based and Index-based insurance. Using the same farm level data as Chapter 2 with fine-scale weather data at the grid-day level and implementing an empirical strategy that formulates two parallel simulations, the study showed that economically significant tradeoffs do exist between APH- and Index-based insurance and that different types of Index products are associated with differing levels of basis risk. Particularly, Index-based insurance that protects against excess accumulation in killing-degree-days generates the most significant gains in farm income and economic rents and is associated with relatively low basis risk. The findings are important given the dual role of a government-led insurance scheme of providing a risk coping mechanism and transferring funds to farmers. The results suggest that where farm-level production data is limited, exploring Index-based insurance that protects against excess accumulation in killing-degree-days can achieve this dual objective.

The final study in Chapter 4 utilized advances in event study and yield projection methodologies to evaluate the potential economic significance of simulated live-streamed yield monitor data vis-à-vis USDA report yields. Chapter 4 relies on high frequency daily price data coupled with farm-level yields that constitutes over 80% of US corn planted acres from 1999-2008. The results showed that corn futures market participants react to USDA reports and that live-streamed yield information does not elicit significant market reaction beyond that. Despite this negative result, Chapter 4 advances the market-price event-study methodology by utilizing sources of information not previously considered. Second, Chapter 4 provides policy implications centered around the ongoing debate about the economic significance of USDA reports in the presence of growing information availability in the private sector.

Appendix A - Study 1

Table A.1: Kansas Farm Level Relationship Between Mean Yield and Yield Risk

	Absolute risk		Relative risk		
	SD	**LAPM**	**CV**	**S- LAPM**	**LCR**
Dryland corn					
Correlation with mean yield	0.105	0.184	-0.649	-0.465	-0.629
Elasticity for mean yield	0.015	0.045	-0.360	-0.282	-0.343
Correlation with relative yield	0.144***	0.586***	-0.856***	-0.707***	-0.919***
	(0.045)	(0.124)	(0.045)	(0.062)	(0.049)
Elasticity for relative yield	0.026	0.161	-0.858***	-0.804***	-0.907***
	(0.084)	(0.234)	(0.102)	(0.124)	(0.112)
Dryland sorghum					
Correlation with mean yield	0.126	0.234	-0.628	-0.434	-0.613
Elasticity for mean yield	0.069	0.123	-0.456	-0.323	-0.443
Correlation with relative yield	0.171***	0.706***	-0.829***	-0.647***	-0.841***
	(0.039)	(0.100)	(0.039)	(0.050)	(0.041)
Elasticity for relative yield	0.116*	0.609***	-0.923***	-0.734***	-0.938***
	(0.062)	(0.159)	(0.070)	(0.082)	(0.073)
Dryland soybeans					
Correlation with mean yield	-0.005	0.121	-0.702	-0.585	-0.685
Elasticity for mean yield	-0.041	-0.013	-0.442	-0.408	-0.433
Correlation with relative yield	0.046	0.383***	-0.954***	-0.808***	-0.962***
	(0.036)	(0.084)	(0.036)	(0.042)	(0.038)
Elasticity for relative yield	-0.038	0.018	-0.981***	-0.934***	-0.995***
	(0.064)	(0.151)	(0.083)	(0.085)	(0.086)
Dryland wheat					
Correlation with mean yield	0.245	0.192	-0.424	-0.359	-0.423
Elasticity for mean yield	0.052	0.036	-0.375	-0.318	-0.381
Correlation with relative yield	0.351***	0.644***	-0.649***	-0.678***	-0.669***
	(0.048)	(0.121)	(0.048)	(0.061)	(0.050)
Elasticity for relative yield	0.114	0.124	-0.878***	-0.930***	-0.915***
	(0.078)	(0.194)	(0.078)	(0.097)	(0.080)

Notes: Table shows the relationship between farm-level mean yield ($\bar{y}_i = \sum_{t=1}^{T} y_{it}$) and mean relative yield ($\tilde{y}_i = \sum_{t=1}^{T}\{y_{it}/y_{ct}\}$) and; measures of absolute risk (Standard deviation [SD], Lower absolute partial moment [LAPM]); and relative risk (Coefficient of variation [CV], Standardized Lower absolute partial moment [S-LAPM], and Empirical loss cost ratio [LCR]). The measures were calculated for each farm using their most recent ten years of successive yields (y_{it}) and the reference yield is taken as the ten-years mean county yield (y_{ct}). The SD and CV were calculated using conventional formulae; LAPM is calculated using formulation provided in Antle (2010) and S-LAPM is the mean ($\bar{y}_i$) standardized version of LAPM; LCR is calculated as $\sum_{t=1}^{10}\{\max[0, \bar{y}_i - y_{it}]/\bar{y}_i\}$. Farm level data was provided by the Kansas Farm Management Association (KFMA).

Table A.2: USA County Level Correlation Between Mean Yield and Yield Risk

	Absolute risk		Relative risk		
	SD	**LAPM**	**CV**	**S-LAPM**	**LCR**
Dryland corn					
Correlation with mean yield	-0.528	-0.457	-0.727	-0.649	-0.717
Elasticity for mean yield	-0.520	-0.466	-0.710	-0.632	-0.704
Correlation with relative yield	-1.464***	-4.659***	-2.464***	-3.329***	-2.255***
	(0.174)	(0.517)	(0.174)	(0.259)	(0.165)
Elasticity for relative yield	-1.468***	-4.400***	-2.384***	-3.116***	-2.208***
	(0.175)	(0.530)	(0.181)	(0.272)	(0.169)
Dryland sorghum					
Correlation with mean yield	0.374	0.377	-0.315	-0.049	-0.361
Elasticity for mean yield	0.140	0.092	-0.487	-0.289	-0.479
Correlation with relative yield	0.611***	1.777***	-0.389*	-0.112	-0.470**
	(0.194)	(0.527)	(0.194)	(0.264)	(0.205)
Elasticity for relative yield	0.405*	1.120*	-0.590***	-0.434	-0.631***
	(0.222)	(0.615)	(0.191)	(0.269)	(0.206)
Dryland soybeans					
Correlation with mean yield	-0.269	-0.325	-0.584	-0.548	-0.553
Elasticity for mean yield	-0.302	-0.339	-0.610	-0.565	-0.581
Correlation with relative yield	-0.642***	-2.293***	-1.642***	-2.146***	-1.596***
	(0.157)	(0.443)	(0.157)	(0.221)	(0.165)
Elasticity for relative yield	-0.719***	-2.432***	-1.701***	-2.198***	-1.670***
	(0.153)	(0.434)	(0.151)	(0.216)	(0.159)
Dryland wheat					
Correlation with mean yield	0.584	0.779	-0.051	-0.057	-0.148
Elasticity for mean yield	-0.009	0.180	-0.555	-0.493	-0.527
Correlation with relative yield	0.954**	2.025***	-0.046	0.013	-0.174
	(0.382)	(0.473)	(0.382)	(0.237)	(0.388)
Elasticity for relative yield	0.058	0.820	-0.774	-0.422	-0.748
	(0.664)	(1.081)	(0.416)	(0.270)	(0.439)

Notes: Table shows the relationship between county-level mean yield ($\bar{y}_c = \sum_{t=1}^{T} y_{ct}$) and mean relative yield ($\widetilde{y_c} = \sum_{t=1}^{T}\{y_{ct}/y_{st}\}$) and measures of absolute risk (Standard deviation [SD], Lower absolute partial moment [LAPM]) and relative risk (Coefficient of variation [CV], Standardized Lower absolute partial moment [S-LAPM], and Empirical loss cost ratio [LCR]). The measures were calculated for each county using NASS Quick Stats data for 2008-2017 and the reference yield is taken as state mean yield (y_{st}) for the same period. The SD and CV were calculated using conventional formulae; LAPM is calculated using formulation provided in Antle (2010) and S-LAPM is the mean ($\bar{y}_i$) standardized version of LAPM; LCR is calculated as $\sum_{t=1}^{10}\{\max[0, \bar{y}_i - y_{it}]/\bar{y}_i\}$.

Table A.3: Example Horizon Thickness Alteration for Root Zone Depth

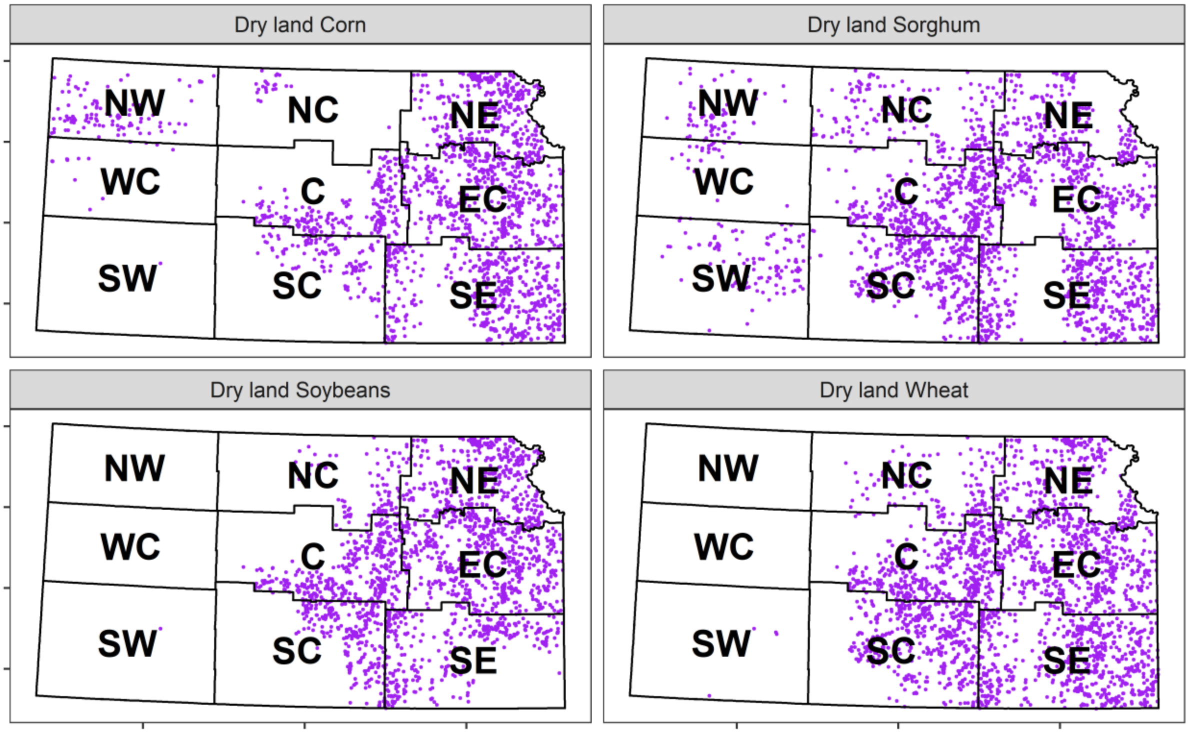
mukey	cokey	chkey	Original information			Altered information		
			Hzdept_r	Hzdepb_r	Hzthk_r	Hzdept_r	Hzdepb_r	Hzthk_r
100017	149541	1	0	25	25	0	25	25
100017	149541	2	25	75	50	25	75	50
100017	149541	3	75	100	25	75	100	25
100017	149542	1	0	50	50	0	50	50
100017	149542	2	50	100	50	50	100	50
100017	149542	3	100	200	100	100	150	50

Hzdept_r is the average distance from the ground surface to the upper boundary of the soil horizon
Hzdepb_r is the average distance from the ground surface to the lower boundary of the soil horizon
Hzthk_r is the average soil horizon thickness (hzdepb_r - hzdept_r)

Table A.4: Table Fields Used in the Calculation of Topographic and Soil Features

Field	Table	Description	Units
Hzdept_r	chorizon	The distance from the top of the soil to the upper boundary of the soil horizon	cm
Hzdepb_r	chorizon	The distance from the top of the soil to the base of the soil horizon.	cm
Hzthk_r	Calculated	A measurement from the top to bottom of a soil horizon throughout its areal extent [hzdepb_r - hzdept_r]	cm
claytotal_r	chorizon	Mineral particles less than 0.002mm in equivalent diameter as a weight percentage of the less than 2.0mm fraction	%
silttotal_r	chorizon	Mineral particles 0.002 to 0.05mm in equivalent diameter as a weight percentage of the less than 2.0mm fraction	%
sandtotal_r	chorizon	Mineral particles 0.05mm to 2.0mm in equivalent diameter as a weight percentage of the less than 2 mm fraction	%
aws0_150	Valu1	Available water storage estimate (aws) in standard zone 5 (0-150 cm depth),	mm
soc0_150	Valu1	Soil organic carbon stock estimate (soc) in standard zone 5 (0-150 cm depth).	g/m^2
rootznemc	Valu1	Root zone depth is the depth within the soil profile that commodity crop roots can effectively extract water and nutrients for growth	cm
nccpi3all	Valu1	National Commodity Crop Productivity Index has the highest value among Corn and Soybeans, Small Grains, or Cotton (weighted average) for major earthy components. Values range from .01 (low productivity) to .99 (high productivity).	index
cec7_r	chorizon	The amount of readily exchangeable cations that can be electrically adsorbed to negative charges in the soil, soil constituent, or other material, at pH 7.0, as estimated by the ammonium acetate method.	meq/100g
slope_r	component	The difference in elevation between two points expressed as a percentage of the distance between those points. (SSM)	%
2mmTot	Calculated	Particles less than 2 mm fraction [claytotal_r + silttotal_r + sandtotal_r]	%
Clay	Calculated	Clay particles as a weight % of the less than 2 mm fraction [claytotal_r / 2mmTot]	ratio
Silt	Calculated	silt particles as a weight % of the less than 2 mm fraction [silttotal_r / 2mmTot]	ratio
Sand	Calculated	sand particles as a weight % of the less than 2 mm fraction [sandtotal_r / 2mmTot]	ratio

Sources: https://www.nrcs.usda.gov/wps/PA_NRCSConsumption/download?cid=stelprdb1241114&ext=pdf
https://www.nrcs.usda.gov/wps/PA_NRCSConsumption/download?cid=stelprdb1241115&ext=pdf
https://www.nrcs.usda.gov/wps/PA_NRCSConsumption/download?cid=nrcseprd1643228&ext=pdf



Source: Constructed by author, using farm data provided by the Kansas Farm Management Association

Figure A.1: Spatial Representation of Farm-Level Data by Crop

Notes: Graph shows the spatial representation of geocoded addresses for farms that produced the specified crops at least in one year from 1973-2018. Farm-level data was provided by the Kansas Farm Management Association (KFMA).

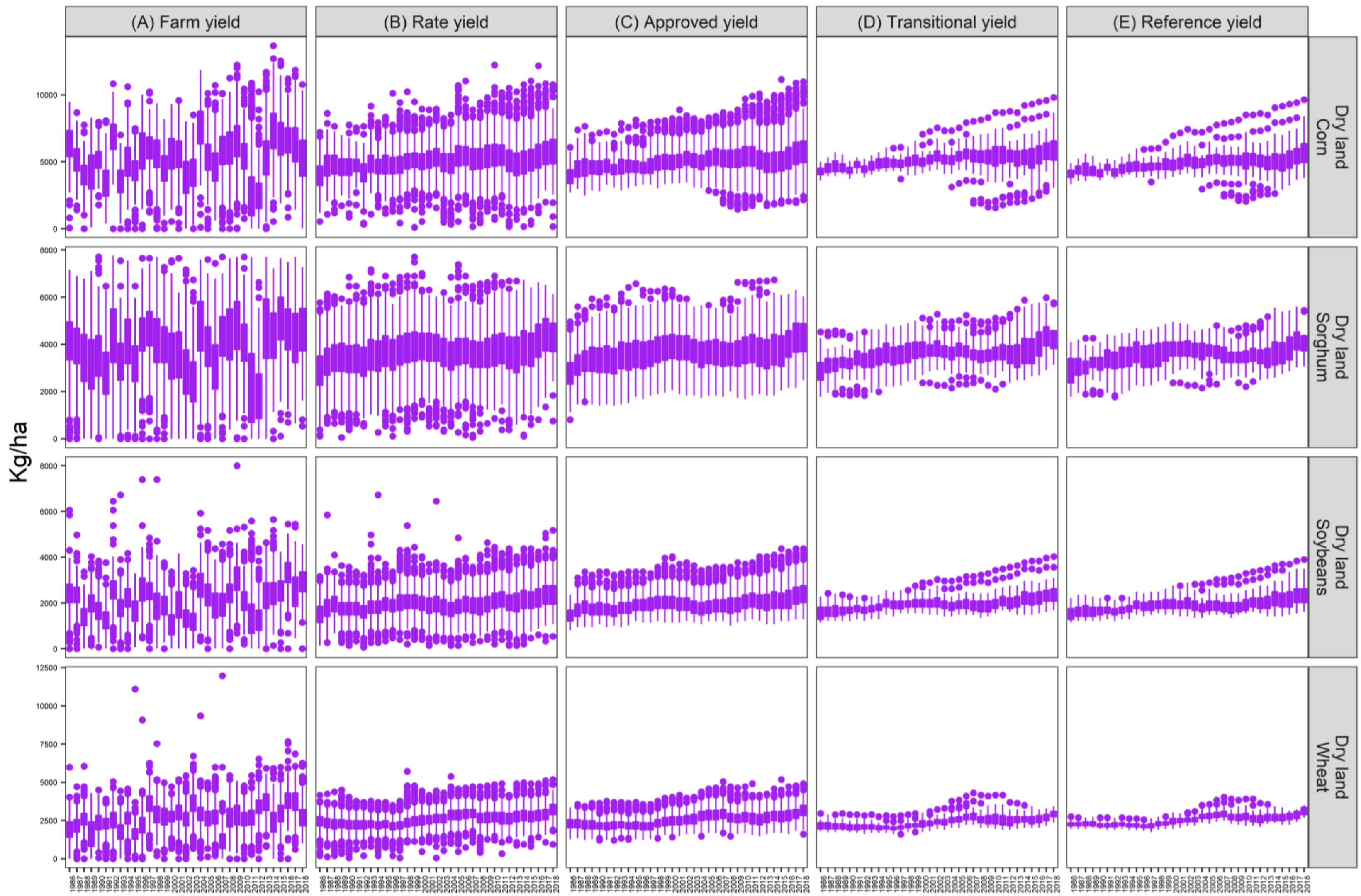


Figure A.2: Boxplots of Yields

Notes: Panel A is from farm-level data provided by the Kansas Farm Management Association (KFMA). Panels B-E is based on the yields in Panel A and following RMA published guidelines.

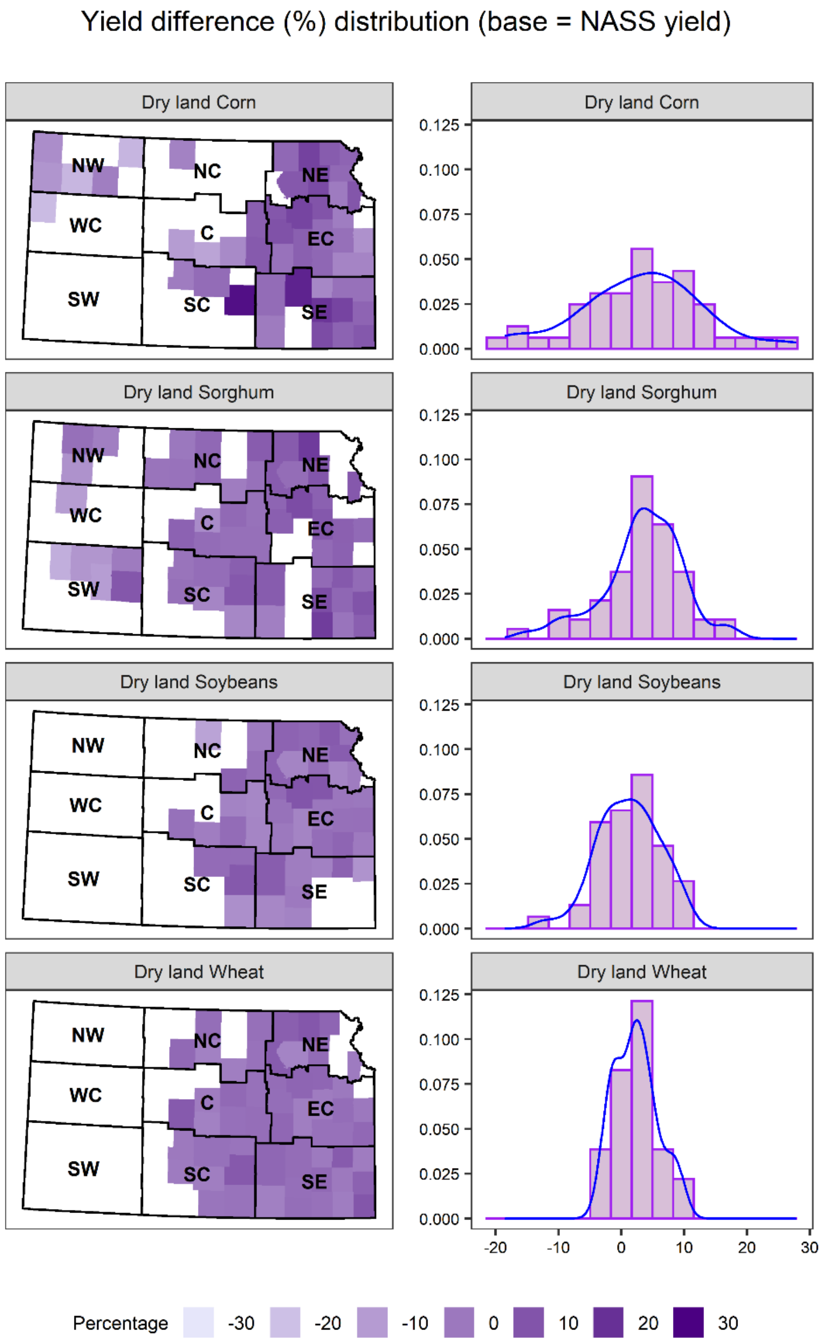


Figure A.3: Representativeness of Kansas Farm Management Association (KFMA) Yields by Crop

Notes: Mean yields for each crop-year-county combination were estimated for the KFMA data. Given these values, the percentage difference for each crop-year-county mean yields relative to their corresponding NASS yield retrieved from NASS Quick Stats was then calculated to make the figure.

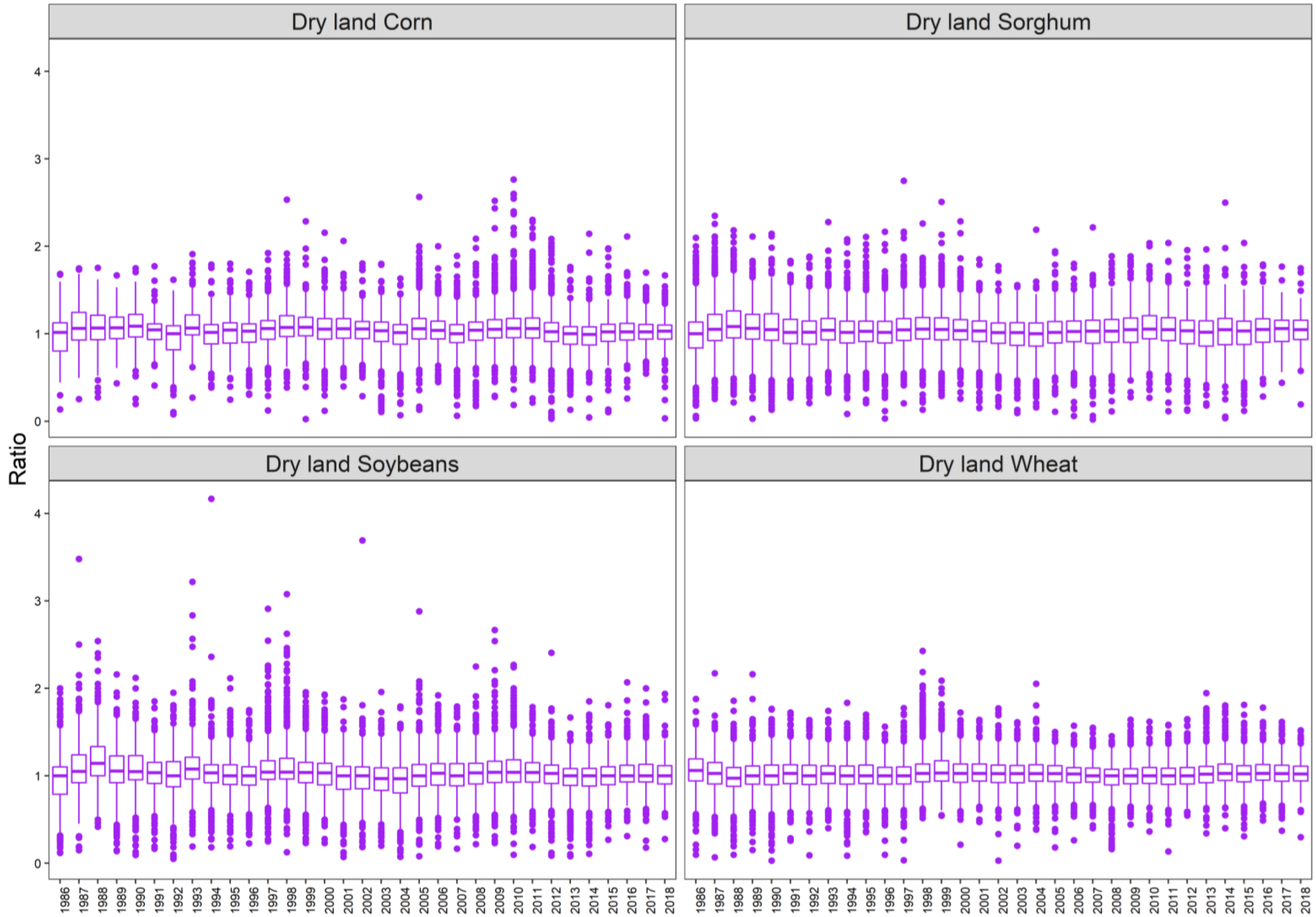


Figure A.4: Relative Yields

Notes: Graph show distribution of the ratio of farm-level rate yields and county level reference yields. The rate yields are the simple average of ten successive years of actual yields and reference yields are the mean of ten successive annual county yields. Farm-level data was provided by the Kansas Farm Management Association (KFMA).

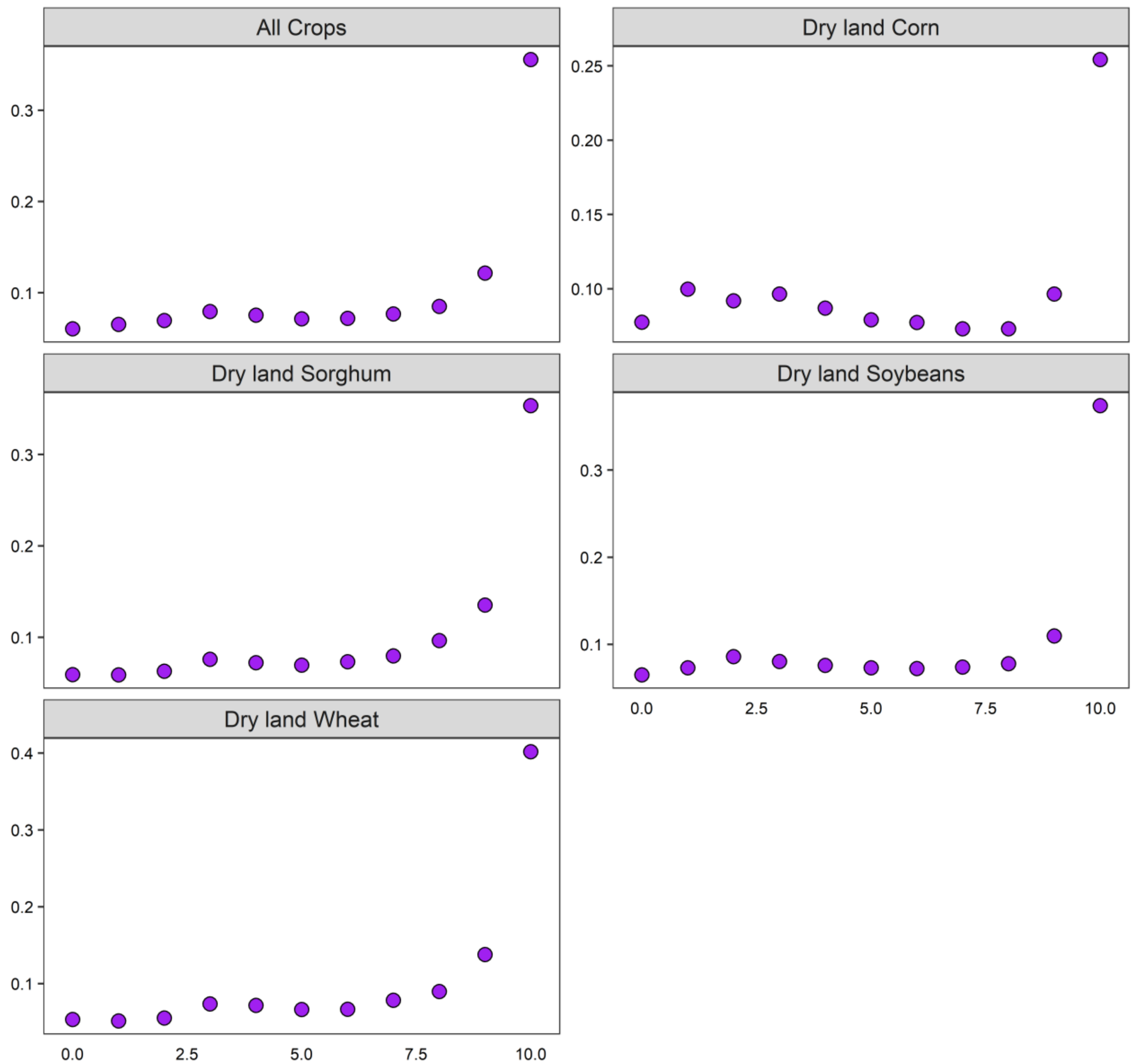


Figure A.5: Sample Distribution by Production History Length

Notes: Figure shows the proportion of the sample that are classified by the length of actual production history in ten years. Farm-level data was provided by the Kansas Farm Management Association (KFMA).

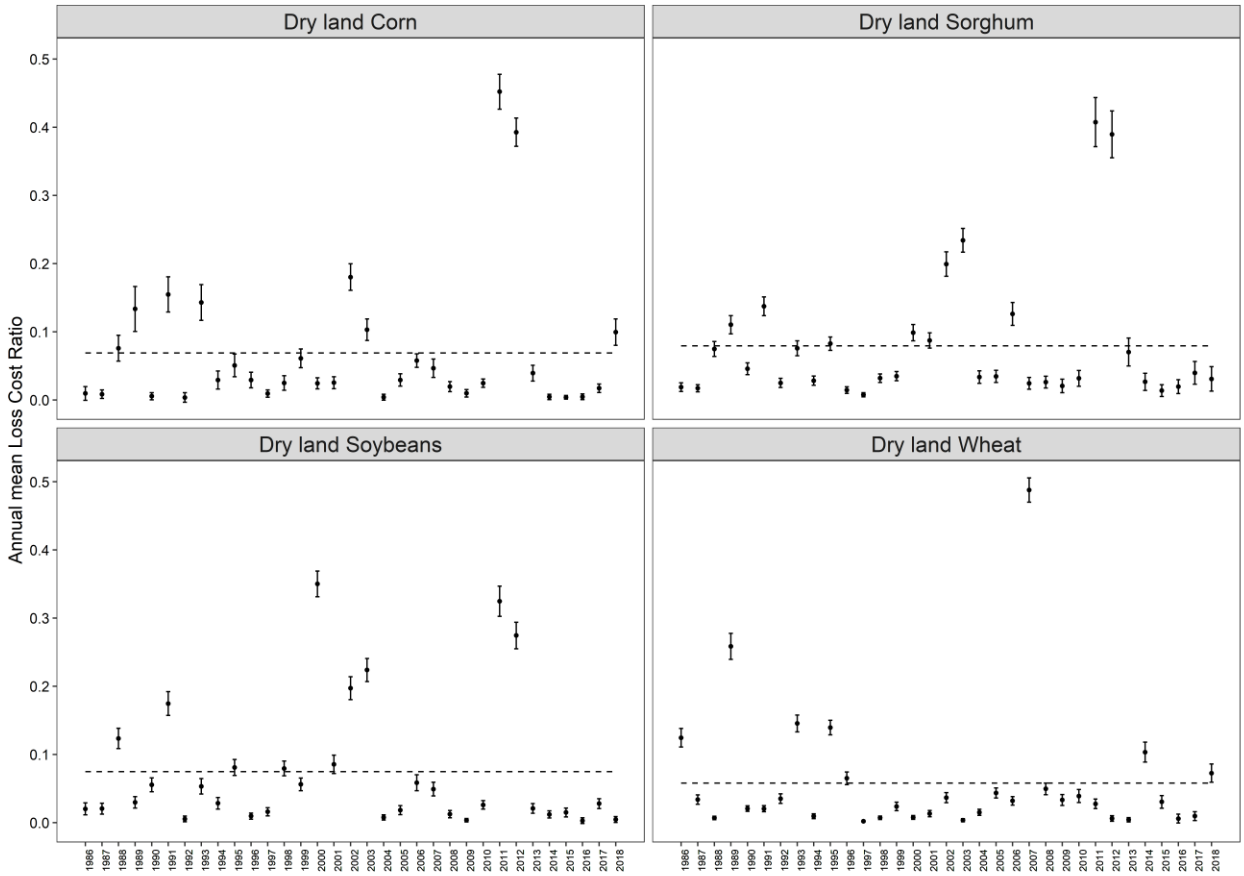


Figure A.6: Mean Empirical Loss Cost Ratios

Notes: For each farm/crop/year, empirical loss cost ratios are calculated as $\sum_{t=1}^{10}\{\max[0, \bar{y}_i - y_{it}]/\bar{y}_i\}$. The horizontal dashed line is the overall mean. Farm-level data was provided by the Kansas Farm Management Association (KFMA).

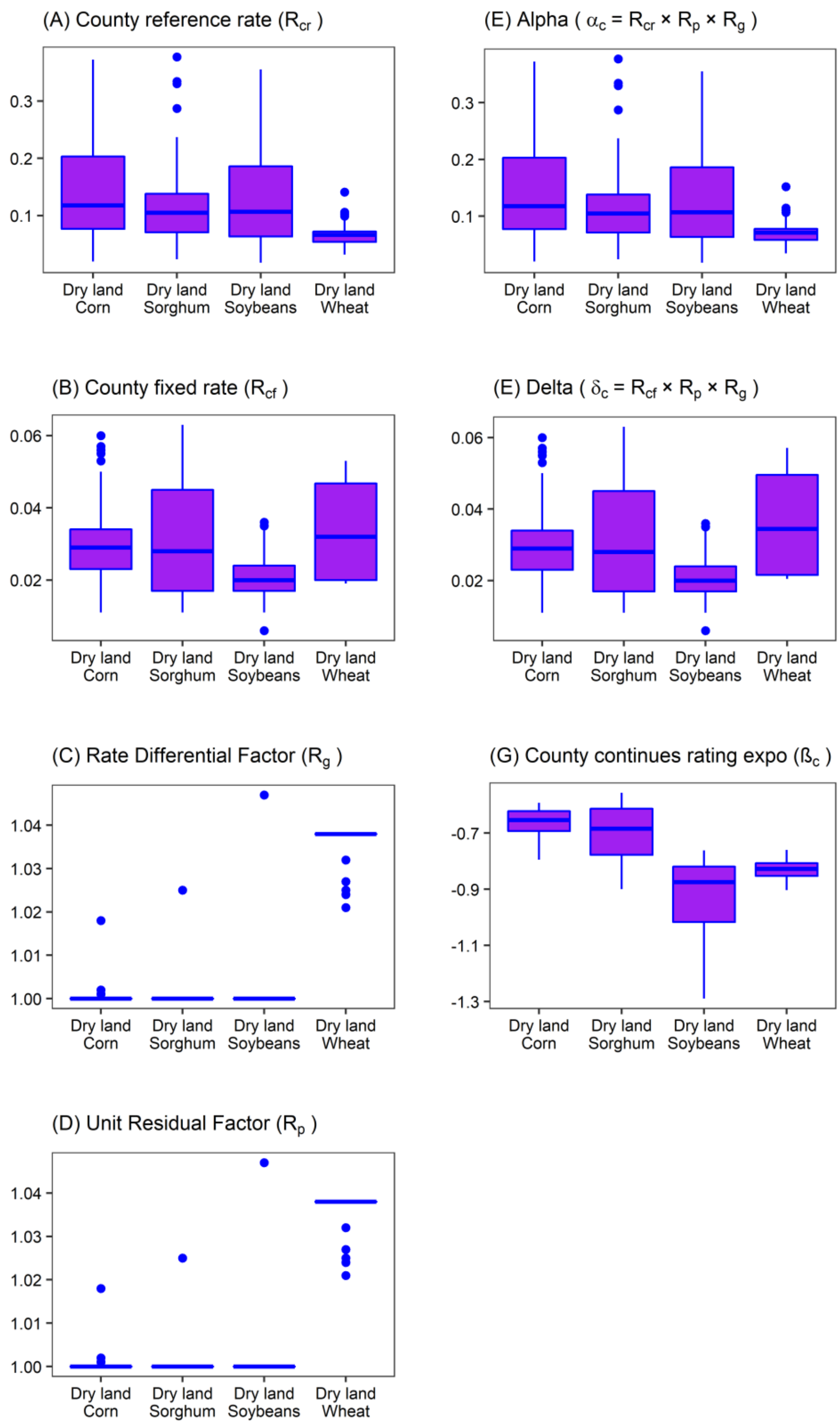


Figure A.7: Boxplot of 2019 Federal Crop Insurance Actuarial Information for Enterprise Unit Dryland Production in Kansas at 75% coverage level

Notes Panel A-E were retrieved from RMA's 2019 Actuarial Data Master found at ftp://ftp.rma.usda.gov/pub/References/actuarial_data_master/2019/

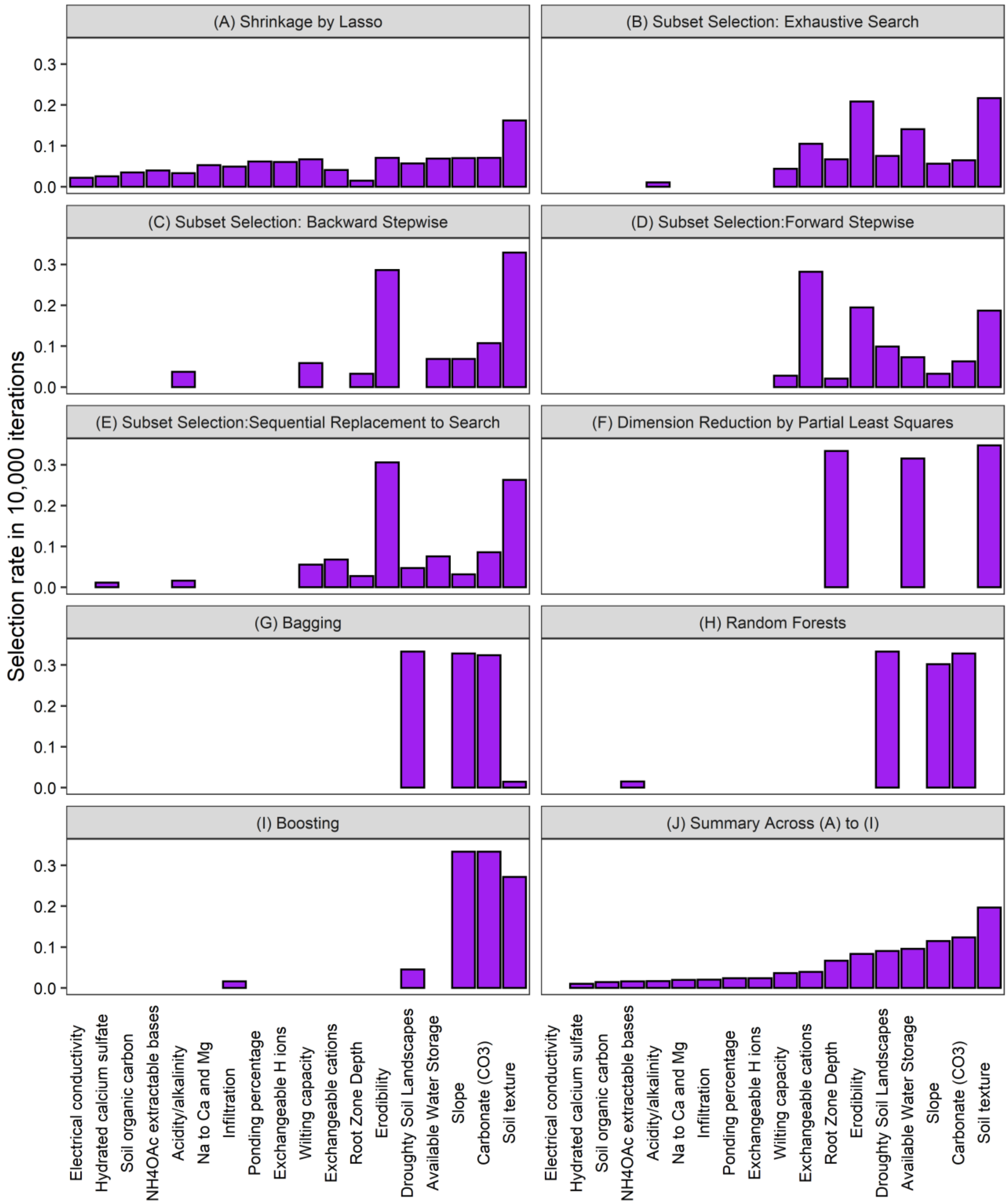


Figure A.8: Statistical Learning Outcome for Soil Attribute Selection

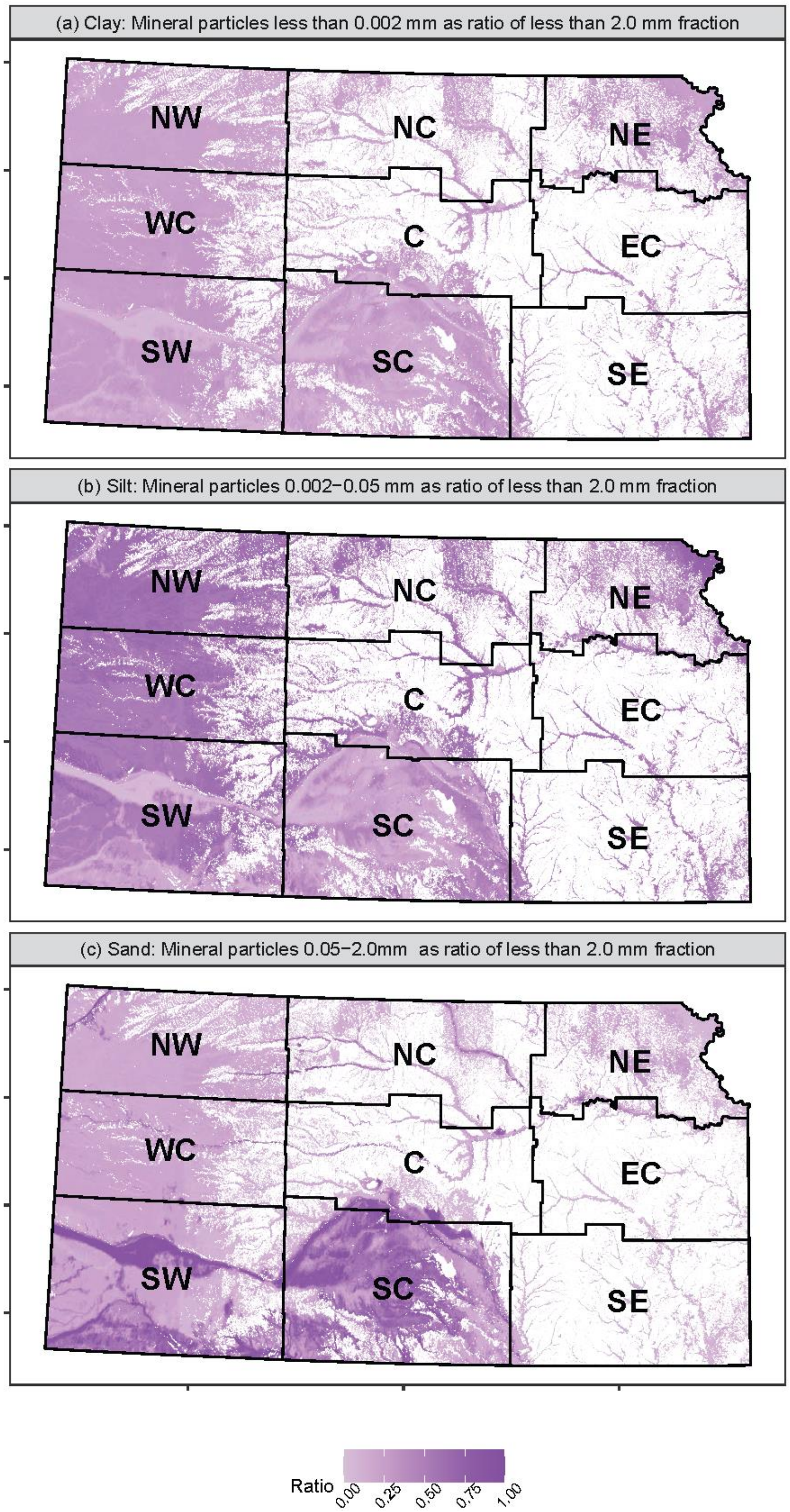


Figure A.9: Map-unit Level Soil Texture Spatial Representation in Kansas

Notes: Constructed by author, using data provided by Gridded Soil Survey Geographic (gSSURGO) Database for the Conterminous United States, available online at https://gdg.sc.egov.usda.gov/

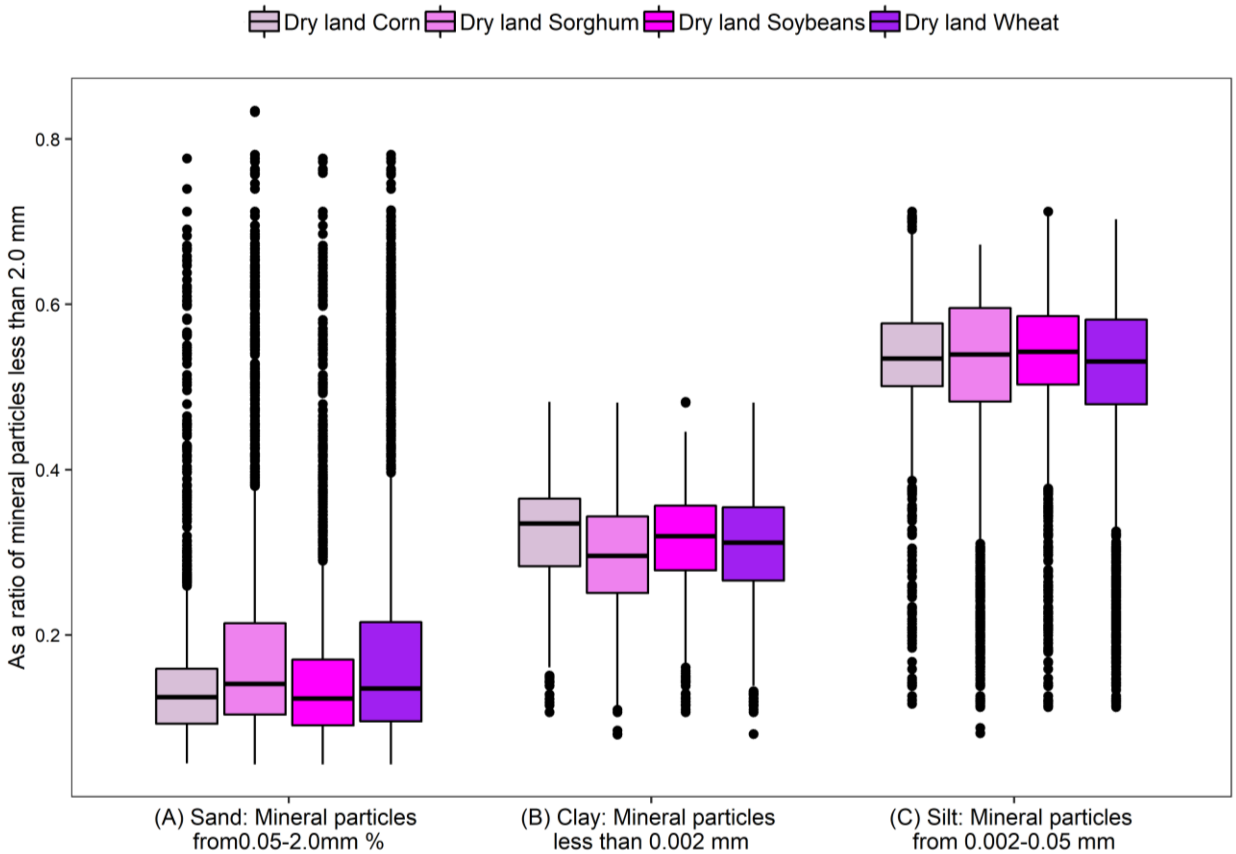


Figure A.10: Farm Level Soil Texture Distribution by Crop

Notes: Constructed by author, using farm data provided by the Kansas Farm Management Association and soil data provided by Gridded Soil Survey Geographic (gSSURGO) Database for the Conterminous United States, available online at https://gdg.sc.egov.usda.gov/.

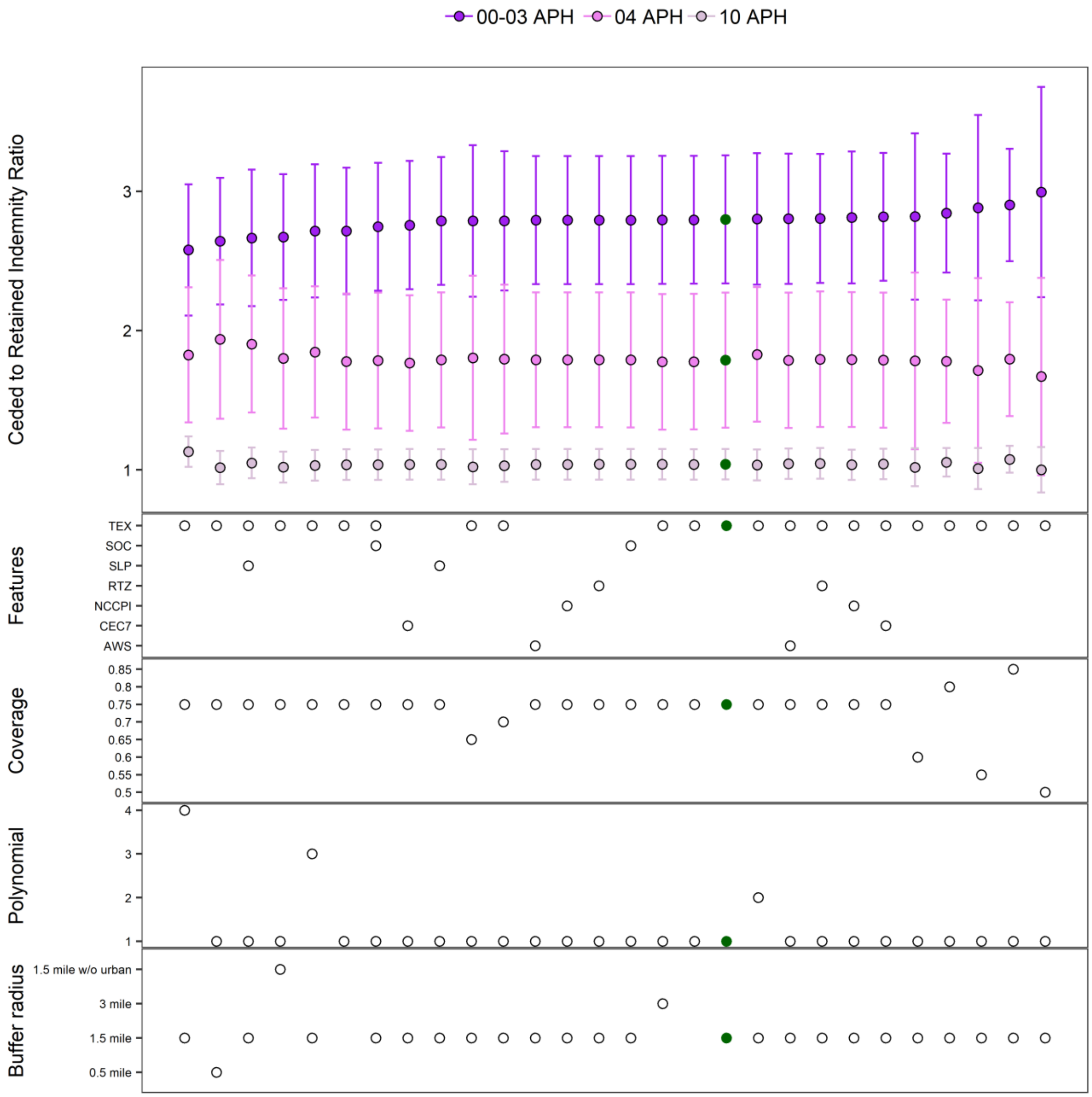


Figure A.11: Robustness of Ceded to Retained Indemnity Ratios Across Model Specifications

Notes: The top panel shows the ceded to retained indemnity ratios for different models defined by the vertical axis of the subsequent panels. The second panel changes the conditioning topographic and soil features for the models; where TEX= soil texture, RTZ= root zone depth, AWS=available water storage, SLP=slope, CEC7=exchangeable cations, and SOC=soil organic carbon. Coverage level specification for loss experience data generation ranged from 50 to 85%; the degree of the polynomial for the conditioning topographic and soil features ranged from 1-4, and the buffer for feature aggregation ranged from 1.5-mile to 10-mile radius. The green-filled dot indicates the position of the preferred model: Linear soil texture model for a coverage level of 75% loss experience data, and feature aggregation using a buffer of a 1.5-mile radius.

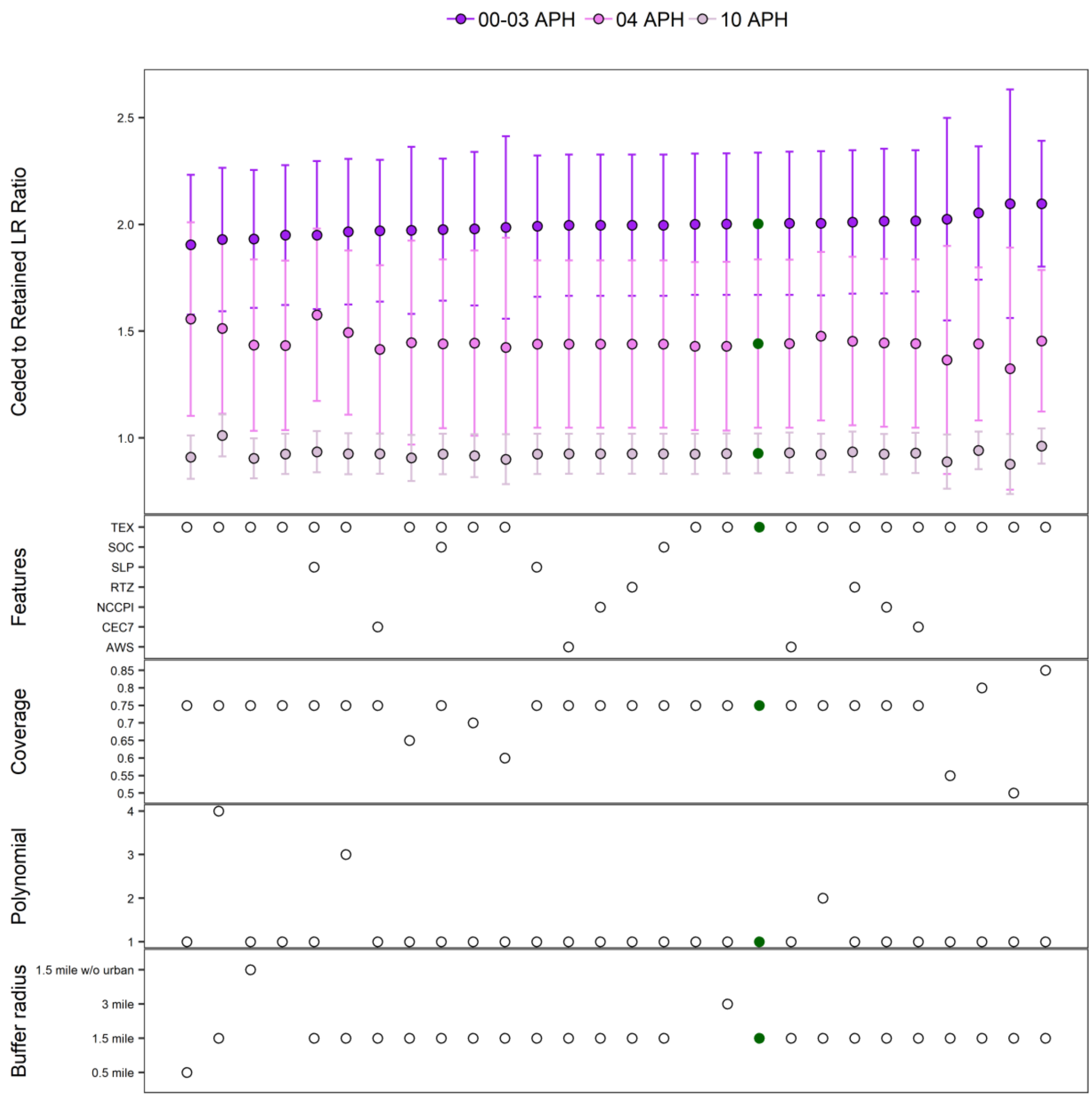


Figure A.12: Robustness of Ceded to Retained LR Ratios Across Model Specifications

Notes: The top panel shows the ceded to retained LR ratios for different models defined by the vertical axis of the subsequent panels. The second panel changes the conditioning topographic and soil features for the models; where TEX= soil texture, RTZ= root zone depth, AWS=available water storage, SLP=slope, CEC7=exchangeable cations, and SOC=soil organic carbon. Coverage level specification for loss experience data generation ranged from 50 to 85%; the degree of the polynomial for the conditioning topographic and soil features ranged from 1-4, and the buffer for feature aggregation ranged from 1.5-mile to 10-mile radius. The green-filled dot indicates the position of the preferred model: Linear soil texture model for a coverage level of 75% loss experience data, and feature aggregation using a buffer of a 1.5-mile radius.

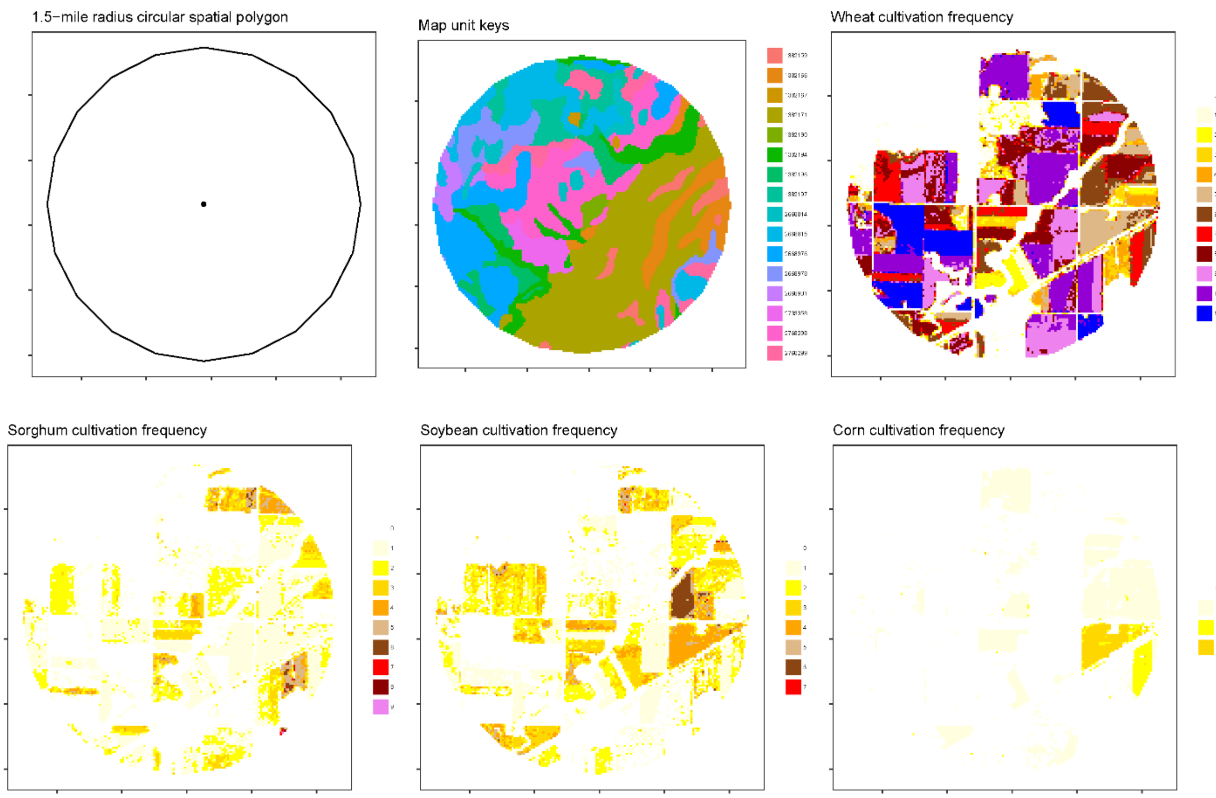


Figure 5.13: Sample of the spatial objects for a given farm

Note A.1: Soil Data Aggregation

Topographic and soil features were determined for each farm using data retrieved from publicly available high-resolution (1:12,000 to 1:63,360) static Soil Survey Geographic (SSURGO) provided by the USDA-NRCS. Particularly, the study utilizes the gridded version of SSURGO (gSSURGO) published as a file geodatabase. The gSSURGO is the product of merging traditional SSURGO digital maps and tabular data into the geographical extent of the conterminous U.S. and adding a corresponding raster map. The raster map is available in a ten-, thirty-, and ninety-meter cell size that approximates Albers Equal Area projected vector polygons. Each raster cell is identified by a map unit that describes an area dominated by a soil group. It is common to find several raster cells with the same map unit. Map units are not homogenous in their topographic and soil features, as they are made up of one to three "components". Components can also be composed of up to three "horizons" which are completely homogenous but vary by depth. The data for the map units, components, and horizons are stored in their tables within the geodatabase and are prefixed by "mu", "co", and "ch", respectively. Each can be related to one another via "keys".

Map units [Components] are identified by "mukey" ["cokey"] and can be found in both the map unit and component [component and horizon] tables to relate them. Horizon tables also have an additional key "chkey" to relate other tables that contain information that can be disaggregated within a horizon. The study utilized information stored in all tables, however, since there is horizontal [vertical] variability of properties within each map unit [component], the information must carefully be aggregated to properly represent this variability. Furthermore, at the horizon and component level, three values for the same properties are available: the low, representative, and high value. The study utilized representative values (post-fixed with "_r") in all calculations. Finally, horizon and/or horizon portions that are deeper than 150cm were also excluded from any aggregation, with the rationale that they are beyond the root zone of most field crops. Table A.3 presents an example of the operation using the horizon thickness.

Using a bottom-up approach, the information stored or calculated by horizons is first aggregated to the component level based on horizon thickness weighted averages. Similarly, component-level information and horizon aggregated data, are further aggregated to the map unit level via component extent weighted averages. Farm assignment of gSSURGO map unit key and aggregation weights method is outlined below (see Figure A.13 for a visual representation).

1. Geocoded farm mailing address to get an approximated spatial coordinate of the farm
2. Created a 1.5-mile radius circular spatial polygon using the spatial coordinate of the farm as its centroid
3. Cropped and masked the 3m gSSURGO map that overlaps with the circular polygon
4. Cropped and masked the 3m crop frequency layer that overlaps with the circular polygon
5. Extract map unit keys from portions of the masked 3m gSSURGO that overlap with the masked 3m crop frequency layer.
6. Calculated the weight for each mukey by counting the number of grids from (5) and then divided that by the total number of grids from (5)

The information for each farm is taken as the weighted average of the information in their assigned gSSURGO map unit key(s). The necessary fields and descriptions of the information contained in their respective tables in the gSSURGO, and derived soil attributes are listed in Table A.4.