Are crop insurance premium-implied yield distributions valid?

Aleksandre Maisashvili, Henry Bryant, George Knapek and James Marc Raulston
Department of Agricultural Economics, Texas A&M University, College Station, Texas, USA

Received 21 September 2018
Revised 14 March 2019
29 May 2019
Accepted 30 May 2019

Abstract

Purpose – The purpose of this paper is to develop methods for inferring if crop insurance premiums imply yield distributions that are valid according to standard laws of probability and broadly consistent with observed empirical evidence. The authors also survey current premium-implied distributions both before and after conditioning on the producer's choice of coverage level.

Design/methodology/approach – Under an assumption of actuarial fairness, the authors derive expressions for upper and lower bounds for premium-implied yield cumulative distribution functions (CDFs) at loss thresholds for each coverage level. When observed premiums imply a CDF that exceeds one or is not non-decreasing, the authors conclude that premiums cannot be actuarially fair. The authors additionally specify very weak conditions for premium-implied yield CDFs to be consistent with two possible reasonable parametric distributions.

Findings – The authors evaluate premiums for the year 2018 for 19,104 county-crop-type-practice combinations, both before and after conditioning on producer's choice of coverage level. The authors find problems in roughly one-third of cases. Problems are exhibited for all crops evaluated, and are strongly associated with areas with lower expected yields and higher yield variability. At least 40m acres are currently insured under premium schedules that cannot possibly be consistent with valid probability distributions.

Originality/value – The authors make two primary contributions. First, the premium-implied yield CDF bounds the authors derive requires fewer assumptions than previous similar work, while simultaneously placing more stringent conditions on premiums to be consistent with actuarial fairness. Second, the authors show that current US crop insurance premiums cannot possibly be actuarially fair for many cases, reflecting tens of millions of insured acres, which implies sub-optimal producer risk mitigation and inequitable expenditures for producers and taxpayers.

Keywords Crop insurance, Actuarial fairness, Yield distributions

Paper type Research paper

Crop insurance has become a critical and widely used risk management tool for US producers of major field crops. Not only do producers rely heavily on crop insurance, but government-provided reinsurance and premium subsidies represent large expenditures for US taxpayers. The fairness of crop insurance premiums is therefore a question of great importance.

In this paper, we revisit the question of fairness of crop insurance premiums. Like Babcock *et al.* (2004), we employ information revealed by premiums at different coverage levels to make inferences regarding actuarial fairness. We additionally decompose premiums into their base components and adjustments intended to account for differences in expected outcomes based on the producer's choice of coverage level and analyze current premiums before and after such adjustments. We are therefore able to disentangle the extent to which the problems we uncover exist in the unadjusted premiums and the extent to which they are caused by conditioning on producer's coverage level selection.

We make two primary contributions. First, we show that premiums and other actuarial information for different coverage levels can be used to infer upper and lower bounds for a yield cumulative distribution function (CDF) at specific yield values. Specifically, crop insurance premiums and their components reveal information about the yield distribution over the range

Agricultural Finance Review
Vol. 79 No. 4, 2019
pp. 467-490

0002-1466
DOI 10.1108/AFR-09-2018-0077

of yield values where indemnities would be triggered. This inference does not require any restrictive assumptions about the yield distribution (e.g. regarding symmetry or skewness).

Second, we use the methods we develop to conduct an updated analysis of the fairness of premiums for major crops under the current rules, data and parameters for their calculation. Because our methods do not require restrictive distributional assumptions, we apply them to a wider range of crops and regions than was previously possible. We apply our methods to premiums both before and after adjustments that condition on the producer's coverage level selection.

We present two motivating examples in the remainder of this introduction. For a typical non-irrigated corn producer in Dallas County, Missouri, the 2018 liability (the maximum possible indemnity) and total premium for yield protection (YP) for the 80 percent coverage level are \$310.46 and \$111.85 per acre, respectively[1]. For the 85 percent coverage level, these values are \$329.87 and \$136.70. Therefore, relative to the 80 percent coverage level, the increment to the 85 percent coverage level is an *uncertain* prospect with a *maximum* payoff of \$19.40 and a *certain* cost of \$24.85[2]. Such a prospect would not be selected (over simply remaining at the 80 percent coverage level) under any risk preferences, assuming only that the agent prefers more money to less.

A second example concerns YP premium increments as the coverage level increases for a typical non-irrigated soybean producer in Butler County, Iowa, for 2018. As the coverage level increases from 75 to 80 percent, the liability increases by \$26.42 per acre and the total premium increases by \$3.68 per acre. As coverage increases from 80 to 85 percent, liability increases by the same \$26.42 (because coverage levels are evenly spaced), but the premium increases by only \$3.48. Any realized yield between the 75 and 80 percent yield thresholds implies a *partial* loss of the amount at risk for the 75–80 percent range, but a *full* loss of the amount at risk for the 80–85 percent range. More generally, for any individual realized yield, the loss associated with a coverage level range must be at least as large as the loss for any lower equal-sized range. Consequently, the *expected* loss associated with the higher coverage level range must be at least as large as that of the lower range. Therefore, for premium increments to fairly reflect expected indemnity increments for successive equal-sized coverage level increases, premium increments must be non-decreasing. The Butler County soybean premiums do not conform to this simple requirement.

Such premium increments are *prima facie* defective, and indeed our first contribution in this paper is to carefully elaborate on the implications of such increments. The premiums stated above are total final premiums, reflecting coverage level conditioning. They reflect the probabilities assigned by the USDA Risk Management Agency (RMA) of a producer *who has selected such a specific coverage level* reporting a realized yield at or below the corresponding indemnity thresholds. Using the methods we develop, we show that to accept that these probabilities are well calibrated is to accept that there is a 128 percent chance that an indemnity will be triggered for the typical Dallas County dryland corn producer who selected 85 percent coverage. Obviously, this implication is incompatible with the laws of probability, and we must conclude that the total premiums for this example are amiss. Regarding the second example, accepting that loss probabilities are well-calibrated one must accept that the probability of a loss for an 80 percent coverage policy is at least 0.14, but also no more than 0.13. Again, we obviously must conclude that premiums cannot possibly correctly reflect expected indemnities.

In the next section we provide some background information. We then describe the calculation of crop insurance premiums and how RMA conditions on the producer's coverage level selection, shows how premiums imply upper and lower bounds for yield CDFs at coverage level thresholds, and devise methods for inferring if implied bounds for yield CDFs are consistent with empirical observations of yield distributions. After that, we survey 2018 crop insurance premiums for major crops across the USA, checking for problems such as those in the examples above for both total and unadjusted premiums.

A final section concludes with a brief discussion of the implications of our findings.

1. Background

Farmers have flexibility in tailoring crop insurance coverage to best suit their needs. Most producers can choose insurance that covers only their individual crop yield (YP), or insurance that covers their approximate market revenue (based on their individual yield and a futures price). Revenue policies allow the revenue guarantee to be improved by the price at harvest time (revenue protection (RP)), or can be based strictly on the price as of planting time (RP with harvest price exclusion, RP–HPE)[3]. Producers can choose to have their individual production units covered individually (optional unit structure), or collectively (enterprise unit structure). Finally, producers can select different coverage levels. In most cases, available coverage levels range from 50 to 85 percent with 5 percent increments. Indemnities are calculated based on the extent to which the realized yield or revenue fall below a threshold associated with the selected coverage level. That is, a lower coverage level corresponds to a higher deductible and greater extent of self-insurance.

Participation in the US crop insurance program has grown consistently since the passage of the Federal Crop Insurance Act of 1980, and it is now viewed as one of the most essential components of the risk management support system provided by the US government (Woodard *et al.*, 2011). Participation in the crop insurance program was given a large boost with premium subsidy increases provided by the 2000 Agricultural Risk Protection Act. In 2018, the program insured about $110bn in liabilities on 335m acres nationwide. The program is delivered by private markets but it is administered by the US Department of Agriculture's RMA. Insurance rates are set non-competitively by the RMA and are heavily subsidized in the form of direct producer premium subsidies, administrative subsidies and favorable government reinsurance. Premium subsidies decrease as the coverage level increases to mitigate potential adverse selection problems.

The crop insurance rate setting process is described in a recent RMA-commissioned report (Coble *et al.*, 2010). RMA develops target and base rates at the county and crop level. For each county and crop, the loss experience is summarized for each year in the historical rating period for specific crop types and production practices. The loss experience is homogenized across the many producer options described above. For revenue policies, liabilities and indemnities are calculated as though they had been yield only policies. All loss experience is converted to a 65 percent common coverage level: liabilities for units insured at higher coverage levels are adjusted down and those for coverage levels below 65 percent are adjusted upward to reflect what would have been incurred at the 65 percent coverage level. Base premium rates are then set based on this homogenized loss experience, and they therefore reflect a fair premium for YP at 65 percent coverage. This approach gives rise to the system of differentials that are used to calculate premiums for coverage levels other than 65 percent. Additional adjustments to this same standardized base rate are used to calculate premiums for revenue policies and alternative unit structures.

Potential problems with RMA's insurance rate-making approach have been described in the past. Babcock *et al.* (2004) derived conditions that must be met for rates to be actuarially fair under the assumption of symmetric or negatively skewed yield probability distributions. They then show that rates for higher coverage levels, calculated using the fixed rate differentials of the time, often could not be consistent with their conditions for corn, wheat and soybeans. They additionally concluded that moral hazard and adverse selection could not justify the fixed rate differentials.

Ramirez *et al.* (2011) proposed a methodology to assess the accuracy of historical lost-cost rating procedures vs alternative parametric premium specification methods. They find that lost-cost rating methodology leaves much to be desired, but that rating could be improved through the use of alternative methods and increased farm-level yield sample sizes. They additionally evaluated the bias and mean square error characteristics of various crop insurance premium specification methods.

Woodard *et al.* (2011) outlined biases and inconsistencies that exist in the rate-making process and guarantee determination, the equity of state excess loads, coverage level differential problems and intra-county risk heterogeneity. Perhaps most prominently, they documented that RMA does not account for yield trends in their rating, while yields were increasing substantially in the prior 15 years for crops such as corn.

Woodard *et al.* (2012) investigated patterns in the loss-ratio experience for the US corn insurance market using a spatial econometric model. They find that systematic geographically related misratings are present and provide estimates of the impacts of observable factors on the magnitude of misrating.

Ramirez and Carpio (2012) use a simulation methodology to estimate actuarially fair premiums for individual producers, given full information about those producers and then compare those premiums to actual premiums being charged. They conclude "that the persistently high government subsidy levels required to keep the program solvent could be solely explained by the inaccuracy in the RMA's premium estimates."

Sherrick *et al.* (2014) analyze the effects of rating methodology changes over time on premiums, risk exposures of market participants (including approved insurance providers) and loss ratios. They isolate the effects of changes to individual elements of the rate making and premium calculations. They find that ratings changes for corn and soybeans have led to their loss ratios more closely approaching targets.

Ramirez and Shonkwiler (2017) create a probabilistic model of producer crop insurance purchasing, under the assumption that neither RMA nor the producer know the correct actuarially fair premium. Using various possible combinations of an individual producer's expected yield and yield variability, they show that actuarially fair premiums can be very different across producers within a single county. They further show that RMA's exponentiation parameter for the yield ratio (see Equation (1) in the next section) results in premiums that are correct for the county on average, but end up effectively over-subsidizing riskier producers, leading to a risk of adverse selection.

Other recent research has looked at the issue of proper calibration of premiums against yield distributions for US area crop insurance (cf. the individual producer crop insurance that is the subject of the present paper). Park *et al.* (2018) fit county yield distribution parameters for a generalized Pareto distribution, including spatial smoothing using a Bayesian procedure. They report that their procedure results in improved loss ratios relative to other procedures.

2. Methods

In this section, we first describe how crop insurance premiums are calculated, paying particular attention to RMA's adjustments that condition expected outcomes on the producer's coverage level choice. Next, we show how crop insurance premiums, assuming actuarial fairness, imply upper and lower bounds for a yield CDF at indemnity thresholds. We then describe how we will use these formulæ to evaluate the fairness of current crop insurance premiums. In the last subsection, we describe the various data we employ in evaluating current premiums.

2.1 Calculation of normalized premiums

We calculate normalized (i.e. the projected price is assumed to be one) premiums according to RMA instructions, for a simple scenario where options such as trend adjustment and yield exclusion are not elected[4]. The base rate for YP at the 65 percent coverage level is:

$$\text{Base Rate} = \left(\frac{Y_{rate}}{Y_{ref}}\right)^{E} \times \text{Reference rate} + \text{Fixed rate} \tag{1}$$

where Y_{ref} is the county reference yield and E is the yield ratio exponent (simply a parameter specified by RMA). All values on the right hand side are retrieved from RMA's ADM database (hereafter simply "ADM," see USDA Risk Management Agency 2018), except for the producer's rate yield, Y_{rate}. The normalized liability for a coverage level c is:

$$\text{Liability} = Y_{appr} \times c, \tag{2}$$

where Y_{appr} is the producer's approved yield and c is written as a proportion (e.g. 0.85 for the 85 percent coverage level). The normalized premium, P_c, is then the base rate, multiplied by the liability, multiplied by the coverage level differential. This latter value is the product of two components from the ADM, the rate differential factor and the unit residual factor, which vary by coverage level and unit structure:

$$\begin{aligned} P_c &= \text{Liability} \times \text{base rate} \times \text{coverage level differential} \\ &= \text{Liability} \times \text{base rate} \times \text{rate differential factor} \times \text{unit residual factor}. \end{aligned} \tag{3}$$

All of our analysis reflects a representative producer, purchasing YP for an optional unit structure, who is typical in the sense that their rate yield and approved yield are both assumed to be equal to the county reference yield: $Y_{appr} = Y_{rate} = Y_{ref}$. Then the expression for the normalized premium for coverage level c reduces to a function of items exclusively from the ADM:

$$\begin{aligned} P_c = Y_{ref} \times c \times (\text{reference rate} + \text{fixed rate}) \\ \times \text{rate differential factor} \times \text{unit residual factor}. \end{aligned} \tag{4}$$

Note that all calculations reflect total premiums, not premiums reduced by subsidies.

We present a brief example of premium calculations reflected in Equations (1)–(4). All values in this example reflect YP coverage for an optional unit for non-irrigated corn producer in Dallas County, Missouri for 2018 (the first motivating example in the introductory section). We fetch the following values from the ADM: Y_{ref} is 98 bushels per acre, E is 1.057, the reference rate is 0.185 and the fixed rate is 0.068. We assume that the representative producer has an approved yield and rate yield equal to the county reference yield, so $Y_{appr} = Y_{rate} = 98$. Per Equation (1), the base rate evaluates to 0.253. That is, absent any other adjustments, the producer would pay about 25 cents per dollar of liability for 65 percent coverage. Adjustments are required for other coverage levels, however. For 80 percent coverage, the normalized liability, per Equation (2), is $98 \times 0.80 = 78.40$ bushels per acre (this would be multiplied by the projected price per bushel to arrive at the non-normalized liability as dollars per acre). For the 80 percent coverage level, we also fetch from the ADM the rate differential factor of 1.234 and the unit residual factor of 1.154. Then, per Equation (3), the normalized total premium is 28.246 bushels per acre. Multiplying this by the appropriate projected price for this example of \$3.96 per bushel results in the non-normalized total premium per acre of \$111.85 that appeared in the introductory section.

The decomposition of the coverage level differential is crucial to understanding how RMA conditions on coverage level selection. The components of the coverage level differentials for the examples in the introduction are presented in Table I. The rate differential factor is the basic adjustment that converts a 65 percent coverage YP base rate to an alternative coverage level. The rate differential factor is always 1.0 for the 65 percent coverage level, less than 1.0 for lower coverage levels and greater than 1.0 for higher coverage levels. The determination of the rate differential factor is very briefly described in Coble *et al.* (2010, section 4.5)[5]. Specifically, they state that the rate differential factors "are derived using a robust median regression method to estimate a model specification where the implied coverage level differential (from the unit level historical loss-cost data) is a function of the following: coverage level (and its squared term), the county-level base rate (and its squared term) and an interaction between the coverage level and the county-level rate at the 65 percent coverage level." They do not describe the calculation of the "implied coverage level differential (from the unit level historical loss-cost data)" (the dependent variable).

To our knowledge, there is no previously available public information regarding the purpose of unit residual factors, or the process by which RMA specifies the values they take. By analyzing the values that appear in the ADM, we find that the unit residual factor is always 1.0 for coverage levels 65 percent and below, and may be greater than 1.0 for higher coverage levels. RMA describes the unit residual factor as accounting for differences in outcomes between producers who have self-selected into higher coverage levels, and those who have not (Worth, 2016). They do not describe the unit residual factor as accounting for moral hazard or adverse selection[6].

Unit residual factors vary across crops and regions in a manner consistent with RMA's description of their purpose, and with the forces that shape producer decision making during crop cultivation. For 85 percent coverage for non-irrigated cotton, they are much higher than for all other crops we analyze. We observe a median value of 1.198 with an interquartile range of 0.012. By contrast, 85 percent coverage unit residual factors for non-irrigated corn have a median of 1.057 with an interquartile range of 0.026. The agronomic characteristics of cotton allow non-irrigated producers in certain regions to establish a crop with a basic level of inputs initially and decide late in the growing season if conditions warrant additional investment. Under some late-season conditions, the net marginal benefit of additional input use will consist of both an increment to expected market revenue and a decrement to expected insurance indemnities. The likelihood of this second component entering into such decisions would obviously be influenced by the coverage level the producer selected prior to the growing season.

Unit residual factors for non-irrigated soybeans are greater than 1.0 for only about 3 percent of cases. Soybeans are a relatively low-input crop, and are not widely cultivated using a non-irrigated practice in low-moisture regions. In this respect, it is unlikely that a soybean crop is subject to abandonment, relative to other crops. Non-irrigated wheat unit residual factors for 85 percent coverage exhibit a different pattern, with more than 49 percent of cases being above the median and modal value of 1.138. The way that unit residual factors vary across crops thus seems consistent with a role in accounting for

Table I.
Example components of coverage level differentials

	Dallas Co., Missouri Non-irrigated corn		Hitchcock Co., Nebraska Non-irrigated soybeans	
Cov. level (%)	Rate Diff. factor	Unit Resid. factor	Rate Diff. factor	Unit Resid. factor
50	0.861	1.000	0.795	1.000
55	0.908	1.000	0.848	1.000
60	0.954	1.000	0.924	1.000
65	1.000	1.000	1.000	1.000
70	1.074	1.055	1.164	1.055
75	1.155	1.104	1.373	1.110
80	1.234	1.154	1.483	1.165
85	1.349	1.214	1.561	1.166

producers' natural effort dilemma, based on the agronomic characteristics and the timing of costly input decisions for each crop. Unit residual factors thus seem to be accounting for a phenomenon that might be referred to as producers' willingness-to-abandon.

In the analysis that follows, we calculate normalized premiums both with and without the unit residual factor. We refer to the yield distribution not conditioned on the producer's choice of coverage level as G (inference using premiums calculated omitting the unit residual factor). We refer to the yield distribution conditioned on the choice of coverage level (inference using full premium calculations, including the unit residual factor) as H. By "conditioned on the choice of coverage level" here, we are referring to the purpose of the unit residual factor stated by Worth (2016), as discussed above.

By distinguishing between these two distributions, we reveal some information regarding the extent to which the problems we discover are caused by the unit residual factor, and the extent to which problems exist in other components of premiums. Babcock *et al.* (2004) reason that the base premiums for the numeraire coverage level (currently 65 percent) should be well-calibrated because they are directly based on the homogenized loss-cost experience. If this is accepted as correct, then problems revealed by the implied distribution G would indicate poorly specified rate differential factors. Problems revealed by the implied distribution of H would indicate poorly specified rate differential factors, or unit residual factors, or both. We acknowledge, however, that our analysis does not specifically test the calibration of premiums for the 65 percent coverage level, and we know of no research that specifically evaluates the assumption of Babcock *et al.* (2004) on this matter. That is, problems with the process of homogenizing the loss experience and specifying the base rates could offer an additional or alternative explanation for the problems we detect.

2.2 Premium-implied yield CDF lower and upper bounds

We adopt notation similar to Babcock *et al.* (2004). For the derivations in this subsection and the following subsection, we denote a yield CDF as F. Ultimately, we will apply the results derived here to yield distributions both reflecting and not reflecting unit residual factors.

For yield-only insurance, we denote the yield level below which an indemnity will be triggered as Y_c, where c reflects some percentage of a production unit's approved yield, Y_{appr}. For example, we might write Y_{85} to denote the indemnity threshold corresponding to an 85 percent coverage level, where $Y_{85} = 0.85 \times Y_{appr}$. For an increase in the coverage level from c to $c+\delta$, where $\delta > 0$, we denote the change in total normalized premium as $\Delta P_{c \to c+\delta}$. By "normalized," we mean that the price paid per unit of yield loss (the "projected price" in RMA's terminology) has been set to one, without loss of generality for the analysis that follows. If premiums are fair, then the change in the normalized premium is:

$$\Delta P_{c \to c+\delta} = Y_{c+\delta} F(Y_{c+\delta}) Y_c F(Y_c) - F(Y_{c+\delta}) E\left(y \middle| y < Y_{c+\delta}\right) + F(Y_c) E\left(y \middle| y < Y_c\right), \quad (5)$$

where F is the yield CDF (this is Equation (2) in Babcock *et al.*, 2004).

For a non-negative random variable, we can represent the conditional expectations in Equation (5) in terms of the yield CDF (proof of this is in the Appendix):

$$E\left(y \middle| y < z\right) = \frac{1}{F(z)} \int_0^z [F(z) - F(y)] dy. \quad (6)$$

Using this representation of the conditional expectation in Equation (5) and conducting some simple cancellation yields:

$$\Delta P_{c \to c+\delta} = Y_{c+\delta} F(Y_{c+\delta}) - Y_c F(Y_c) - \int_0^{Y_{c+\delta}} [F(Y_{c+\delta}) - F(y)] dy + \int_0^{Y_c} [F(Y_c) - F(y)] dy. \quad (7)$$

Integrating under the constants $F(Y_{c+\delta})$ and $F(Y_c)$, and consolidating the remaining integrals, we get:

$$\Delta P_{c\to c+\delta} = \int_{Y_c}^{Y_{c+\delta}} F(y)dy. \tag{8}$$

Thus, if premiums are fair, then differences in crop insurance total premiums for different coverage levels reveal the area under the premium-implied yield CDF between the two corresponding indemnity thresholds.

We can exploit this last expression to infer lower bounds for F at indemnity thresholds. Given that F must be non-decreasing in y and $\delta > 0$ we must have:

$$F(Y_{c+\delta}) \geqslant \frac{1}{Y_{c+\delta} - Y_c} \int_{Y_c}^{Y_{c+\delta}} F(y)dy = \frac{\Delta P_{c\to c+\delta}}{Y_{c+\delta} - Y_c}, \tag{9}$$

(proof of this is in the Appendix). It can be similarly shown that we must have a premium-implied upper bound of:

$$F(Y_c) \leqslant \frac{\Delta P_{c\to c+\delta}}{Y_{c+\delta} - Y_c}. \tag{10}$$

For the bounds of F at the indemnity thresholds corresponding to $Y_{55}, Y_{60}, \ldots, Y_{85}$, we can use the next lower threshold associated with each coverage level. For example, for the premium-implied lower bound of $F(Y_{85})$, we use the observed difference in premiums for the 80 and 85 percent coverage levels. For the lower bound of F at Y_{50}, we assume a premium of 0 is associated with $Y_0 = 0$. That is, it is costless for the producer to fully self-insure.

We use this approach, along with normalized 2018 premiums to draw Figures 1–3 (described later in this subsection).

Expressions (9) and (10) can be combined for some constant δ (i.e. equally-spaced changes in the coverage level and loss threshold) to show that we must have:

$$\Delta P_{c-\delta\to c} \leqslant \Delta P_{c\to c+\delta}. \tag{11}$$

That is, successive equally-spaced increases in the coverage level should be associated with ever greater premium increments. In Figure 2, the violation of expression (11) manifests as a

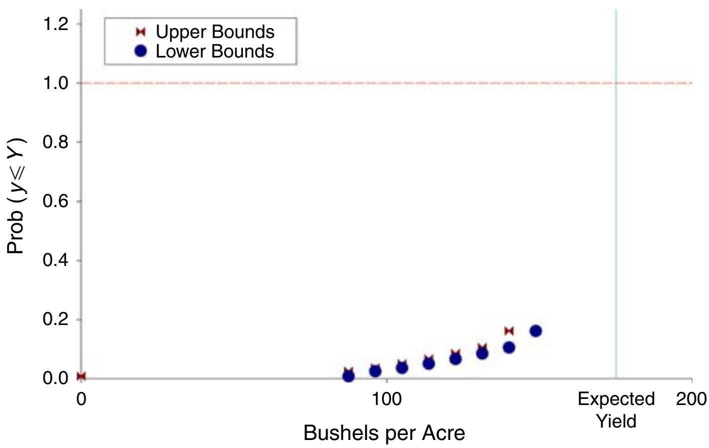


Figure 1.
2018 premium-implied bounds for the corn yield cumulative distribution function at indemnity thresholds, Allamakee county, Iowa

premium-implied CDF lower bound at the 75 percent coverage level that is higher than the corresponding premium-implied upper bound.

Obviously, any time the premium-implied lower bound of a yield CDF is greater than one for any Y_c, the implied underlying yield distribution is invalid, and we must conclude in such an instance that premiums cannot be actuarially fair. We must reach the same conclusion if premium increments ever decrease as the coverage level increases in fixed increments. These requirements constitute our first evaluation criterion.

What can we infer from a violation of our first criterion? A premium-implied yield CDF that exceeds one implies that at least some premiums are too high to be consistent with actuarial fairness. However, our simple method does not reveal which rates are too high. The location of the whole distribution may be wrong (implying base rates are too high), or it could be that only the rates for some upper coverage levels where the implied CDF exceeds one are too high. Decreasing premium increments (violations of expression (11)) could imply that the premium for the lower associated coverage level is too high, or that the premium for the higher associated coverage level is too low, or both.

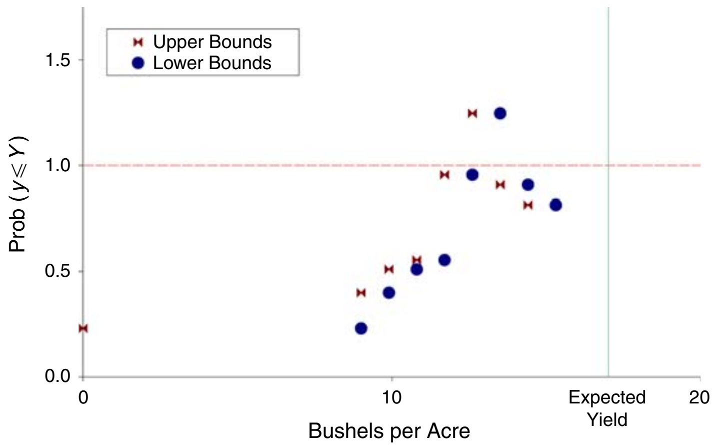


Figure 2.
2018 premium-implied bounds for the soybean yield cumulative distribution function at indemnity thresholds, Hitchcock County, Nebraska

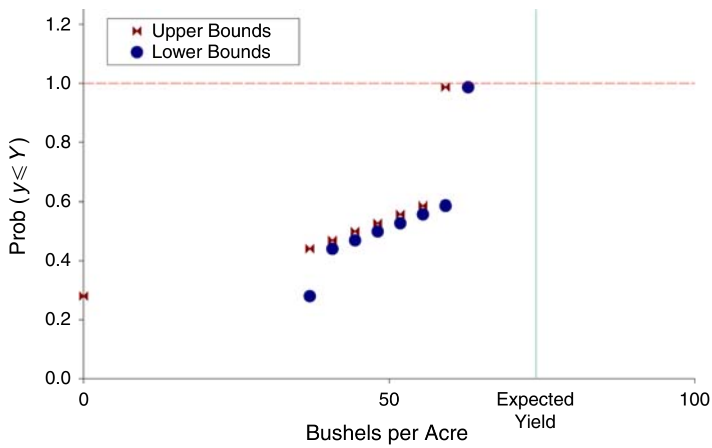


Figure 3.
2018 premium-implied bounds for the corn yield cumulative distribution function at indemnity thresholds, Hughes County, South Dakota

As a first example of the premium-implied bounds we derive above, we consider premiums, including unit residual factors, for a typical non-irrigated corn producer in Allamakee County, Iowa. CDF bounds implied by premiums for this example are shown in Figure 1. The round markers reflect the lower bound of the CDF associated with the liability thresholds for each coverage level, while the bow tie markers reflect the upper bounds. The leftmost round marker reflects the indemnity threshold for the 50 percent coverage level, while the rightmost round marker reflects the 85 percent coverage level. The vertical line labeled "expected yield" reflects the producer's approved yield used in calculating premiums, which we assumed to be equal to RMA's county reference yield for non-irrigated corn production in this county (175 bushels per acre). This example does not violate our first evaluation criterion, and appears consistent with a perfectly plausible crop yield CDF. It is easy to imagine a continuation of a curve conscribed by the implied bounds which would take a value of around 0.5 as it crossed the expected yield. That is, premiums are consistent with similar values for the mean yield and median yield.

As a second example of premium-implied bounds, we consider premiums, before conditioning on the selection of coverage level (i.e. not including unit residual factors), for a typical non-irrigated soybean producer in Hitchcock County, Nebraska. This example reflects both of the problems demonstrated in the two examples in the introductory section. We infer that actuarial fairness would imply upper and lower bounds for the yield CDF at liability thresholds for this typical producer that are presented in Figure 2. The third round marker from the right reflects the 75 percent coverage level. If actuarially fair, the premiums given above would imply that there is at least a 125 percent chance that the realized yield will fall below the 75 percent indemnity threshold, and that the probability that it will exceed that level is less than 0. The corresponding 75 percent coverage upper-bound bow tie marker is *below* the lower-bound round marker, denoting that if premiums were actuarially fair, there is at most a 91 percent chance that the yield will fall below this yield threshold. These implications are obviously incompatible with a valid probability distribution, and we must conclude that even these premiums that do not include unit residual factors are not actuarially fair for this example.

2.3 Reasonable probabilities of loss for parametric yield distributions

So far, we have made no assumptions about yield distributions. The evaluation criterion described above should be uncontroversial, however it is quite weak. For example, in Hughes County, South Dakota, we observe a premium-implied lower bound for $H(Y_{85})$ for non-irrigated corn of 0.987. This does not violate our first criterion. Nonetheless, the crop insurance premiums imply that at least 98 percent of the probability mass for the yield distribution has been assigned to values less than 85 percent of the expected yield. The bounds and expected yield for this example are presented in Figure 3. In terms of the actual yield values, accepting that these premiums are actuarially fair would be to accept that there is a 98 percent chance that a yield realization will be less than 63 bushels, even though the mean yield is 74 bushels. Any distribution consistent with both the premium-implied bounds and expected yield for this example would be quite odd, with some probability mass assigned to extremely high yield values. This is dramatically at odds with our experience that corn yields are neither bimodal nor substantially positively skewed. We are deeply suspicious of such examples, even though they do not violate our weak first criterion.

We therefore are motivated to devise a second criterion. Specifically, we seek to determine maximum reasonable values that $F(Y_{85})$ might take for parametric distributions that are consistent with empirical observations of crop yield realizations[7]. We seek to make this criterion conservative, however, essentially insisting only that any bimodality or positive skewness in the yield distribution is due only to truncation at zero[8]. To this end, below we calculate the maximum possible value of $F(Y_{85})$, given truncation at zero, for two parametric distributions commonly applied to crop yields. First, we calculate this maximum for a normal distribution. Second, we devise a process to determine this maximum for a beta distribution that is consistent with the skewness and bimodality assumption above and with some information from the ADM for each county-crop-type-practice combination we analyze[9].

However, we do not assume that crop yields are drawn from either of these specific distributions, we merely use these two maximum values for $F(Y_{85})$ as a starting point. We then take the maximum of these two maximum values, and inflate that by an additional 10 percent. We assert that this inflated maximum of maximums is a conservative threshold for $F(Y_{85})$ for some smooth, unimodal (over yield values greater than zero) probability distribution. We conclude that observed premium-implied lower bounds for $F(Y_{85})$ that exceed this threshold are indicative of premiums at the 85 percent coverage level that are unreasonably high given some uncontroversial, established knowledge about crop yield distributions.

For a normal distribution truncated from below at zero, we derive (see Appendix) the following expression for $F(Y_{85})$:

$$F(Y_{85}) = \frac{\Phi\left(0.15\gamma + 0.85\frac{\phi(\gamma)}{z}\right) - \Phi(\gamma)}{z} \tag{12}$$

where ϕ and Φ are the standard normal PDF and CDF, respectively, $\gamma = -\mu/\sigma$, $z = 1-\Phi(\gamma)$, and μ and σ are the mean and variance of the untruncated distribution. Unconstrained maximization over γ yields a maximum possible value for $F(Y_{85})$ for the normal distribution truncated from below at zero of max $[F_{norm}(Y_{85})] = 0.566$. Because the expression above is a function of only the *ratio* of μ and σ, and we maximized the expression without any constraints on the value of this ratio, 0.566 is the maximum possible value $F_{norm}(Y_{85})$ can take for any possible values for μ and σ.

For the beta distribution, we employ the four parameter specification. For the underlying non-truncated distribution, we have shape parameters α, $\beta > 0$, and location parameters a and b, with $a < b$, where the distribution has support over $[a, b]$. We assume that both positive and zero yields are possible, and therefore require $a \leqslant 0$ and $b > 0$. The underlying non-truncated distribution will be positively skewed for $\alpha < \beta$, and negatively skewed for $\alpha > \beta$.

We thus need four pieces of information to fully specify the beta distribution for a given county-crop-type-practice combination. We specify three constraints on the distribution. First, our assumption that any positive skewness is due strictly to truncation from below at zero implies $\alpha \geqslant \beta$. Given that we are trying to maximize F at a specific value, we want to shift the probability mass as far to the left as possible. This implies that at our optimum we will have exactly $\alpha = \beta$. Second, as in the previous subsection, we assume the value of the county reference yield from the ADM is the mean of the truncated yield distribution for a typical producer: $E(y|y > 0) = Y_{ref}$. Third, we specify that $F(0)$ must be no greater than the corresponding premium-implied upper bound calculated per expression (10). We numerically maximize $F(Y_{85})$ over the parameters α, β, a and b, subject to the three constraints. This process generates a unique max $[F_{\text{beta}}(Y_{85})]$ for each county-crop-type-practice combination we consider, for a beta distribution truncated from below at zero, that is consistent with empirical knowledge about yield distributions and consistent with the RMA's own actuarial data.

We now briefly return to the South Dakota example from the beginning of this subsection. Allowing any mean and SD for the underlying, non-truncated normal distribution, the largest possible value for $F_{norm}(Y_{85})$ is 0.566. Using the procedure described above and ADM data specific to this example, we numerically determine the maximum possible value for $F_{\text{beta}}(Y_{85})$ is 0.627. The specific cutoff value applied as our second criterion then is $1.1 \times \max(0.566, 0.627) = 0.69$. The premium-implied lower bound for the CDF at the 85 percent coverage level for this example is 0.987. Since this is greater than 0.69, we determine that this example violates our second criterion. This process is used to determine violations for criterion two in the results tables and maps. We believe this is still quite conservative. For this example, we were prepared to accept up to 68 percent of the probability mass being to the left of a value that is 15 percent below the mean – arguably an extremely odd characteristic for a crop yield distribution – without declaring a violation.

2.4 Summary of premium evaluation criteria

We summarize our two criteria for evaluating the actuarial fairness of current crop insurance premiums. The first criterion is the simplest, namely that premiums must reflect a valid yield distribution. The premium-implied lower bound for $F(Y_c)$ from expression (9) must not be greater than one for any c. If it is ever greater than one, we infer that premiums are, at least at some coverage level, clearly too high. Also, we require that premium increments never decrease as the coverage level increases in fixed increments, per expression (11).

As our second criterion, we compare premium-implied lower bounds for $F(Y_{85})$ to 110 percent of the maximum of the maximal reasonable values of $F(Y_{85})$ for the normal and beta distributions. Here, the maximal reasonable such value for the beta distribution is unique to each county-crop-type-practice combination analyzed.

We evaluate premiums for YP for each county-crop-type-practice combination by checking the two criteria in order. We do not check the second criterion if the first fails. For each combination, we apply the two criteria to implied yield distributions both without the inclusion of unit residual factors (G), and with (H).

2.5 Data

Our data source for all analysis that follows is RMA's Actuarial Data Master (ADM) for the 2018 insurance year (USDA Risk Management Agency, 2018). The ADM provides all of the inputs for calculation of normalized premiums for all valid coverage levels.

3. Survey of current premiums

We analyze 19,104 county, crop, type, practice combinations, both with and without unit residual factors for each. The crop, type and practice combinations we analyze are selected based on their importance in US agriculture among crops covered by RMA's "COMBO" policies for individual producers. Based on summary of business (SOB) data, we calculate that the combinations we analyze reflect about \$79bn of annual liability out of the total liability of about \$84bn that is insured under the COMBO rating process that this paper concerns[10]. Area policies and insurance for specialty crops are covered by different rate making processes that we do not consider here.

Percentages of cases that exhibit criteria violations for each combination are presented in Table II. The geographical distributions of violations for selected combinations are depicted in Figures 4–8. Analogous maps for other combinations are available in online supplemental material.

We observe more violations for premiums without unit residual factors. There are 4,765 violations of our first criterion across all county, crop, type and practice combinations analyzed (24.94 percent of cases). There are 592 violations of our second criterion when there was no violation of the first criterion (3.10 percent of cases). By contrast, for premiums including unit residual factors, we observe 3,988 criterion one violations (20.88 percent) and 2,708 criterion two violations (14.18 percent).

For premiums without residual factors, nearly all first criterion violations reflect decreasing premium increments as the coverage level increases (i.e. a violation of expression (11)). Only about 1 percent of such violations reflect an implied CDF that exceeds one.

Crop and practice	Without unit residual factors (%) Criterion 1 Violation	Criterion 2 Violation	No Violation	With unit residual factors (%) Criterion 1 Violation	Criterion 2 Violation	No Violation
Winter wheat						
Irrigated	49	4	47	41	21	38
Non-Irr and summerfallow	51	5	45	37	28	36
Continuous cropping	52	0	48	37	28	36
Spring wheat						
Irrigated	43	4	53	29	22	49
Non-Irr and summerfallow	44	3	54	36	22	43
Continuous cropping	43	2	54	33	51	16
Corn						
Irrigated	13	1	86	9	10	81
Non-irrigated	15	1	85	13	23	65
Soybeans						
Irrigated (incl. NFAC)	14	2	84	14	2	84
Non-Irr (incl. NFAC)	16	8	76	16	8	76
FAC (Non-irrigated)	17	8	75	17	8	75
FAC (Irrigated)	22	1	77	22	1	77
Grain sorghum						
Irrigated	22	4	74	11	7	81
Non-irrigated	19	2	78	9.6	5.9	84.5
Upland cotton						
Irrigated	5	0	95	5	16	79
Non-irrigated	7	0	93	14	23	63

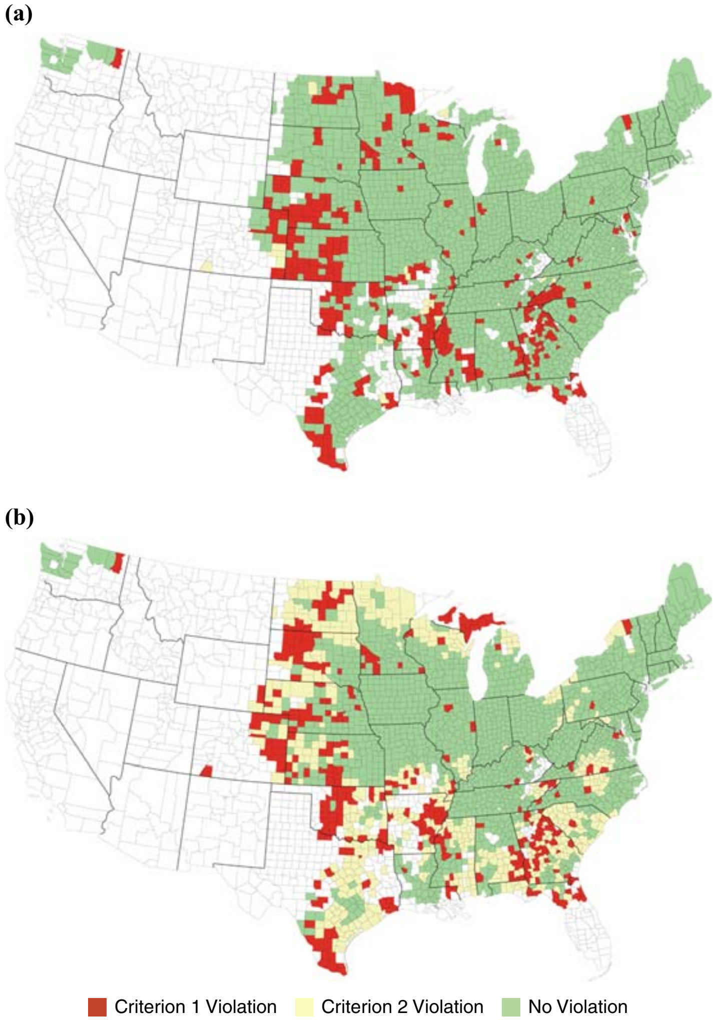
Table II. Percentages of criterion violations

Because unit residual factors can only increase premiums at coverage levels above 65 percent, they reduce violations of expression (11), while increasing occurrences of implied CDFs exceeding one. For premiums with unit residual factors, 89 percent of first criterion violations reflect decreasing premium increments, and 16 percent of such violations reflect an implied CDF that exceeds one. Note that these two problems may occur simultaneously, as Figure 2 demonstrates.

There are two exceptions to the general pattern described above. First, for soybeans, unit residual factors are rarely greater than one, even for higher coverage levels. As a result, final premiums that include unit residual factors reflect criteria violations that are nearly identical to premiums without such adjustments.

Second, for cotton premiums including unit residual factors, there are many more occurrences of premium-implied CDFs exceeding one. For non-irrigated cotton, the inclusion of unit residual factors increases first criterion violations from 37 counties to 75. Among the 75 violations when unit residual factors are included, 71 reflect $F(Y_{85}) > 1$, while only 14 reflect decreasing successive premium increments. So, for cotton, unit residual factors largely corrected the latter problem, at the expense of increasing the former and increasing the overall number of counties in which premium-implied distributions are invalid.

Geographic distributions of criteria violations vary across crop, type and practice combinations. For corn, violations largely appear in the southeast and in the western side of the nation's corn cultivating regions: the Dakotas down through Texas (Figure 4). For soybeans, violations are fairly diffuse, affecting even the productive Midwest region (Figure 5). Soybean violations are especially common in Oklahoma. Wheat violations are very common and are widely dispersed across the whole country, including the

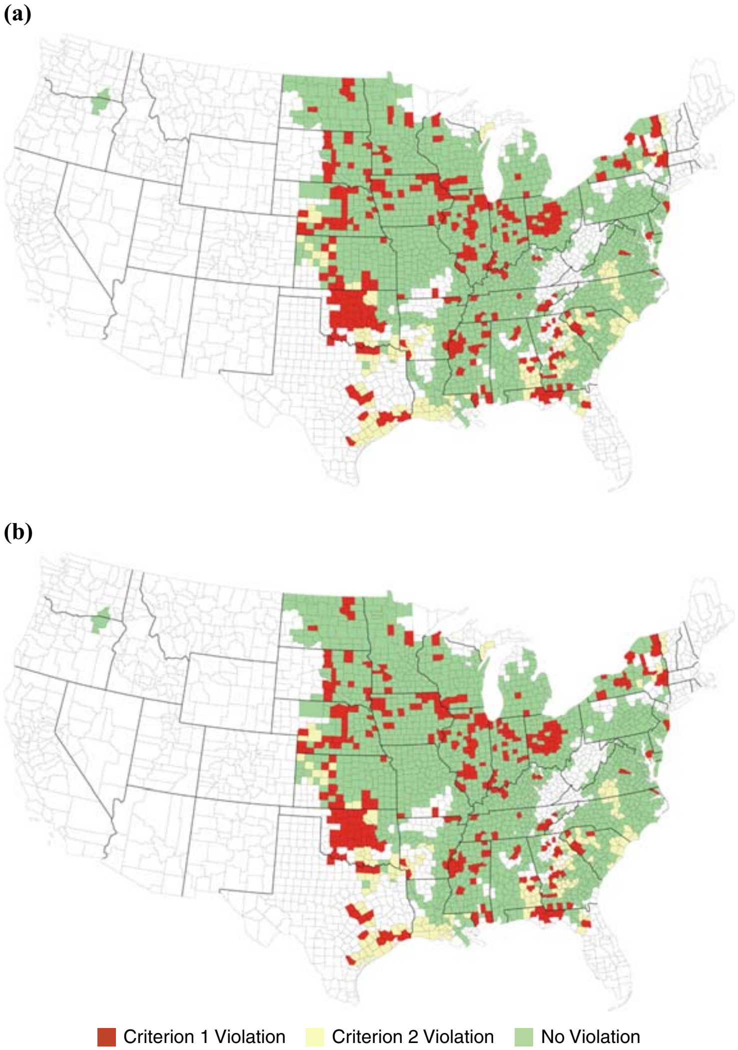


Notes: (a) Without unit residual factors; (b) unit residual factors

Figure 4.
Criteria violations for non-irrigated corn

Pacific Northwest (Figure 6). Texas wheat premiums are particularly dysfunctional. Grain sorghum violations follow a pattern similar to corn (Figure 7). Cotton premium violations are largely confined to Texas and Oklahoma, but large proportions of counties in these states are affected (Figure 8).

3.1 Correspondence of violations with loss experience and producer choices
Given that there appear to be regional patterns in the violations, we are motivated to explore the possibility that violations vary systematically in correspondence with crop productivity and yield variability. We evaluate this possible correspondence using logit models with a binary dependent variable indicating any violation (i.e. either of our two criteria) from premiums including unit residual factors. We calculate the average loss-cost ratio



Notes: (a) Without unit residual factors; (b) unit residual factors

Figure 5. Criteria violations for non-second crop, non-irrigated soybeans

(LCR; indemnities divided by liabilities) for each county over the years 2009 through 2018 using SOB data, and use this as one explanatory variable[11]. We interpret the LCR as largely reflecting yield variability for a county. We include the reference yield (Y_{ref}) as an additional explanatory variable to account for spacial variation in yield levels. We fit these models via maximum likelihood.

Results for non-irrigated corn are presented in Table III. The model is a good fit with an area under the receiver operating characteristic curve (AROC) of 0.87. The minimum possible AROC of 0.5 would imply the logit model had no explanatory ability at all, while the maximum possible AROC of 1.0 would imply the model could perfectly discriminate between counties with and without criteria violating premiums. Additionally, the model

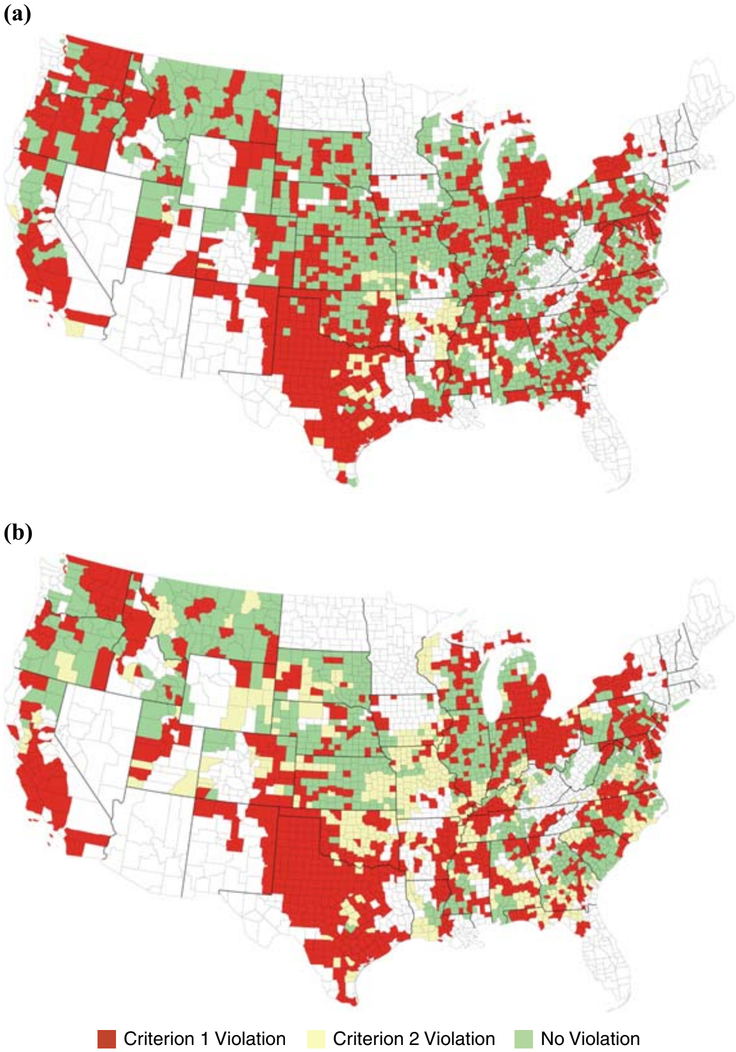
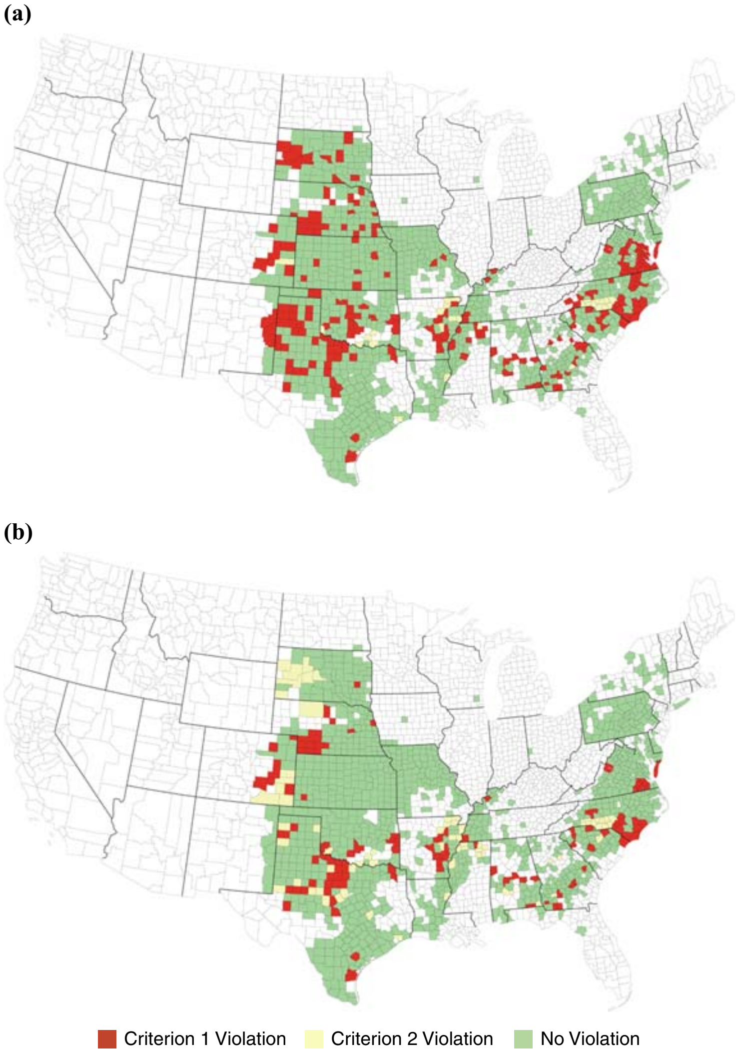


Figure 6.
Criteria violations for non-irrigated and summerfallow winter wheat

Notes: (a) Without unit residual factors; (b) unit residual factors

has a McFadden's pseudo R^2 (McFadden, 1974) of 0.326, where McFadden (1977) characterizes a pseudo R^2 of 0.2–0.4 as representing "an excellent fit," even though the measure has a maximum possible value of 1.0. Violation probabilities emanating from this logit model are presented visually in Figure 9. This figure corresponds to the actual occurrences of violations shown in the lower panel of Figure 4. It appears that our simple model with just two explanatory variables does a good job of capturing the regional pattern of violations we observe.

One or two-sided restrictions on an individual explanatory variable coefficient in our logit model can be tested using the ratio of the coefficient estimate and its standard error, which is distributed standard normal (Greene, 2000). We reject at the 1 percent level the null



Notes: (a) Without unit residual factors; (b) unit residual factors

Figure 7.
Criteria violations for non-irrigated grain sorghum

hypotheses that the LCR has a non-negative effect on the probability of a criteria violation for dryland corn, and that Y_{ref} has a non-positive effect.

The pattern observed in the logit models for corn hold across all non-irrigated crops. Detailed logit results for dryland crops other than corn are presented in the online supplemental information. The logit results for all crops are summarized in Table IV. We find that for all crops the LCR coefficient is positive and highly significant. For all crops, the Y_{ref} coefficient is negative, although it is not significant for soybeans and wheat. We therefore establish a clear pattern of our criteria violations occurring in the context of lower expected yields and especially occurring in the presence of higher yield variability.

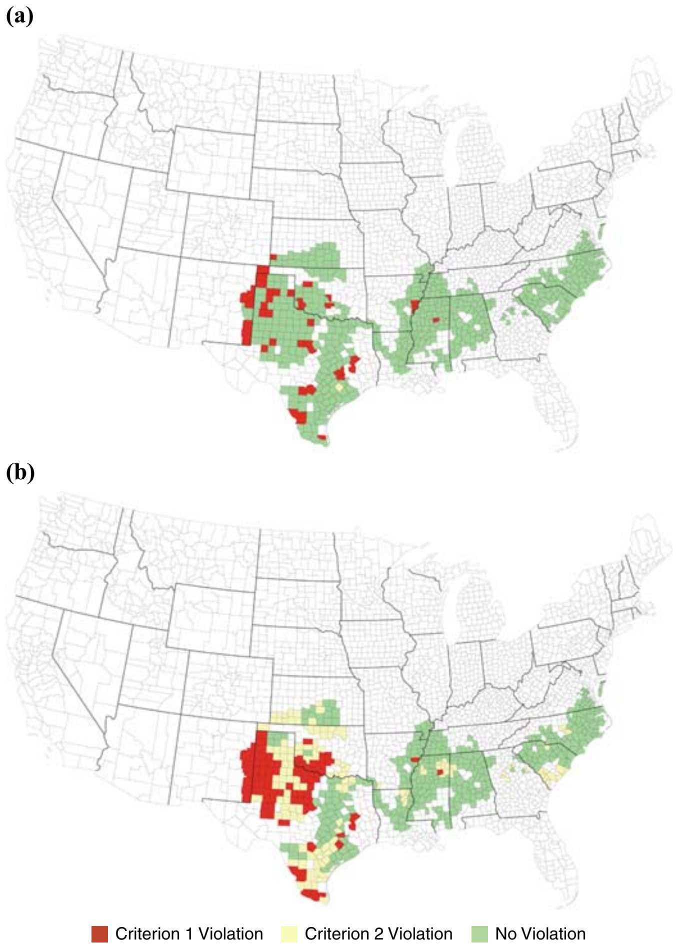


Figure 8.
Criteria violations for non-irrigated cotton

Notes: (a) Without unit residual factors; (b) unit residual factors

Table III.
Logit model of any criteria violation for dryland corn

Independent variable	Coefficient estimate (SE)	Units
Constant	3.353 (0.327)***	
Loss-cost ratio	4.014 (0.638)***	Dollar of indemnity per dollar of liability
County reference yield	−0.041 (0.003)***	Bushels per acre

Notes: Number of observations: 2,037; McFadden (1974) Pseudo R^2: 0.326; area under ROC curve: 0.87. *,**,***Statistically significant at 10, 5 and 1 percent level, respectively

We additionally find a significant association between criteria violations and producer coverage, as measured by dollars of liability per acre. We investigate this using ordinary least squares (OLS) models of liability per acre as a function of our any-violation dummy and the county reference yield. Liability per acre will obviously be higher in counties with a higher reference yield simply because yield is directly used in the liability calculation.

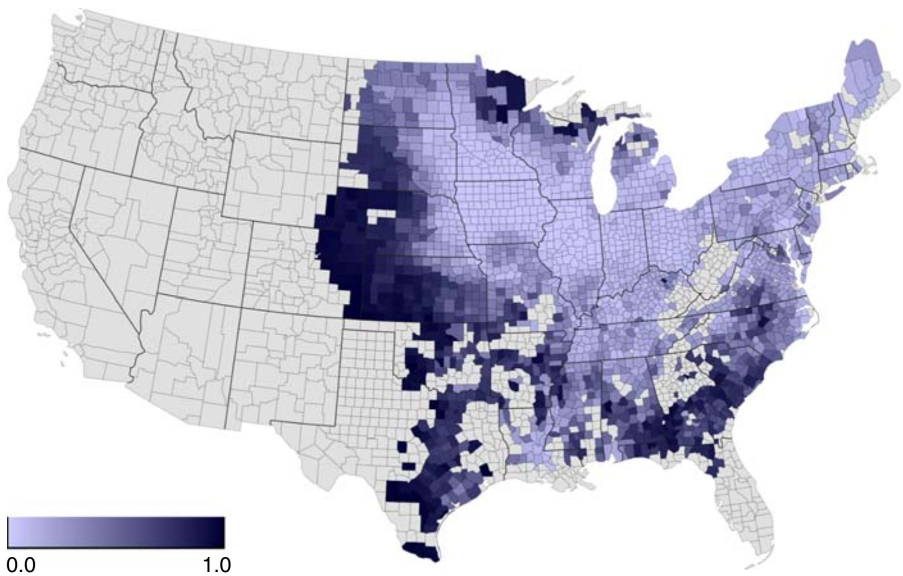


Figure 9. Violation probabilities from logit model for dryland corn

Crop	n	Pseudo R^2	AROC	LCR p-value	Y_{ref} p-value
Corn	2,037	0.33	0.87	< 0.001	< 0.001
Cotton	463	0.43	0.91	0.006	< 0.001
Grain sorghum	634	0.07	0.70	0.018	< 0.001
Soybeans	1,816	0.09	0.64	< 0.001	0.309
Winter wheat	1,698	0.03	0.59	< 0.001	0.790

Notes: LCR, loss-cost ratio; AROC, area under receiver operating characteristic curve. p-valules are associated with H_0: the variable has no effect on the probability of a violation; Pseudo R^2 is from McFadden (1974)

Table IV. Summary of logit models of any criteria violation for dryland crops

However, we also allow for the possibility that typical producer behavior (especially coverage level selection) may be different in high productivity areas (larger, more leveraged farms where operating loan terms require high coverage) and low productivity areas (small, less leveraged farms). We therefore include in these models a squared Y_{ref} term to allow for a possible non-linear influence of Y_{ref} on liability per acre.

Results for non-irrigated corn are presented in Table V. Y_{ref} exerts a significant positive effect on liability per acre, as expected. The coefficient on criteria violations is negative and significant at the 1 percent level. Results for other non-irrigated crops are summarized

Independent variable	Coefficient estimate (SE)	Units
Constant	138.802 (12.315)***	
Criterion 1 or 2 violation	−12.713 (3.462)***	1 if either criterion 1 or 2 is violated
Yref	0.059 (0.203)	Bushels per acre
*Yref*2	0.015 (0.001)***	

Notes: $n =$ 2,037; adjusted $R^2 =$ 0.84; the dependent variable is dollars of liability per acre. *,**,***Statistically significant at 10, 5 and 1 percent level, respectively

Table V. Ordinary least squares model of the association between criteria violations and liability per acre for dryland corn in Table VI (detailed results for all crops are presented in the online supplement). The effect of criteria violations on liability per acre is negative for all crops, and significant at the 10 percent level or better for all crops except grain sorghum. We therefore establish a clear negative association between criteria violations and the extent of producer coverage.

3.2 Potential causes of violations

The process by which rate differential factors are specified is only very briefly summarized in Coble *et al.* (2010). To our knowledge, the process by which unit residual factors are specified is not publicly documented at all. We are therefore prevented from investigating these key potential causes of the violations we find. We are thus relegated to only weak inferences regarding causes.

We have established that premium-implied distributions that are invalid or suspect correspond to a high yield variability, and also correspond to less coverage (lower liabilities) being purchased by producers. The violations we are able to detect are largely associated with premiums that are too high, especially for upper coverage levels. The correspondence between violations and producers electing lower coverage is potentially explained simply by downward sloping demand. The correspondence between high yield variability and our violations might be explained by a flaw in the specification of base rates, rate differential factors, or unit residual factors (or some combination of these).

However, rate setting and producer purchasing decisions will affect one another over time. Premiums that are too high for higher coverage levels (which is clearly true when $F(Y_{85}) > 1$) would decrease purchase of policies at those coverage levels, which in turn affects the process by which the loss experience is standardized to a common hypothetical coverage level of 65 percent, as described in (Coble *et al.*, 2010). Briefly, RMA cannot know what loss would have been incurred at the 65 percent coverage level when a producer elected a lower coverage level and no indemnity was triggered. This may bias the determination of base rates, which would then further influence the robust median regression process for fitting rate differential factors. A limited number of loss observations at higher coverage levels (we know this occurs with cotton in Texas, for example) may also directly influence the rate differential fitting process, introducing an econometric leverage effect, whereby a small number of observations ends up exerting a large influence on the rate differentials, especially for higher coverage levels. We speculate that any bias in the determination of base rates and rate differentials would further influence the specification of unit residual factors, but we have no way of knowing, as the process for their specification is not publicly available. If a lack of loss experience reflecting higher coverage levels leads to premiums that are too high for higher coverage levels, then a self-reinforcing process would obviously arise.

3.3 Affected acres

Using the 2018 SOB data, we estimate that at least 39m acres are insured under premium schedules that violate our first criterion[12]. This is only a lower bound, however, as the SOB data reflect only 220m acres for the state-county-crop-type-practice combinations we cover,

Table VI. Summary of OLS models of the association between criteria violations and liability per acre

Crop	n	R^2	Coefficient estimate for criterion violation (SE)
Corn	2,037	0.84	−12.713 (3.462)***
Cotton	463	0.60	−26.283 (13.441)**
Grain Sorghum	634	0.62	−5.728 (4.746)
Soybeans	1,816	0.85	−4.088 (2.425)*
Winter Wheat	1,698	0.75	−4.599 (1.844)***

Notes: The dependent variable for all models is dollars of liability per acre. *,**,***Statistically significant at 10, 5 and 1 percent level, respectively due to data confidentiality motivated omissions, and given that we have not analyzed all crops. Since more than 335m acres are insured for the USA, the number of acres insured in cases where premium schedules violate our first criterion may be tens of millions of acres more than 39M. Tens of millions more acres are insured under premium (including unit residual factors) schedules that violate our second criterion.

4. Conclusions

In this paper, we develop methods for inferring whether crop insurance premiums imply yield distributions that are valid according to standard laws of probability and whether they imply yield distributions that are broadly consistent with observed empirical characteristics of yield distributions. We survey premium-implied distributions both before and after conditioning on the producer's choice of coverage level. Overall, we observe more criteria violations for crops and areas where willingness-to-abandon and the likelihood of crop abandonment are higher: non-irrigated cotton and corn in semi-arid regions. Winter wheat premiums exhibit problems in much of the country. These observations are true for premiums both with and without conditioning on producers' coverage level choices. Thus, the pattern of results we observe clearly indicates problems with the processes used by RMA to specify both rate differential factors and unit residual factors.

Premiums that are too high work against the stated policy objectives of increasing sign-up and buy-up. For producers that do purchase crop insurance at premiums that are too high, government expenditure on subsidies is elevated. However, such premiums would obviously also lead to decreased buy-up, which would reduce expenditure on subsidies, so the net effect is unclear. Certainly there is an equity issue if some producers enjoy actuarially fair premiums while others do not. Also, premium-implied probability distributions that cannot possibly be properly calibrated must inevitably result in expected losses that are different than costs, counter to RMA's rate setting mandate.

Our results do not indicate problems for most county-crop-practice combinations under the loss-cost approach currently employed by RMA. It is not our purpose to advocate for an explicit distribution-fitting approach to specifying crop insurance premiums. We believe, rather, that our results highlight areas in which RMA's methodology requires adjustment. Even in the absence of an explicit distribution-fitting approach to rate setting, we argue that premium-implied loss probabilities should be broadly consistent with probability theory and with well-established empirical data regarding crop yield distributions.

Clearly our study has limitations. Our criteria do not determine if premium-implied loss probabilities are well-calibrated overall to actual yield distributions for a particular situation. We merely detect premium increments that are clearly inconsistent with standard probability concepts. Our criteria are not at all well-suited to detecting problems over the range of yield realizations reflecting larger losses (lower coverage levels), although such problems clearly would also be possible. In short, we detect only problems that are fairly obvious.

Notes

1. By "typical", we mean that the producer has a rate yield that is equal to RMA's published county reference yield for this county. Ignoring some special cases, the rate yield is essentially the sample mean of a production unit's historical yield observations. We also follow Babcock *et al.* (2004) in assuming that the producer's approved yield (the yield used in indemnity calculations) is equal to their rate yield, for analytical tractability. This assumption is exactly true when the producer does not elect options such as yield adjustment, trend adjustment and yield exclusion.
2. These are total (i.e. unsubsidized) premiums. The change in the producer (subsidized) premium relative to the change in liability would be even larger due to subsidies decreasing as the coverage level increases.

3. There are many other policies and options available that we do not describe here. There are policies based on area, rather than individual, yield realizations. Minimal "catastrophic" individual coverage is also offered to all producers for a nominal fee. Individual coverage options are also offered, such as trend adjustment of a producers expected yield and exclusion of some of a producer's historical yields in expected yield calculations.
4. We additionally abstract from limits on year-to-year changes in the base rate and coverage level differential.
5. Confusingly, the value that is described in the RMA's premium calculations as the "rate differential factor" (just one of two components of their "coverage level differential") is described in (Coble *et al.*, 2010) as the "coverage level differential."
6. Nonetheless, there are at least three types of producer private information relevant to expected indemnities. First, producers will have private information regrading growing conditions for an upcoming crop, such as soil moisture and pre-established nutrient availability. Second, RMA certainly does not have knowledge of the producer's effort level for cultivating the upcoming crop. Third, even though the individual producer's expected yield is a part of premium calculations (Y_{rate} and Y_{appr}), their yield variability is not.
7. Note that RMA does not explicitly specify parametric probability distributions for yields as part of the rate setting process (Coble *et al.*, 2010; Worth, 2016). Nonetheless, actuarial fairness requires YP premiums to be well-calibrated with the probabilities of various yield values being realized, and RMA is thus implicitly specifying information regarding yield distributions.
8. Non-positive skewness in the yield distribution was an explicit assumption in the criteria applied by Babcock *et al.* (2004). We relax this assumption somewhat by allowing positive skewness that is caused by truncation from below at zero.
9. The β distribution is a commonly assumed representation of crop yields (Nelson and Preckel, 1989; Borges and Thurman, 1994; Babcock and Hennessy, 1996; Coble *et al.*, 1996; Hennessy *et al.*, 1997; Knight *et al.*, 2010; Ubilava *et al.*, 2011), and was determined to be a better fit for corn and soybean yields than several other parametric distributions in Sherrick *et al.* (2004).
10. These liability values include YP and revenue policies. However, revenue policies use YP rates as a starting point, and make adjustments to reflect the price risk that is additionally incurred for revenue policies. Revenue policy premiums are therefore very unlikely to accurately reflect relevant risks when the underlying YP parameters do not.
11. Note we are considering the correspondence between the loss-cost ratio (LCR) and our violations, not the loss ratio (LR). The LCR is defined in the literature and in this paper as indemnities divided by liabilities (Woodard *et al.*, 2011, 2012; Ramirez *et al.*, 2011). The LR is defined in the literature and the SOB as indemnities divided by premiums.
12. This counts YP, RP and RP–HPE policies, while our analysis concerns only YP premiums. As described in the previous footnote, problems with YP premiums automatically imply problems with corresponding revenue policy premiums.

References

Babcock, B.A. and Hennessy, D.A. (1996), "Input demand under yield and revenue insurance", *American Journal of Agricultural Economics*, Vol. 78 No. 2, pp. 416-427.

Babcock, B.A., Hart, C.E. and Hayes, D.J. (2004), "Actuarial fairness of crop insurance rates with constant rate relativities", *American Journal of Agricultural Economics*, Vol. 86 No. 3, pp. 563-575.

Borges, R.B. and Thurman, W.N. (1994), "Marketing quotas and random yields: marginal effects of inframarginal subsidies on peanut supply", *American Journal of Agricultural Economics*, Vol. 76 No. 4, pp. 809-817.

Coble, K.H., Knight, T.O., Pope, R.D. and Williams, J.R. (1996), "Modeling farm-level crop insurance demand with panel data", *American Journal of Agricultural Economics*, Vol. 78 No. 2, pp. 439-447.

Coble, K.H., Knight, T.O., Goodwin, B.K., Miller, M.F., Rejesus, R.M. and Duffield, G. (2010), "A comprehensive review of the RMA APH and COMBO rating methodology, final report", Contract report, prepared by Sumaria Systems for the Risk Management Agency, Washington, DC.

Greene, W.H. (2000), *Econometric Analysis*, 4th International ed., Prentice Hall, Upper Saddle River, NJ, pp. 201-215.

Hennessy, D.A., Babcock, B.A. and Hayes, D.J. (1997), "Budgetary and producer welfare effects of revenue insurance", *American Journal of Agricultural Economics*, Vol. 79 No. 3, pp. 1024-1034.

Knight, T.O., Coble, K.H., Goodwin, B.K., Rejesus, R.M. and Seo, S. (2010), "Developing variable unit-structure premium rate differentials in crop insurance", *American Journal of Agricultural Economics*, Vol. 92 No. 1, pp. 141-151.

McFadden, D. (1974), "Conditional logit analysis of qualitative choice behavior" in Zarembka, P. (Ed.), *Frontiers in Econometrics*, Academic Press, New York, NY, pp. 105-142.

McFadden, D. (1977), *Quantitative Methods for Analyzing Travel Behavior of Individuals: Some Recent Developments*, Institute of Transportation Studies, University of California, Berkeley, CA.

Nelson, C.H. and Preckel, P.V. (1989), "The conditional beta distribution as a stochastic production function", *American Journal of Agricultural Economics*, Vol. 71 No. 2, pp. 370-378.

Park, E., Brorsen, B.W. and Harri, A. (2018), "Using Bayesian Kriging for spatial smoothing in crop insurance rating", *American Journal of Agricultural Economics*, Vol. 101 No. 1, pp. 330-351.

Ramirez, O.A. and Carpio, C.A. (2012), "Premium estimation inaccuracy and the actuarial performance of the US crop insurance program", *Agricultural Finance Review*, Vol. 72 No. 1, pp. 117-133.

Ramirez, O.A. and Shonkwiler, J.S. (2017), "A probabilistic model of the crop insurance purchase decision", *Journal of Agricultural and Resource Economics*, Vol. 42 No. 1, pp. 10-26.

Ramirez, O.A., Carpio, C.E. and Rejesus, R.M. (2011), "Can crop insurance premiums be reliably estimated?", *Agricultural and Resource Economics Review*, Vol. 40 No. 1, pp. 81-94.

Sherrick, B.J., Schnitkey, G.D. and Woodward, J.D. (2014), "Crop insurance loss experience, ratings changes, and impacts on participants", *Agricultural Finance Review*, Vol. 74 No. 4, pp. 443-463.

Sherrick, B.J., Zanini, F.C., Schnitkey, G.D. and Irwin, S.H. (2004), "Crop insurance valuation under alternative yield distributions", *American Journal of Agricultural Economics*, Vol. 86 No. 2, pp. 406-419.

Ubilava, D., Barnett, B.J., Coble, K.H. and Harri, A. (2011), "The SURE program and its interaction with other federal farm programs", *Journal of Agricultural and Resource Economics*, Vol. 36 No. 3, pp. 630-648.

USDA Risk Management Agency (2018), "Actuarial data master", available at: ftp://ftp.rma.usda.gov/pub/References/actuarial_data_master/ (accessed August 9, 2018).

Woodard, J.D., Sherrick, B.J. and Schnitkey, G.D. (2011), "Actuarial impacts of loss cost ratio ratemaking in US crop insurance programs", *Journal of Agricultural and Resource Economics*, Vol. 36 No. 1, pp. 211-228.

Woodard, J.D., Schnitkey, G.D., Sherrick, B.J., Lozano-Gracia, N. and Anselin, L. (2012), "A spatial econometric analysis of loss experience in the US crop insurance program", *Journal of Risk and Insurance*, Vol. 79 No. 1, pp. 261-286.

Worth, T. (2016), "Personal communication, Tom Worth, Chief Actuary", USDA Risk Management Agency, Kansas City, MO, October 20.

Further reading

Glauber, J.W. (2004), "Crop insurance reconsidered", *American Journal of Agricultural Economics*, Vol. 86 No. 5, pp. 1179-1195.

Ramirez, O.A., Carpio, C.E. and Collart, A.J. (2015), "Are the federal crop insurance subsidies equitably distributed? Evidence from a Monte Carlo simulation analysis", *Journal of Agricultural and Resource Economics*, Vol. 40 No. 3, pp. 457-475.

Appendix

Expected value, conditional on a realization below a threshold, in terms of the CDF for a non-negative random variable

This is an adaptation of the well-known derivation of the CDF variant of the unconditional expected value of a non-negative random variable. We assert that:

$$E(y|y<z) = \frac{1}{F(z)}\int_0^z [F(z)-F(y)]dy. \tag{A1}$$

For a corresponding PDF f, the right-hand side above can be re-written:

$$\frac{1}{F(z)}\int_0^z \left[\int_0^z f(t)dt - \int_0^y f(t)dt\right] dy = \frac{1}{F(z)}\int_0^z \int_y^z f(t)dtdy. \tag{A2}$$

Changing the order of integration, this expression is:

$$\frac{1}{F(z)}\int_0^z \int_0^t f(t)dydt = \frac{1}{F(z)}\int_0^z yf(y)dy, \tag{A3}$$

which is the standard definition of the given conditional expected value in terms of the PDF.

CDF lower bound at an indemnity threshold

Suppose we had:

$$F(Y_{c+\delta}) > \frac{1}{Y_{c+\delta}-Y_c}\int_{Y_c}^{Y_{c+\delta}} F(y)dy. \tag{A4}$$

The integral on the right-hand side of this inequality would be maximized if $F(y)$ were as large as possible over the interval $[Y_c, Y_{c+\delta}]$. For an arbitrary value for $F(Y_{c+\delta})$, given that F must be non-decreasing, we would have $F(y) = F(Y_{c+\delta})$ over this whole interval. The integral then evaluates to $F(Y_{c+\delta})\times(Y_{c+\delta}-Y_c)$, implying the contradiction:

$$F(Y_{c+\delta}) > F(Y_{c+\delta}).$$

Expression for $F(Y_{85})$ for a normal variable truncated from below at zero

Denote the mean and variance of the untruncated distribution as μ and σ, respectively. We define $\gamma = -\mu/\sigma$ and $z = 1-\Phi(\gamma)$, where Φ is the standard normal CDF, with corresponding PDF f. Then the truncated distribution has mean:

$$m = \mu + \left[\frac{\phi(\gamma)}{z}\right],$$

and CDF:

$$F(y) = \frac{\Phi(y-\mu)/\sigma - \Phi(\gamma)}{z}.$$

Recognizing that $Y_{85} = 0.85m$, the formula in the main text for $F(Y_{85})$ is produced after some substitution and simplification.

Corresponding author
Henry Bryant can be contacted at: henry@tamu.edu