Journal of the Royal Statistical Society Series C: Applied Statistics, 2024, **00**, 1–17
https://doi.org/10.1093/jrsssc/qlae020

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
Series C
Applied Statistics

Original Article

Testing for distributional structural change with unknown breaks: application to pricing crop insurance contracts

Hanjun Lu[1] and Alan P. Ker[2]

[1]Department of Food, Agricultural, and Resource Economics, University of Guelph, Guelph, Canada
[2]Department of Agricultural, Food, and Resource Economics, Michigan State University, East Lansing, MI, USA

Address for correspondence: Alan P. Ker, Department of Agricultural, Food, and Resource Economics, Michigan State University, East Lansing, MI, USA. Email: keralan1@msu.edu

Abstract

Agriculture in developed countries is produced under heavily subsidized insurance. The pricing of these insurance contracts, termed premium rates, directly influences farmers profits, their financial solvency, and indirectly, global food security. Changing climate and technology have likely caused significant shifting of mass in crop yield distributions and, if so, has rendered some of the historical yield data irrelevant for estimating premium rates. Insurance is primarily interested in lower tail probabilities and as such the detection of structural change in tail probabilities or higher moments is of great concern for the efficacy of crop insurance programs. We propose a test for structural change with an unknown break(s) which has power against structural change in any moment and can be tailored to a specific range of the underlying distribution. Simulations demonstrate better finite sample performance relative to existing methods and reasonable performance at identifying the break. The asymptotic distribution is shown to follow the Kolmogorov distribution. Our proposed test finds structural change in most major U.S. field crop yields leading to significant premium rate differences. Results of an out-of-sample premium rating game indicate that incorporating structural change in crop yields leads to more accurate premium rates.

Keywords: crop insurance, tail probabilities, structural change, unknown break

1 Introduction

1.1 Background

The Agricultural Act of 2014 solidified crop insurance as the cornerstone of domestic U.S. farm policy. In 2022, total U.S. government liabilities associated with the crop insurance program were \$194.6 billion and the Congressional Budget Office estimates crop insurance will cost the public roughly \$100 billion over the next decade. Moreover, publicly subsidized crop insurance is the dominant farm policy in most of the developed world. The pricing of these insurance contracts, termed premium rates, directly influences farmers profits, their financial solvency, and indirectly, global food security. The actuarially fair premium rate, denoted π, of an insurance contract is defined as the expected loss divided by total liability. Defining the random variable crop yield as Y, the actuarially fair premium rate for insurance coverage below a yield guarantee, denoted y_G, is:

$$\pi = \frac{1}{y_G} \int_0^{y_G} (y_G - y)\, \mathrm{d}F_Y, \tag{1}$$

Received: May 27, 2023. **Revised:** March 17, 2024. **Accepted:** April 5, 2024

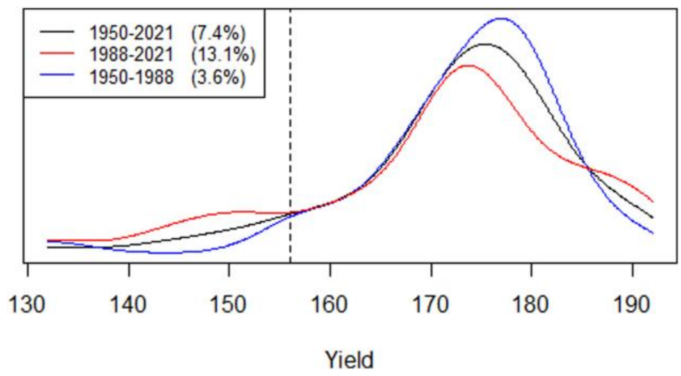

Figure 1. Conditional corn yield densities, La Crosse County, Wisconsin.

where the yield distribution, F_Y, needs to be estimated. The standard approach in the crop insurance literature is to estimate a temporal process using the historical yields, if necessary adjust the residuals for heteroskedasticity, and finally estimate the unknown yield distribution using the temporal and variance adjusted yields (Goodwin & Hungerford, 2015; Goodwin et al., 2000; Harri et al., 2011; Ker et al., 2016; Mieno et al., 2018; Woodard & Sherrick, 2011). Premium rates are derived using the estimated distribution plugged into equation (1). The above approach necessarily assumes the adjusted yields are strongly stationary and from the unknown distribution F_Y.

Climate and technological change bring into question the identically distributed assumption of the historical data. Testing and correcting for structural change in the conditional mean using, for example, splines or some other temporal model and correcting for change in the variance using appropriate heteroskedasticity adjustments are easily doable. However, testing for structural change in certain parts of the distribution—such as the lower tail which is particularly important for pricing insurance contracts—does not currently exist.[1] Ignoring structural change in the tails of the yield distribution will lead to biased tail probabilities, and by default, biased premium rates. In this article, we propose a test, based on a sup-type measure of distributional equivalence,[2] which can identify structural change in any moment, not just the conditional mean and variance. Furthermore, a weighting function can be imposed to only test over certain regions of the distribution. While we apply this test and argue its relevance for pricing crop insurance contracts and by default global food security, such a test is useful for pricing any insurance contract (e.g. property insurance, fire insurance, flood insurance, etc.) given that climate change appears to be altering loss/tail probabilities and as such calls into question the usefulness of historical data to estimate these probabilities.

Figure 1 provides an example of how structural change in higher moments affects the estimation of yield densities and premium rates, and thus illustrates the economic importance of our proposed test. Corn yields in La Crosse County, Wisconsin are found to contain a structural change in 1988 both beyond the conditional mean as well as beyond the conditional variance. That is, the premiums calculated are different solely because moments higher than order two are different.[3,4] Three estimated yield densities are plotted in Figure 1; one using the entire yield series, the second using data prior to the structural change in 1988, and finally one using data post the 1988 structural change. The premium rates are calculated with equation (1). At the 90% coverage level, the premium rate using all yield data is 7.4%. Conversely, the premium rate prior to the 1988 break is 3.6% whereas the premium rate after the break is 13.1%, almost double that of using the entire

[1] In the insurance world that precipitated this proposed test, we are concerned with significant portions of the lower tail and not just the extreme lower tail as in 'extreme-value theory'. We use the terms 'higher moments' and 'tail probabilities' to denote structural change beyond the mean.

[2] A 'sup-type' measure is the supremum or maximum over a given set. In our case, this set is across the various possible breaks.

[3] La Crosse County corn is chosen only as a representative example where a structural change is found.

[4] The United States Department of Agriculture (USDA) Risk Management Agency (RMA) methodology is used to adjust for change in conditional variance, which is detailed in Section 4.

yield series as is currently done. Given the U.S. crop insurance program is a multi-billion dollar program, these premium rate differences suggested by the structural change tests represent billions in transfers of public monies.

1.2 Structural change tests

Since the seminal work of Page (1954) and Chow (1960), testing for structural change in the conditional mean function has received significant attention in the literature. For example, Andrews (1993, 2003) considered three tests (LR test, LM test, and Wald test) for parameter instability with a known and unknown break. Bai & Perron (1998, 2003) considered structural change using a supremum version of the Chow-test across a set of possible breaks. Ghysels et al. (1998) and Hall and Sen (1999a, 1999b) proposed predictive tests to identify possible structural change for models estimated using Generalized Method of Moments (GMM). This work was followed by Gagliardini et al. (2005) which considered endogenous structural change while Li et al. (2021) focused on smooth structural change. Many tests have considered models under more specific conditions. A non-exhaustive list includes: models with possible unit roots (Hansen, 1992; Harvey et al., 2009; Perron & Yabu, 2009); models with common breaks (Oka & Perron, 2018); models with endogeneity (Kurozumi, 2017); models with cointegration (Kejriwal & Perron, 2010); quantile regressions models (Oka & Qu, 2011; Qu, 2008); models estimated via the Lasso method (Chan et al., 2014; Harchaoui & Lévy-Leduc, 2010; Qian & Su, 2016); panel data models (Bai, 2010; Baltagi et al., 2016; Li et al., 2016), forecasting models (Xu & Perron, 2017); and factors models (Bai et al., 2020). While most tests for structural change have focused on the conditional mean, a few have considered structural change in the conditional variance. For example, a Wald-type test and a bounds test for a single known break in the conditional variance were proposed by Greene (2012) and Kobayashi (1986), respectively.[5] More recently, Esfandiar et al. (2010) and Mumtaz et al. (2016) proposed an 'MZ' test and a sup-type 'MZ' test which can be used when the break in the conditional variance is known or unknown, respectively.

Tests for distribution equivalence also abound. The Cramér–von Mises (CvM) test (Anderson, 1962; Cramér, 1928; Mises, 1972) examined the goodness of fit using an integrated square difference (ISD) metric between two empirical cumulative distribution functions (CDF) or an empirical CDF and a given CDF. The Anderson–Darling test (Anderson & Darling, 1952) is similar to the CvM test, except that the ISD is weighted by the pointwise variances of empirical CDFs. Moreover, the test can be applied to greater than two empirical CDFs. An alternative to the above two tests is the Kolmogorov–Smirnov (KS) test (Darling, 1957) which focuses on the supremum of the absolute difference between two empirical CDFs. A closely related test is Kuiper's test (Kuiper, 1960), where the test statistic is the sum of the absolute values of the single largest positive difference and single largest negative difference from the empirical CDFs. The Mann–Whitney U-test (Mann & Whitney, 1947) and the Wilcoxon signed-rank test (Wilcoxon, 1945) assume the distributions only differ in location under the alternative. Conversely, the Cucconi test (Cucconi, 1968) and the Lepage test (Lepage, 1971) assume the distributions only differ in location and scale in the alternative space. More recently, Li (1996) proposed a test of distributional equivalence between two unknown densities using an ISD measure based on kernel estimates. The test was later expanded to mixed discrete and continuous data (Li et al., 2009). Inoue (2001) tested for distributional equivalence in time series for an unknown break by using the weighted KS test and the weighted CvM test.

The purpose of this article is to design a test for structural change that has power against an unknown break in any moment or specific part of the distribution and is relatively easy to apply. Our purposed test does not require the break to be known and has power against structural change in the conditional mean, the conditional variance, and all upper moments. The proposed test provides two contributions to the existing literature. First, it allows superimposing non-uniform weights easily so that it can be tailored to evaluate any subset of the distribution domain. This is of great relevance to crop insurances and other types of insurances where only tail probabilities matter. Second, the proposed test has greater power, relative to existing tests, for structural change

[5] The Wald test has the property that the probability of type I error is larger than the critical level used, which can be partially fixed by the bounds test.

above the conditional mean and provides a reasonable estimate for the unknown break when there is one.[6] We construct the proposed test by combining distributional tests of equivalence with a sup-type measure. We argue that this is important for the insurance industry and particularly so in light of the effect climate change is having on tail probability events. We focus on crop insurance as this is not only the impetus for deriving the test but it also represents multi-trillion dollars in world-wide liability and is imperative to food security.

The next section outlines the proposed test and its accompanying asymptotics. The third section outlines our simulation results while the fourth and fifth sections present the results of our application to agricultural crop yields and implications for insurance. The final section summarizes our article.

2 Proposed test

Assume an independently distributed time series $y_1,\ldots,y_T$, where observations $y_1,\ldots,y_t$ follows probability density function (pdf) f and the remaining $y_{t+1},\ldots,y_T$ follows pdf g. Assume that both f and g are continuous and bounded. The null hypothesis of interest is $H_0 : f(y) = g(y)$ almost everywhere (a.e.). While there are many measures of distance between $f(y)$ and $g(y)$, we considered the commonly used ISD:

$$ISD = \int_{-\infty}^{\infty} (f(y) - g(y))^2 \, dy \tag{2}$$

Intuitively, a sup-type measure based on a distributional equivalence measure could be used to test for unknown breaks, that is t unknown, in any part of the distribution or any moment. The proposed test statistic, ISD^*, can be obtained based on the supremum of ISD by replacing $f(y)$ and $g(y)$ by their kernel estimates in equation (2):

$$\begin{aligned} ISD^* &= \sup_{t \in TT} \widehat{ISD}_t \\ &= \sup_{t \in TT} \int_{-\infty}^{\infty} (\hat{f}(y) - \hat{g}(y))^2 \, dy \end{aligned} \tag{3}$$

where TT is a subset of $(1, \ldots, T)$ which contains all possible breaks. As is common, TT is a subset that ensures sufficient data exists to estimate the densities on either side of the possible breaks. f and g can be consistently estimated by their corresponding kernel estimates:

$$\hat{f}(y) = \frac{1}{T_1 h_1} \sum_{i=1}^{t} K\left(\frac{y_i - y}{h_1}\right) \quad \text{and} \quad \hat{g}(y) = \frac{1}{T_2 h_2} \sum_{i=t+1}^{T} K\left(\frac{y_i - y}{h_2}\right)$$

where K is a kernel function,[7] h_1 and h_2 are smoothing parameters, T_1 and T_2 are sample sizes for the two subsets. Note, H_0 will be rejected at a given critical level if ISD^* is sufficiently large. If H_0 is rejected, the corresponding t (denoted as t^*) will be used to estimate the unknown break.

To demonstrate the practical applicability of ISD^*, we present and prove the following theorem. It states that the asymptotic distribution of ISD^* under the null follows the Kolmogorov distribution. It implies that the asymptotic distribution of ISD^* is of the same form under the null hypothesis for any continuous and bounded densities f and g.

Theorem 1 Under the null hypothesis, assuming that $h_1, h_2 \to 0$ and $T_1 h_1, T_2 h_2 \to \infty$, then for any positive number ε,

$$Pr(\sqrt{T} ISD^* < \varepsilon) \xrightarrow{T_1, T_2 \to \infty} \sum_{x=-\infty}^{\infty} (-1)^x e^{-2x^2 \varepsilon^2}$$

[6] While the methodology proposed in Inoue (2001) theoretically achieves similar purpose as our proposed test, ours shows better performance in practice, especially in small samples.

[7] Some commonly used kernel functions include: normal, uniform, triangle, quartic, and Epanechnikov. Kernel functions are symmetric with mean 0 and integrate to 1 (Li & Racine, 2007).

Proof. ISD can be expanded to:

$$\begin{aligned} ISD &= \int_{-\infty}^{\infty} (f(y) - g(y))^2 \, \mathrm{d}y \\ &= \int_{-\infty}^{\infty} (f(y)\, \mathrm{d}F(y) + g(y)\, \mathrm{d}G(y) - f(y)\, \mathrm{d}G(y) \\ &\quad - g(y)\, \mathrm{d}F(y)) \end{aligned}$$

A feasible test statistic can be obtained by replacing $f(y)$ and $g(y)$ by their kernel estimates and replacing $F(y)$ and $G(y)$ by their empirical distribution functions $\hat{F}(y)$ and $\hat{G}(y)$ (Li, 1996; Li et al., 2009):

$$ISD^* = \sup_{t \in TT} \int_{-\infty}^{\infty} (\hat{f}(y)\, \mathrm{d}\hat{F}(y) + \hat{g}(y)\, \mathrm{d}\hat{G}(y) - \hat{f}(y)\, \mathrm{d}\hat{G}(y) - \hat{g}(y)\, \mathrm{d}\hat{F}(y)) \tag{4}$$

With $\int L(y)\, \mathrm{d}\hat{F}(y) = \frac{1}{T_1} \sum_{i=1}^{t} L(y_i)$ and $\int L(y)\, \mathrm{d}\hat{G}(y) = \frac{1}{T_2} \sum_{j=t+1}^{T} L(y_j)$, where $L()$ is any location-scale function, ISD^* can be formulated as:

$$\begin{aligned} ISD^* &= \sup_{t \in TT} \left\{ \sum_{i=1}^{t} \hat{f}(y_i)/T_1 + \sum_{j=t+1}^{T} \hat{g}(y_j)/T_2 \right. \\ &\quad \left. - \sum_{j=t+1}^{T} \hat{f}(y_j)/T_2 - \sum_{i=1}^{t} \hat{g}(y_i)/T_1 \right\} \\ &= \sup_{t \in TT} \left\{ \sum_{i=1}^{t} (\hat{f}(y_i) - \hat{g}(y_i))/T_1 - \sum_{j=t+1}^{T} (\hat{f}(y_j) - \hat{g}(y_j))/T_2 \right\} \end{aligned} \tag{5}$$

Let $A_t = \sum_{i=1}^{t} (\hat{f}(y_i) - \hat{g}(y_i))/T_1$ and $B_t = \sum_{j=t+1}^{T} (\hat{f}(y_j) - \hat{g}(y_j))/T_2$. Next, we show that the proposed test statistic does not depend on the form of A_t and B_t. Suppose there are q possible breaks, and let $A_{(t)}$ be $a = A_{(t_1)} \le A_{(t_2)} \cdots \le A_{(t_q)} = b$. For any p that $a \le p \le b$, it has: $0 \le \frac{p-a}{b-a} \le 1$. Let $z = \frac{p-a}{b-a}$ and $a_{(t)} = \frac{A_{(t)}-a}{b-a}$.[8] Define $a^{-1}(z) = inf\{t : a_{(t)} \ge z\}$. The properties of $a^{-1}(z)$ are provided in online supplementary material. Let M and N be two independent empirical distribution functions of the standard uniform distribution. By online supplementary material, Proposition 1 (provided in online supplementary material), if the null hypothesis of no structural change is true, the following holds:

$$\begin{aligned} ISD^* &= \sup_{t \in TT} \{A_t - B_t\} = \sup_{t \in TT} \left\{ \left(\frac{A_t - a}{b - a} - \frac{B_t - a}{b - a} \right) (b - a) \right\} \\ &= \sup_{t \in TT} \{(a_t - b_t)(b - a)\} \\ &= \sup_{0 \le z \le 1} \{(b - a)(N - M)(z)\} \end{aligned}$$

Above implies that the distribution of ISD^* will stay the same and not depend on the original form of $f(y)$ and $g(y)$ under the null hypothesis. The uniform process $\sqrt{T}((b - a)(N - M)(z))$ is shown to converge to a

[8] Assuming that T_1/T_2 is a constant when $T_1, T_2 \to \infty$, the nuisance parameters (T_1, T_2, and h) are neutralized when standardizing $A_{(t)}$ to $a_{(t)}$.

Brownian bridge process. Komlóos et al. (1975) found a rate of convergence such that:

$$Pr\left(\sup_{0\le z\le 1}(\sqrt{T}((b-a)(N-M)(z))-B_z)>\frac{x+c\log T}{\sqrt{T}}\right)<Ke^{-\lambda x}$$

for all T and x, where c, K, and λ are positive constants. Bretagnolle and Massart (1989) made assumptions on those parameters and provided a proof of Komlóos et al. (1975). That is, the convergence of the uniform process to the Brownian bridge holds when $c=12$, $K=2$, $\lambda=\frac{1}{6}$ and $T\ge 2$. With Komlóos et al. (1975) convergence and Bretagnolle and Massart (1989) constants, under the null hypothesis, $\sqrt{T}ISD^*$ converges in distribution to $\sup_{0\le z\le 1}B_z$, whose distribution is known (see Darling, 1957 and Dudley, 2002—Proposition 12.3.3):

$$Pr(\sqrt{T}ISD^*<\varepsilon)\xrightarrow{T_1,\,T_2\to\infty}\sum_{x=-\infty}^{\infty}(-1)^x e^{-2x^2\varepsilon^2}.$$

Therefore, the proposed test is independent of the form of f or g under the null and follows the Kolmogorov distribution. □

We note that nothing in the above proofs prevent the restriction of the ISD measure to a subset of f and g. In this sense, the proposed test can be used to consider structural change in a specified range. One simply takes the integral within the test statistic over the subset rather than taking the integral over entire support. This is beneficial within an insurance context where the lower tail or just the lower tail below some insurance guarantee is of primary interest.

3 Finite sample simulations

In this section, we assess the performance of the proposed test in finite samples via simulations. The proposed test is applied to data from distributions differing in the ith moment ($i=1, 2, 3, 4$). We consider the first four moments and simulate from four sets of distributions with each set comprising five distinct distributions. Table 1 gives the details on alternative density functions g. For example, the column labelled 'I' shows the first set of five distributions that differ only in the first moment; the column labelled 'IV' has another set of five distributions that differ in the forth moment or higher. The null density f is $N(1, 1)$ in all cases. A number of caveats regarding g deserve attention. First, with respect to simulations for structural change in either the first and second moment (column 'I' and 'II', respectively), densities f and g are both normal and so only differ across the first moment or second moment. Second, with respect to simulations for structural change in the third and fourth moments (column 'III' and 'IV', respectively), mixtures of normals are necessarily used for g. When considering structural change in the third moment, the first two moments are equal between f and g, but all higher moments are not. When considering structural change in the fourth moment, the first three moments are identical between f and g while the fourth and higher moments are not. Third, by construction, the difference between f and g increases from density g_1 through density g_5. The difference between f and g is zero when g takes density g_1, which represents the null hypothesis.

We consider sample sizes of 70, 150, 500, and 1,000 and the significance level of the test is 0.05 in all simulations.[9] We specifically consider a sample size of 70 as county-level crop yield data began to be collected in the 1950s, resulting in a sample size of 70 for most annual crop yield data at county level. A sample size of 70 for a test of equivalence of nonparametrically estimated densities with an unknown break is borderline nonsensical (prefer more data) but represents a worst case scenario for our proposed test. The results presented are derived using the leave-one-out cross-validation method for bandwidth selection. The kernel function used is the Gaussian

[9] Results for sizes 0.01 and 0.10 are structurally similar.

Table 1. Simulated alternative densities g

	Simulation Scenarios			
	I	II	III	IV
g_1	$N(1, 1)$	$N(1, 1)$	$N(1, 1)$	$N(1, 1)$
g_2	$N(2, 1)$	$N(1, 2)$	$0.13N(2.64, 0.6) + 0.87N(0.75, 0.79)$	$0.49N(1.0, 0.6) + 0.51N(1.0, 1.27)$
g_3	$N(4, 1)$	$N(1, 4)$	$0.19N(2.57, 0.6) + 0.81N(0.64, 0.67)$	$0.59N(1.0, 0.6) + 0.41N(1.0, 1.39)$
g_4	$N(6, 1)$	$N(1, 8)$	$0.23N(2.53, 0.6) + 0.77N(0.55, 0.55)$	$0.66N(1.0, 0.6) + 0.34N(1.0, 1.50)$
g_5	$N(8, 1)$	$N(1, 16)$	$0.26N(2.50, 0.6) + 0.74N(0.48, 0.41)$	$0.71N(1.0, 0.6) + 0.29N(1.0, 1.60)$

Note. Scenario 'I' refers to densities that only differ in the first moment. Scenario 'II' refers to densities that only differ in the second moment. Scenario 'III' refers to densities that only differ in the third and above moments. Scenario 'IV' refers to densities that only differ in the fourth and above moments.

function. For each g and sample size combination, 1,000 simulations are undertaken. For comparison, the Bai–Perron test (Bai & Perron, 1998), the sup-type 'MZ' test (Mumtaz et al., 2016), the weighted KS test and the weighted CvM test (Inoue, 2001) are considered as well. Those tests are the most commonly applied test when the break is unknown. The true break, denoted as B, is located in the middle of the sample. We consider the set of possible breaks to be $TT = (10, 11, 12, \ldots, T-12, T-11, T-10)$. While we have derived the asymptotic distribution of our test statistic under the null, as discussed earlier, Li (1996) has shown randomization methods can be used in practice to recover a distribution of the test statistic under the null hypothesis, which tends to perform better in terms of reducing probability of type II error. Specifically, to recover a realization from the distribution of our test statistic under the null, we first randomize the data. Based on the randomized data, we estimate the ISD_t at each possible break in TT. We then take the supremum of the ISD. This supremum represents one realization from the distribution of ISD^* under the null hypothesis. We repeat this 1,000 times to recover 1,000 realizations from the null distribution. We use the 1,000 realizations to recover the empirical distribution function of the null distribution. We then compare our test statistic to the empirical null distribution. As noted by Li (1996), it is important to use the same smoothing parameters as used in the original estimate of ISD^*.

The rejection rates from the proposed test and other tests at samples of 70 and 1,000 are located in Table 2.[10] ISD^*-R is when randomization method is used to recover the distribution of our test statistic under the null hypothesis; ISD^*-A is when the asymptotic distribution of our test statistics under the null hypothesis is used. As mentioned, the difference between f and g increases from density g_1 through density g_5 within each set of simulated distributions. This is shown in the third column of Table 2 which states the true ISD between f and g. Not surprising, the ISD between f and g declines dramatically as the moment of structural change increases while all lower moments remain equal. This intuitively makes sense because a given percentage change in a higher moment will have a relatively smaller impact on the density at a given support point than a lower moment.

A number of interesting points are worth consideration from Table 2. First, we find results similar to Li (1996) in that the randomization method performs better than the asymptotic null distribution and there is a tendency to over-reject which goes away as the sample size increases. Therefore, we focus on the randomization method (column ISD^*-R) rather than the asymptotic null (column ISD^*-A) in the remainder of our discussion. Similarly, for SupMZ, weighted KS, and weighted CvM we also use the randomization method to recover the distribution of the test statistic under the null hypothesis. Second, the results for all three distributional-based tests (ISD^*-R, weighted KS, weighted CvM) are surprisingly powerful for $T = 70$. Third, the results are consistent across all sample sizes. Fourth, the power of all three distributional tests is consistent with the true ISD in that the greater the true ISD, the more power the test has independent of which moment the structural change actually exists.

[10] Results for other sample sizes are structurally similar and located in the online supplementary material.

Table 2. Rejection rates from the proposed test (randomization and asymptotic), the Bai–Perron test, the sup-type 'MZ' test ('SupMZ'), the weighted KS test ('weighted KS') and the weighted Cramér–von Mises test ('weighted CvM') at sample sizes of 70 and 1,000

$g(y)$		True ISD	ISD^*-R	ISD^*-A	Bai–Perron	SupMZ	Weighted KS	Weighted CvM
(a) Rejection rates ($T=70$)								
I	g_1	0.000	0.114	0.111	0.044	0.158	0.036	0.025
	g_2	0.125	0.898	0.259	0.896	0.266	0.938	**0.975**
	g_3	0.505	**1.000**	0.881	**1.000**	0.990	**1.000**	**1.000**
	g_4	0.563	**1.000**	0.470	**1.000**	**1.000**	**1.000**	**1.000**
	g_5	0.564	**1.000**	0.456	**1.000**	**1.000**	**1.000**	**1.000**
II	g_1	0.000	0.122	0.089	0.058	0.118	0.056	0.035
	g_2	0.066	**0.894**	0.125	0.054	0.844	0.242	0.500
	g_3	0.159	**1.000**	0.210	0.092	**1.000**	0.850	**1.000**
	g_4	0.218	**1.000**	0.263	0.132	**1.000**	**1.000**	**1.000**
	g_5	0.250	**1.000**	0.304	0.112	**1.000**	**1.000**	**1.000**
III	g_1	0.000	0.128	0.115	0.060	0.102	0.080	0.050
	g_2	0.008	**0.164**	0.132	0.058	0.118	0.050	0.075
	g_3	0.023	**0.350**	0.143	0.042	0.144	0.126	0.065
	g_4	0.055	**0.698**	0.172	0.054	0.200	0.226	0.180
	g_5	0.135	**1.000**	0.272	0.056	0.244	0.684	0.645
IV	g_1	0.000	0.158	0.115	0.036	0.110	0.046	0.030
	g_2	0.007	0.104	**0.132**	0.048	0.046	0.040	0.065
	g_3	0.013	**0.194**	0.139	0.026	0.088	0.076	0.070
	g_4	0.018	**0.184**	0.141	0.054	0.102	0.074	0.065
	g_5	0.022	**0.196**	0.163	0.048	0.098	0.078	0.060
(b) Rejection rates ($T=1{,}000$)								
I	g_1	0.000	0.046	0.021	0.052	0.140	0.060	0.035
	g_2	0.125	**1.000**	0.995	**1.000**	**1.000**	**1.000**	**1.000**
	g_3	0.505	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	g_4	0.563	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	g_5	0.564	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
II	g_1	0.000	0.058	0.016	0.052	0.124	0.064	0.045
	g_2	0.066	**1.000**	0.662	0.056	**1.000**	**1.000**	**1.000**
	g_3	0.159	**1.000**	**1.000**	0.088	**1.000**	**1.000**	**1.000**
	g_4	0.218	**1.000**	**1.000**	0.070	**1.000**	**1.000**	**1.000**
	g_5	0.250	**1.000**	**1.000**	0.078	**1.000**	**1.000**	**1.000**
III	g_1	0.000	0.056	0.020	0.044	0.120	0.054	0.050
	g_2	0.008	**0.624**	0.154	0.056	0.122	0.246	0.215
	g_3	0.023	**1.000**	0.458	0.056	0.150	0.958	**1.000**
	g_4	0.055	**1.000**	0.646	0.052	0.248	**1.000**	**1.000**
	g_5	0.135	**1.000**	0.964	0.036	0.184	**1.000**	**1.000**
IV	g_1	0.000	0.062	0.025	0.046	0.102	0.042	0.035
	g_2	0.007	**0.214**	0.104	0.048	0.170	0.090	0.007
	g_3	0.013	**0.824**	0.240	0.026	0.196	0.392	0.370

(*continued*)

Table 2. Continued

$g(y)$	True ISD	$ISD^{*}-R$	$ISD^{*}-A$	Bai–Perron	SupMZ	Weighted KS	Weighted CvM
g_4	0.018	**0.878**	0.256	0.054	0.236	0.538	0.550
g_5	0.022	**0.922**	0.406	0.048	0.262	0.658	0.780

Note. The first two columns display the simulated distributions as detailed in Table 1. The third column indicates the true ISD between *f* and *g*. The remaining columns indicate, out of 1,000 simulations, how frequently each test rejects the null hypothesis. The highest rejection rate in cases where null hypothesis is false is highlighted in bold.

Focusing on structural change in each specific moment, we find more results worth noting. First, with respect to structural change in the first moment only, $ISD^{*}-R$ performs very similarly to that of the Bai–Perron test for all sample sizes. Second, with respect to the second moment, $ISD^{*}-R$ performs very similarly to SupMZ and of course Bai–Perron performs quite poorly (again for all sample sizes). Third, with respect to structural change in the third moment, for all sample sizes $ISD^{*}-R$ outperforms weighted KS and weighted CvM, while Bai–Perron and SupMz perform poorly. Finally, with respect to the fourth moment, $ISD^{*}-R$ again outperforms weighted KS and weighted CvM. Overall, we find the simulation results as expected and quite encouraging for our proposed test.

We also evaluate to what extent the proposed test can recover the true break. Recall that the true break is set in the middle of the sample ($B = T/2$). Two criteria are used to assess the accuracy of identifying the break. Criterion 1 checks if $t^{*} = B$ when the null is rejected where t^{*} is the estimated break. Criterion 2 calculates the mean squared error (MSE) of estimated break. Results are presented in Table 3. By criterion 1, the estimated break is equal to the true break in many cases and not surprisingly tends to drops off as the ISD decreases. Similarly, the MSE (criterion 2) between the estimated break and the true break tends to decrease as the ISD increases. Both results are as expected. Interestingly, as the sample size increases we would also expect that the two criteria would improve. While they tend to, they do not to the extent one may expect because of two competing forces. On the one hand, an increase in sample size tends to increase the accuracy of the two estimated densities and in turn the accuracy of the ISD measure and thus the supremum of that measure. This would improve the two criteria as sample size increases. On the other hand, as the sample size increases the ISD measure while more accurate, tails off from its supremum at a much slower rate. This is because the contamination of the one estimated density with a fixed number of observations from the other density is becoming less as the sample size grows. That is, one contaminated observation represents 2.86% for a sample of size 35 versus only 0.20% for a sample of size 500. Therefore, ISD tails off slower from its supremum (as measured by number of points away from the true break) as the sample size grows. This would worsen the two criteria as sample size increases. These two competing effects are common in predicting structural break points. Therefore, the results in Table 3 are as expected.

4 Application to crop yields

In this section, the proposed test is applied to county data for corn, soybean, and winter wheat yields in the U.S. Of the roughly 250 million U.S. crop acres in 2022, corn and soybeans account for roughly 35% each and wheat accounts for roughly 20%. Furthermore, in 2022 there were 1.5 million insurance policies purchased for these crops carrying a total premium of $12 billion and a total liability of $122 billion. Corn serves both as a food staple and as livestock feed. The U.S. is the largest global producer of corn; in 2022, it accounted for 353 million metric tonnes of 1,169 million metric tonnes produced globally. Similarly, the U.S. is the world's largest producer and second largest exporter of soybean. In 2022, the U.S. produced 120 million metric tons of the 354 million metric tons worldwide. We include wheat since it is the third most important U.S. field crop in terms of area planted and volume produced. Planted area for winter wheat was around 34 million acres in 2022.

As with most of the literature, we use county-level yield data from the United States Department of Agriculture (USDA) National Agricultural Statistics Service. The most complete data are

Table 3. Accuracy of predicting break using the proposed test at sample size 70 and 1,000

Sample Size		True ISD	Criterion 1		Criterion 2	
			$T = 70$	$T = 1{,}000$	T=70	$T = 1{,}000$
I	g_1	0.000	n/a	n/a	n/a	n/a
	g_2	0.125	22.5%	26.6%	7.64	6.47
	g_3	0.505	74.6%	68.6%	1.00	1.46
	g_4	0.563	95.4%	87.2%	0.21	0.32
	g_5	0.564	96.4%	98.2%	0.21	0.12
II	g_1	0.000	n/a	n/a	n/a	n/a
	g_2	0.066	21.9%	22.9%	7.42	7.74
	g_3	0.159	39.2%	57.0%	3.99	3.15
	g_4	0.218	48.2%	56.8%	4.09	2.43
	g_5	0.250	45.0%	57.3%	4.66	2.41
III	g_1	0.000	n/a	n/a	n/a	n/a
	g_2	0.008	8.5%	7.7%	11.62	13.18
	g_3	0.023	11.4%	14.0%	8.65	10.01
	g_4	0.055	17.5%	20.8%	8.24	7.44
	g_5	0.135	39.4%	41.8%	4.48	3.42
IV	g_1	0.000	n/a	n/a	n/a	n/a
	g_2	0.007	7.7%	9.3%	11.02	10.87
	g_3	0.013	6.1%	11.2%	12.50	12.36
	g_4	0.018	10.3%	8.2%	11.81	11.96
	g_5	0.022	12.9%	12.8%	10.49	10.78

Note. The first two columns display the simulated distributions as detailed in Table 1. The third column indicates the true ISD between f and g. The forth and fifth columns indicate, out of 1,000 simulations, how frequently the estimated break, t^*, is the same as the true break at sample size 70 and 1,000, respectively (Criterion 1). The sixth and seventh columns display the mean squared error of estimated break at sample size 70 and 1,000, respectively (Criterion 2).

available from 1950 to 2021. To be included in the analysis, counties had to have a complete set of data. In total, our data set consists of 901 crop-county combinations. We employ USDA-RMA two-step methodology (detailed in Liu & Ker, 2021) in rating their area crop insurance programs. Their approach detrends yields using a two-knot linear spline.[11] After estimating the trend, the degree of heteroskedasticity is estimated using Harri et al. (2011) and crop yields are adjusted with the estimated degree of heteroskedasticity.[12]

A number of caveats are worth discussing with respect to the adjusted yield data. First, to avoid endpoint issues, the set of possible breaks removes the first and final 10 years. Second, in addition to evaluating the entire domain of the yield distribution, we restrict our proposed test over subsets of the domain that are of particular interest for insurance purposes. Our proposed test statistic is easily restricted to a given subset of the support by simply taking the integral of the squared difference over that subset rather than taking the integral over entire support. Specifically, defining the expected yield as y^e, we take the integrals over subsets $(0, .7y^e)$, $(0, .8y^e)$, $(0, .9y^e)$, and $(0, y^e)$. These choices, referred to as 'lower tail' and 'λ lower tail' ($\lambda = .7, .8, .9$), represent the standard coverage levels for crop insurance. For comparison purposes, the proposed test is symmetrically applied to upper tail as well.

[11] Equation for two-knot linear spline: $y_t = \alpha_1 + \alpha_2 t + \delta_1 d_1(t - k_1) + \delta_2 d_2(t - k_2) + \epsilon_t$, where y_t is crop yield at time $t \in [1, \ldots, T]$; k_1 and k_2 represents knots on which certain restrictions are imposed to prevent knot positions from being too close to the end points and each other; d_1 and d_2 are two indicator functions where $d_1 = 1$ if $t \geq k_1$ and $d_2 = 1$ if $t \geq k_2$.

[12] The degree of heteroskedasticity is estimated by: $ln(\hat{\epsilon}_t^2) = \beta + \gamma ln(\hat{y}_t) + v_t$, where $\hat{\epsilon}_t$ and $\hat{y}_t$ are residuals and fitted values from the spline estimation. $\hat{\gamma}$ represents the degree of heteroskedasticity.

Table 4. Structural change test results

	Above the first moment			Above first two moments		
Yield support evaluated	Corn	Soybean	Winter wheat	Corn	Soybean	Winter wheat
Entire	374	189	76	187	84	37
Lower tail	345	170	74	181	60	36
0.9 Lower tail	268	126	67	178	49	37
0.8 Lower tail	131	75	58	140	38	30
0.7 Lower tail	53	46	47	75	28	24
Upper tail	335	177	72	162	89	33
0.9 Upper tail	300	164	64	156	86	30
0.8 Upper tail	282	142	50	152	81	25
0.7 Upper tail	281	127	44	150	63	26
Total	442	335	124	442	335	124

Note. Column 1 depicts the support interval that the structural change test is calculated over. Columns 2–4 show test results after adjusting crop yields for differences in the first moment. Columns 5–7 show test results after adjusting crop yields for differences in the first two moments.

Table 4 presents the number of rejections at the 5% significance level. Randomization method is used for finding the rejection critical threshold as discussed in Section 3. The first set of three columns represent tests on the yield data correcting only for the temporal process (first moment). The number represents the rejections of the null (i.e. no structural change). The second set of the three columns represents tests on the yield data correcting for the temporal process and heteroscedasticity (first two moments). The number again represents the rejections of the null (i.e. no structural change). Given that the first set of three columns test for structural change beyond the conditional mean while the second set of three columns test for structural change beyond the conditional mean and variance, the number of rejections is necessarily smaller. The rows represent different domains over which the proposed test is calculated.

In general, we find a surprising number of rejections given the relatively small sample size (1950–2021).[13] A number of interesting points are worth noting. First, when evaluating at the entire yield range, we find strong evidence of structural change above the first moment in corn yields (84.6%), followed by wheat (61.3%), and soybean (56.4%). As expected, the number of rejections notably decreases when we consider structural change beyond first two moments, but the numbers still support the existence of structural change beyond mean and variance in corn yields (42.3%), wheat (29.8%), and soybean (25.1%). Second, the number of rejections decreases as the support over which the test statistic is calculated shrinks. This is not surprising given the power of the test will decrease. Also, the upper and lower tail rejections are roughly equivalent. Third, when we plot the p-values and rejection decisions for corn yields as illustrated in Figure 2 (plots for soybean and wheat are in online supplementary material), there does appear to be spatial correlation among the rejections. Grey counties are those without data. Finally, with hundreds of county-level crop yields being tested, there are concerns regarding the multiple testing issue. We have implemented the Holm–Bonferroni correction (Holm, 1979) for the purpose of multiple testing adjustment. As anticipated, this adjustment has led to a reduction in the count of counties for which the null hypothesis is rejected. For example, with respect to corn that had 374 rejections for the entire distribution, after the correction that number drops to 279. The results when the Holm–Bonferroni correction is implemented are presented in online supplementary material. Overall, the results consistently substantiate our conclusion that structural change beyond the mean has taken place in crop yield distributions.

[13] For comparison, we also employed the weighted KS test and weighted CvM test (Inoue, 2001) to assess structural change beyond the first two moments in our crop yield data (see online supplementary material). These tests yield fewer rejections than our proposed test. However, the relative number of rejections can not be interpreted as the presence or lack thereof of structural change is unknown with real yield data.

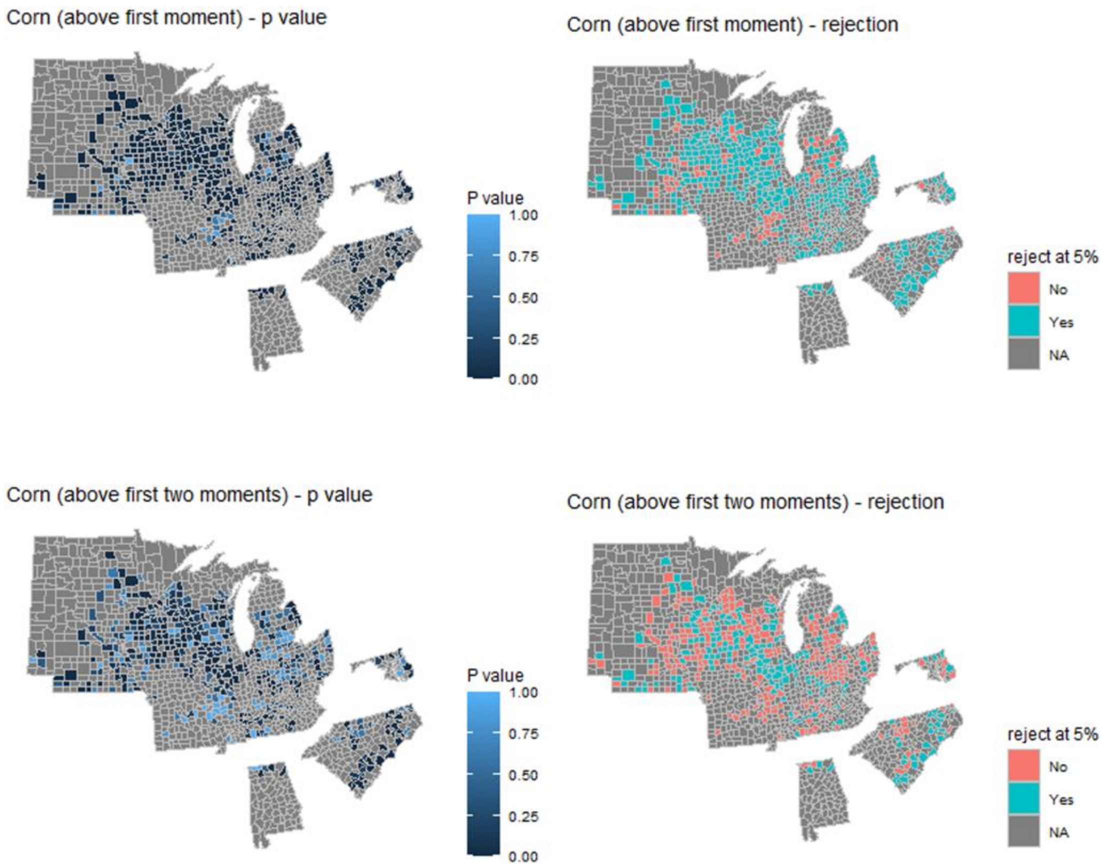

Figure 2. *P*-values and rejection decisions for county-level corn yields. *Note*. Structural change test above first moment (top 2 plots) and above first two moments (bottom 2 plots).

The locations of the estimated breaks can be easily recovered from the proposed test. Figure 3 shows when structural change is found given testing across the entire support of yield distribution.[14] For corn and soybean but markedly less so for wheat, the estimated breaks are found near the beginning of the sample period. This is likely caused by increasing adoption of hybrid varieties for corn and soybean. Conversely, there is very little production of hybrid wheat. Interestingly, this raises the question of multiple break points. To estimate a second break, we focus on counties which have a break estimated around 1965, 1985, and 2000 so that there are enough observations on both sides of the possible break for estimation. Of the 164 corn counties found to have a structural break, a second break is found in 53 of them. Of the 62 soybean counties found to have a structural break, a second break is found in 21 of them. Finally, of the 31 wheat counties found to have a structural break, a second break is found in 10 of them.

5 Implications for rating crop insurance contracts

The motivation for the proposed test is with respect to rating crop insurance contracts. We illustrate the impact of structural change on the resulting premium rates by calculating premium rates based on the entire historical yield series and those based on the historical yield data post the structural change. Again, we use the USDA-RMA methodology to adjust the yield series and recover the premium rates. We consider both the 90% and 70% coverage levels. The absolute percentage difference in the premium rates are depicted in Figure 4; rate differences for all three crops and testing for structural change above the first and first two conditional moments are illustrated. The results are notable because of the sheer magnitude. For all three crops, the premium rate differences between the historical data and the historical data post structural change are substantial; the median

[14] Online supplementary material illustrates the results when we subset to specified regions. As expected, the results do not change markedly.

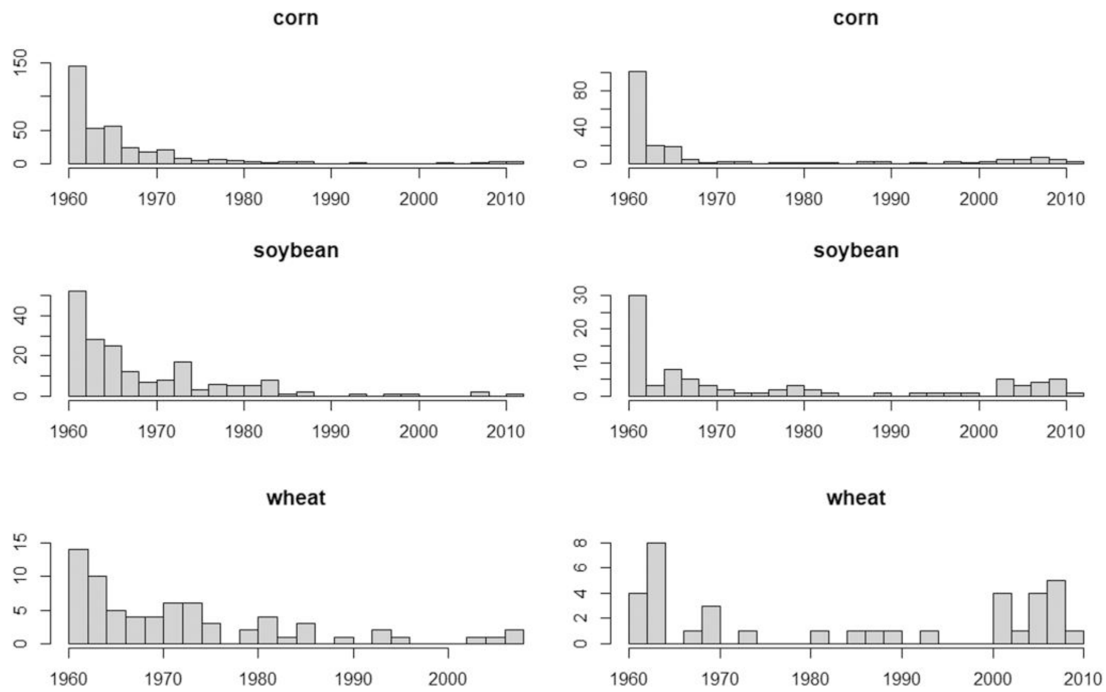

Figure 3. Estimated breaks. Plots on the left side are based on yield data adjusted for the first moment only. Plots on the right side are based on yield data adjusted for both first two moments.

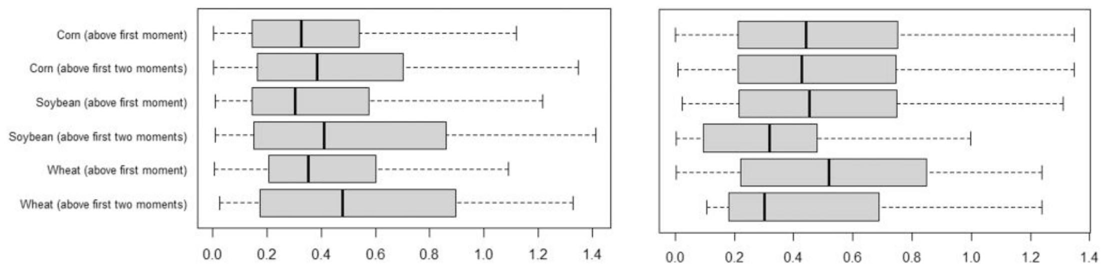

Figure 4. Empirical premium rate differences at 90% coverage level (left) and 70% coverage level (right).

rate differences are between 30% and 50%. For a multi-billion dollar program, that represents a significant transfer of monies between farmers, insurance companies, and the government.

The impact of structural change on the resulting premium rates can be evaluated by a repeated game of out-of-sample rating accuracy designed by Ker McGowan (2000) and detailed therein. Essentially, two competing methodologies, denote A and B, are pitted against each other in a repeated out-of-sample rating game of insurance contracts. Methodology A retains all contracts that it believes methodology B has over-priced and thus expects to make money. Conversely, methodology A cedes back to B any contracts that it believes methodology B has under-priced and expects to lose money. That is, methodology A retains all contracts whose premium rates are less than methodology B (over-priced contracts) and cedes all contracts whose premium rates are greater than methodology B (under-priced contracts). Only yield information from prior is used to estimate the premium rates for the competing methodologies A and B so that an out-of-sample metric not an in-sample metric is employed. The loss ratios (total losses over total premiums) of the set of retained and ceded contracts are recovered using actual realized yields. Randomization methods are used to recover the null for the out-of-sample metric. This game has been widely used in the crop insurance literature to compare competing rating methodologies (Annan et al., 2014; Harri et al., 2011; Ker & Coble, 2003; Ker & Tolhurst, 2019; Ker et al., 2016; Liu & Ramsey, 2023; Park et al., 2019, 2022; Ramsey, 2020; Tolhurst & Ker, 2015).

Table 5. Out-of-sample rating game

Crop	Number of counties	Retained (%)	Loss ratio (G)	Loss ratio (P)	*p*-value	Efficacy test
At 90% Coverage level						
Corn	442	34.19	1.20	1.03	0.0000	0.0005
Soybean	335	46.98	1.53	0.48	0.0000	0.0000
Winter wheat	124	40.00	2.06	1.22	0.1598	0.0005
At 70% Coverage level						
Corn	442	40.43	0.85	0.53	0.0000	0.0005
Soybean	335	69.46	0.41	0.27	0.0950	0.0005
Winter wheat	124	41.82	3.24	1.45	0.0850	0.0037

In our application of the game, we use identical rating methodologies (the current RMA methodology) but one set of rates is based on the entire yield series (methodology B) while the other is based on yield data only post structural break found via our proposed test (methodology A). That is, premium rates in 2007 are estimated using data from 1950 to 2006 for methodology B and structural break to 2006 for methodology A (assuming a structural break exists). The actual loss in 2007 is calculated by using the corresponding realized 2007 yields. The game is repeated for 2008,…,2020. To account for the first mover advantage, a second test of efficacy is conducted by considering private gains of methodology A relative to gains methodology B would make if their roles were reversed. This second test is detailed in Ker et al. (2016).

Table 5 presents the results of the out-of-sample rating accuracy game and the efficacy test for coverage 90% and 70% coverage levels and all three crops. In all cases, loss ratios for methodology A (rates based on structural change tests) are less than methodology B (rates based on entire yield series) and statistically significant in five of six cases.[15] This suggests that money can be made by private insurers adverse selecting against the government based on premium rates from truncated (as per our proposed test) yield series. Furthermore, the efficacy tests suggest that rates based on yield data post structural change (as identified by our proposed test) are more accurate than rates based on the entire yield series. These results are very encouraging for our proposed test as they are based on an out-of-sample metric tailored for insurance purposes.

6 Conclusion

Tests for structural change with known and unknown break(s) in the first (Chow, 1960) and (Bai & Perron, 1998) and second moment (Esfandiar et al., 2010) abound. However, insurance is primarily interested in lower tail probabilities and as such the detection of structural change in tail probabilities or higher moments is of great concern for the pricing and efficacy of insurance programs. In this article, we proposed a test for structural change with an unknown break(s) which has power against structural change in any moment or tail. Intuitively, we combined a sup-type measure with nonparametric tests for distributional equivalence. Furthermore, our proposed test is both easy to apply and can be tailored to concentrate on a specific range of the underlying distribution if desired (i.e. subsection of the lower tail for insurance purposes).[16] The asymptotic distribution is shown to follow the Kolmogorov distribution although in practice simulating the null based on the randomization method performs better. Our simulations demonstrate that the proposed test not only has power against structural change in higher moments, but also performs reasonably well at identifying the break.

[15] Rather than compare our proposed test to no structural change in the rating game, we can explicitly test our proposed test against that of the weighted KS and weighted CvM tests. The results (see online supplementary material) are encouraging in that our test outperforms the weighted KS and weighted CvM tests.

[16] Note that while the proposed test can be used to detect structural change in the extreme tails of a distribution, it is likely to perform quite poorly. Our test statistic is based on kernel estimates which are entirely data-driven. These estimators do not perform well at estimating extreme tail probabilities where there little to no data.

The impetus for developing this structural change test was rating crop insurance contracts with historical yield/loss data in the presence of changing technology and climate. In the light of the sizeable resources directed toward these programs and its importance to the continued viability of production agriculture throughout the world and thus food security, accurate estimation of premium rates for crop insurance contracts is of utmost importance to consumers, farmers, insurance companies, and governments. Note, publicly subsidized crop insurance is the dominant farm policy in most of the developed world. Applying our proposed test to major U.S. field crop yields (corn, soybean, winter wheat), we find structural change above the first and second moments in 70.9% and 34.2% of the counties with a median absolute premium rate difference of around 35% and 50%, respectively. Given crop insurance is a multi-billion dollar program, these premium rate differences represent significant public dollars. Although we have applied our proposed test to major U.S. crop yields, its application is of general interest as it can be applied to any type of crop yield data, other forms of insurance loss data (i.e. property, casualty, etc.), and even financial market data and climate data.

Finally, there are a number of avenues to consider in further developing the proposed test. First, relaxing the independently distributed assumption to allow strong mixing processes. This likely will require block sampling methods in the test statistic. Second, extending to multivariate data structures for both continuous and mixed data. This will require the use of product kernels. Third, correcting the over-rejection bias in smaller samples when using the randomization method. This likely will require the use of an adaptive smoothing parameter. Finally, consider a smooth/continual transition in structural change. This may require discarding a subset of the data.

Conflict of interest: The authors have no relevant financial or non-financial interests to disclose.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this article.

Data availability

The data that support the findings of this study are openly available in United States Department of Agriculture - National Agricultural Statistics Service at https://quickstats.nass.usda.gov/.

Supplementary material

Supplementary material is available online at *Journal of the Royal Statistical Society: Series C*.

References

Anderson T. W. (1962). On the distribution of the two-sample Cramér-von Mises criterion. *The Annals of Mathematical Statistics*, *33*(3), 1148–1159. https://doi.org/10.1214/aoms/1177704477

Anderson T. W., & Darling D. A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *The Annals of Mathematical Statistics*, *23*(2), 193–212. https://doi.org/10.1214/aoms/1177729437

Andrews D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, *61*(4), 821–856. https://doi.org/10.2307/2951764

Andrews D. W. K. (2003). Tests for parameter instability and structural change with unknown change point: A corrigendum. *Econometrica*, *71*(1), 395–397. https://doi.org/10.1111/1468-0262.00405

Annan F., Tack J., Harri A., & Coble K. (2014). Spatial pattern of yield distributions: Implications for crop insurance. *American Journal of Agricultural Economics*, *96*(1), 253–268. https://doi.org/10.1093/ajae/aat085

Bai J. (2010). Common breaks in means and variances for panel data. *Journal of Econometrics*, *157*(1), 78–92. https://doi.org/10.1016/j.jeconom.2009.10.020

Bai J., Han X., & Shi Y. (2020). Estimation and inference of structural changes in high dimensional factor models. *Journal of Econometrics*, *219*(1), 66–100. https://doi.org/10.1016/j.jeconom.2019.08.013

Bai J., & Perron P. (1998, January). Estimating and testing linear models with multiple structural changes. *Econometrica*, *66*(1), 47–78. https://doi.org/10.2307/2998540

Bai J., & Perron P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, *18*(1), 1–22. https://doi.org/10.1002/jae.659

Baltagi B. H., Feng Q., & Kao C. (2016). Estimation of heterogeneous panels with structural breaks. *Journal of Econometrics*, *191*(1), 176–195. https://doi.org/10.1016/j.jeconom.2015.03.048
Bretagnolle J., & Massart P. (1989). Hungarian constructions from the nonasymptotic viewpoint. *The Annals of Probability*, *17*(1), 239–256. https://doi.org/10.1214/aop/1176991506
Chan N. H., Yau C. Y., & Zhang R.-M. (2014). Group LASSO for structural break time series. *Journal of the American Statistical Association*, *109*(506), 590–599. https://doi.org/10.1080/01621459.2013.866566
Chow G. C. (1960, July). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, *28*(3), 591–605. https://doi.org/10.2307/1910133
Cramér H. (1928). On the composition of elementary errors. *Scandinavian Actuarial Journal*, *1928*(1), 13–74. https://doi.org/10.1080/03461238.1928.10416862
Cucconi O. (1968). Un nuovo test non parametrico per il confronto fra due gruppi di valori campionari. *Giornale degli Economisti e Annali di Economia*, *27*, 225–248. https://www.jstor.org/stable/23241361
Darling D. (1957). The Kolmogorov-Smirnov, Cramer-von Mises test. *Annals of Mathematical Statistics*, *28*(4), 823–838. https://doi.org/10.1214/aoms/1177706788
Dudley R. M. (2002). *Real analysis and probability*. Cambridge University Press.
Esfandiar M., Asad Z., & Mumtaz A. (2010). Tests for structural change, aggregation, and homogeneity. *Economic Modelling*, *27*(6), 1382–1391. https://doi.org/10.1016/j.econmod.2010.07.009
Gagliardini P., Trojani F., & Urga G. (2005). Robust GMM tests for structural breaks. *Journal of Econometrics*, *129*(1-2), 139–182. https://doi.org/10.1016/j.jeconom.2004.09.006
Ghysels E., Guay A., & Hall A. (1998, February). Predictive tests for structural change with unknown breakpoint. *Journal of Econometrics*, *82*(2), 209–233. https://doi.org/10.1016/S0304-4076(97)00057-2
Goodwin B. K., & Hungerford A. (2015). Copula-based models of systemic risk in U.S. agriculture: Implications for crop insurance and reinsurance contracts. *American Journal of Agricultural Economics*, *97*(3), 879–896. https://doi.org/10.1093/ajae/aau086
Goodwin B. K., Roberts M. C., & Coble K. H. (2000). Measurement of price risk in revenue insurance: Implications of distributional assumptions. *Journal of Agricultural and Resource Economics*, *25*(1), 195–214. https://doi.org/10.22004/ag.econ.30830
Greene W. H. (2012). *Econometric analysis* (7th ed.). Pearson.
Hall A., & Sen A. (1999, 7). Structural stability testing in models estimated by generalized method of moments. *Journal of Business and Economic Statistics*, *17*(3), 335–348. https://doi.org/10.1080/07350015.1999.10524822
Hall A., & Sen A. (1999, December). Two further aspects of some new tests for structural stability. *Structural Change and Economic Dynamics*, *10*(3-4), 431–443. https://doi.org/10.1016/S0954-349X(99)00011-9
Hansen B. E. (1992). Tests for parameter instability in regressions with I(1) processes. *Journal of Business and Economic Statistics*, *10*, 321–335. https://doi.org//10.2307/1391545
Harchaoui Z., & Lévy-Leduc C. (2010). Multiple change-point estimation with a total variation penalty. *Journal of the American Statistical Association*, *105*(492), 1480–1493. https://doi.org/10.1198/jasa.2010.tm09181
Harri A., Coble K. H., Ker A. P., & Goodwin B. J. (2011). Relaxing heteroscedasticity assumptions in area-yield crop insurance rating. *American Journal of Agricultural Economics*, *93*(3), 707–717. https://doi.org/10.1093/ajae/aar009
Harvey D. I., Leybourne S. J., & Taylor A. R. (2009). Simple, robust, and powerful tests of the breaking trend hypothesis. *Econometric Theory*, *25*(4), 995–1029. https://doi.org/10.1017/S0266466608090385
Holm S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70. https://www.jstor.org/stable/4615733
Inoue A. (2001). Testing for distributional change in time series. *Econometric Theory*, *17*(1), 156–187. https://doi.org/10.1017/S0266466601171057
Kejriwal M., & Perron P. (2010). Testing for multiple structural changes in cointegrated regression models. *Journal of Business and Economic Statistics*, *28*(4), 503–522. https://doi.org/10.1198/jbes.2009.07220
Ker A. P., & Coble K. (2003). Modeling conditional yield densities. *American Journal of Agricultural Economics*, *85*(2), 291–304. https://doi.org/10.1111/1467-8276.00120
Ker A. P., & McGowan P. (2000). Weather-based adverse selection and the U.S. crop insurance program: The private insurance company perspective. *Journal of Agricultural and Resource Economics*, *25*, 386–410. https://doi.org/10.22004/ag.econ.30907
Ker A. P., & Tolhurst T. N. (2019). On the treatment of heteroscedasticity in crop yield data. *American Journal of Agricultural Economics*, *101*(4), 1247–1261. https://doi.org/10.1093/ajae/aaz004
Ker A. P., Tolhurst T. N., & Liu Y. (2016). Bayesian estimation of possibly similar yield densities: Implications for rating crop insurance contracts. *American Journal of Agricultural Economics*, *98*(2), 360–382. https://doi.org/10.1093/ajae/aav065
Kobayashi M. (1986). A bounds test of equality between sets of coefficients in two linear regressions when disturbance variances are unequal. *Journal of the American Statistical Associations*, *81*(394), 510–513. https://doi.org/10.1080/01621459.1986.10478297

Komlóos J., Major P., & Tusnáady G. (1975). An approximation of partial sums of independent RV'-s, and the sample DF. I. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, *32*(1-2), 111–131. https://doi.org/10.1007/BF00533093
Kuiper N. H. (1960). Tests concerning random points on a circle. *Indagationes Mathematicae (Proceedings)*, *63*, 38–47. https://doi.org/10.1016/S1385-7258(60)50006-0
Kurozumi E. (2017). Monitoring parameter constancy with endogenous regressors. *Journal of Time Series Analysis*, *38*(5), 791–805. https://doi.org/10.1111/jtsa.12236
Lepage Y. (1971). A combination of Wilcoxon's and Ansari-Bradley's statistics. *Biometrika*, *58*(1), 213–217. https://doi.org/10.1093/biomet/58.1.213
Li D., Qian J., & Su L. (2016). Panel data models with interactive fixed effects and multiple structural breaks. *Journal of the American Statistical Association*, *111*(516), 1804–1819. https://doi.org/10.1080/01621459.2015.1119696
Li H., Zhou J., & Hong Y. (2021, 10). Estimating and testing for smooth structural changes in moment condition models. (SSRN) https://ssrn.com/abstract=3792747.
Li Q. (1996). Nonparametric testing of closeness between two unknown distribution functions. *Econometric Reviews*, *15*(3), 261–274. https://doi.org/10.1080/07474939608800355
Li Q., Maasoumi E., & Racine J. S. (2009). A nonparametric test for equality of distributions with mixed categorical and continuous data. *Journal of Econometrics*, *148*(2), 186–200. https://doi.org/10.1016/j.jeconom.2008.10.007
Li Q., & Racine J. S. (2007). *Nonparametric econometrics: Theory and practice*. Princeton University Press.
Liu Y., & Ker A. P. (2021). Simultaneous borrowing of information across space and time for pricing insurance contracts: An application to rating crop insurance policies. *Journal of Risk and Insurance*, *88*(1), 231–257. https://doi.org/10.1111/jori.12312
Liu Y., & Ramsey A. F. (2023). Incorporating historical weather information in crop insurance rating. *American Journal of Agricultural Economics*, *105*(2), 546–575. https://doi.org/10.1111/ajae.12329
Mann H. B., & Whitney D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, *18*(1), 50–60. https://doi.org/10.1214/aoms/1177730491
Mieno T., Walters C. G., & Fulginiti L. E. (2018). Input use under crop insurance: The role of actual production history. *American Journal of Agricultural Economics*, *100*(5), 1469–1485. https://doi.org/10.1093/ajae/aay040
Mises R. V. (1972). *Wahrscheinlichkeit, statistik und wahrheit*. Springer Vienna.
Mumtaz A., Gulfam H., & Asad Z. (2016). Detecting structural change with heteroskedasticity. *Communications in Statistics–Theory and Methods*, *46*(21), 10446–10455. https://doi.org/10.1080/03610926.2016.1235200
Oka T., & Perron P. (2018). Testing for common breaks in a multiple equations system. *Journal of Econometrics*, *204*(1), 66–85. https://doi.org/10.1016/j.jeconom.2018.01.003
Oka T., & Qu Z. (2011). Estimating structural changes in regression quantiles. *Journal of Econometrics*, *162*(2), 248–267. https://doi.org/10.1016/j.jeconom.2011.01.005
Page E. S. (1954, June). Continuous inspection schemes. *Biometrika*, *41*(1-2), 100–115. https://doi.org/10.1093/biomet/41.1-2.100
Park E., Brorsen B. W., & Harri A. (2019). Using Bayesian Kriging for spatial smoothing in crop insurance rating. *American Journal of Agricultural Economics*, *101*(1), 330–351. https://doi.org/10.1093/ajae/aay045
Park E., Harri A., & Coble K. H. (2022). Estimating crop yield densities for counties with missing data. *Journal of Agricultural and Resource Economics*, *47*, 634–655. https://doi.org/10.22004/ag.econ.313319
Perron P., & Yabu T. (2009). Testing for shifts in trend with an integrated or stationary noise component. *Journal of Business and Economic Statistics*, *27*(3), 369–396. https://doi.org/10.1198/jbes.2009.07268
Qian J., & Su L. (2016). Shrinkage estimation of common breaks in panel data models via adaptive group fused Lasso. *Journal of Econometrics*, *191*(1), 86–109. https://doi.org/10.1016/j.jeconom.2015.09.004
Qu Z. (2008). Testing for structural change in regression quantiles. *Journal of Econometrics*, *146*(1), 170–184. https://doi.org/10.1016/j.jeconom.2008.08.006
Ramsey A. F. (2020). Probability distributions of crop yields: A bayesian spatial quantile regression approach. *American Journal of Agricultural Economics*, *102*(1), 220–239. https://doi.org/10.1093/ajae/aaz029
Tolhurst T. N., & Ker A. P. (2015). On technological change in crop yields. *American Journal of Agricultural Economics*, *97*(1), 137–158. https://doi.org/10.1093/ajae/aau082
Wilcoxon F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, *1*(6), 80–83. https://doi.org/10.2307/3001968
Woodard J. D., & Sherrick B. J. (2011). Estimation of mixture models using cross-validation optimization: Implications for crop yield distribution modeling. *American Journal of Agricultural Economics*, *93*(4), 968–982. https://doi.org/10.1093/ajae/aar034
Xu J., & Perron P. (2017). Forecasting in the presence of in and out of sample breaks. (IDEAS Working Paper Series from RePEc).