
# Is There Too Much History in Historical Yield Data

*Y. Liu[1]; A. Ker[2]*

*1: University of Guelph, The Department of Food, Agricultural & Resource Economics, Canada, 2: University of Guelph, Food, Agricultural and Resource Economics, Canada*

*Corresponding author email: yliu17@uoguelph.ca*

***Abstract:***

*County crop yield data from USDA-NASS are extensively used in the literature as well as practice. In many applications, yield data are adjusted for the first two moments then assumed independent and identically distributed. For most major crop-region combinations, yield data exist from 1955 onwards and reflect significant innovations in both seed and farm management technologies. These innovations have likely changed the yield distribution raising doubt regarding the identically distributed assumption. We consider the question of how much historical yield data should be used in empirical analyses. First, we use distributional tests to assess if and when the adjusted yield data result from different DGPs. Second, we consider the application to crop insurance by using an out-of-sample rating. Third, we estimate DGPs and then simulate to quantify the additional error. Overall, the results indicate that using yield data more than 30 years old can substantially increase estimation error. Given that discarding data is unappetizing, we propose three methodologies that can re-incorporate the discarded data. Our results suggest gains in efficiency by using these methodologies. While our results are most applicable to the crop insurance literature, we certainly feel they suggest proceeding with caution when using historical yield data in other applications.*

*Acknowledegment:*

**JEL Codes:** C53, Q14

#1570

# Is There Too Much History in Historical Yield Data


Abstract

County crop yield data from United States Department of Agriculture - National Agricultural Statistics Service (USDA-NASS) has and continues to be extensively used in the literature as well as practice. The most notable example is crop insurance; the Risk Management Agency (RMA) uses the data to set guarantees, estimate premium rates, and calculate indemnities for their area programs. Examples from the literature include investigation of rating methodologies, issues related to land use, modeling the climate-yield relationship, and supply analysis. In many of these applications, and certainly with respect to RMA and the crop insurance literature, yield data are detrended and adjusted for possible heteroskedasticity and then assumed to be independent and identically distributed. For most major crop-region combinations, county yield data exist from 1955 onwards and reflect very significant innovations in both seed and farm management technologies. Despite correcting for movements in the first two moments of the yield data generating process (DGP), these innovations have likely moved mass all around the support of the yield distribution raising doubt regarding the identically distributed assumption. This manuscript considers the rather nebulous question of how much historical yield data should be used in empirical analyses. The answer is obviously dependent on the empirical application, crop-region combination, econometric methodology, and chosen loss function. Nonetheless, we attempt to tackle this question in three ways using county-level yield data for corn, soybean, and winter wheat. First, we use distributional tests to assess if and when the adjusted yield data may result from different DGPs. Second, we consider the application to crop insurance by using an out-of-sample rating game -- commonly employed in the literature -- to compare rates from the full versus restricted data sets. Third, we estimate flexible time-varying DGPs and then simulate to quantify the additional error when the identically distribution assumption is imposed. Overall, the results indicate that despite accounting for time-varying movements in the first two moments, using yield data more than 30 years old can substantially increase estimation error. Given that discarding data is unappetizing, particularly so in applications with relatively small T, we investigate three methodologies that can re-incorporate the discarded data while both explicitly acknowledging the unknown DGPs are different and not requiring knowledge about the extent or form of those differences. Our results suggest gains in efficiency can be realized by using these methodologies. While our results are most applicable to the crop insurance literature, we certainly feel they suggest proceeding with caution when using historical yield data in other applications as well.



*Some key words:* NASS yield data, changing technology, borrowing information, crop insurance

## Introduction

County crop yield data from United States Department of Agriculture - National Agricultural Statistics Service (USDA-NASS) has and continues to be extensively used in the literature as well as practice. The most notable example is crop insurance; the Risk Management Agency (RMA) uses the data to set insurance guarantees, estimate premium rates, and calculate indemnities for their area programs. Examples from the literature include investigation of rating methodologies, issues related to land use, modeling the climate-yield relationship, and supply analysis. In many of these applications, and certainly with respect to RMA and the crop insurance literature, yield data are detrended and adjusted for possible heteroskedasticity and then assumed to be independent and identically distributed. For most major crop-region combinations, county yield data exist from 1955 onwards and reflect a number of significant innovations in both seed and farm management technologies. Despite correcting for changes in the first two moments of the yield data generating process (DGP), these innovations have likely moved mass all around the support of the yield distribution raising doubt regarding the identically distributed assumption. This assumption is employed in not only the literature, but in the RMA rating methodology for their area yield and revenue programs. However, it is difficult to conceive that yield losses in the 1950s, 1960s, and even 1970s can inform anything about losses in 2018. Quite interestingly, RMA uses NASS data from 1955 and onwards to estimate premium rates for their area yield and revenue products but instead use data from 1991 and onwards to rate the newer area-based shallow loss products.[1]

To the best of our knowledge, the literature has used the entire series of yield data dating back to the 1950s in their empirical analyses. However, there has been a multitude of significant changes in the seed technology, farm management, and even climate. According to Reilly and Fuglie (1998); Fernandez-Cornejo et al. (2004); Duvick (2005); Egli (2008); Fernandez-Cornejo et al. (2014); Assefa et al. (2017); Egli (2017), average per acre yields for corn, soybean and wheat in the United States have more than doubled from 1950 to mid-1990s. More than half of average yield gains are attributed to genetic improvements and the other half have come from improved agronomic practices and other factors. At the same time, the intertwined combination of these factors, especially the changing climate, also increased the crop yields variability over time. Naylor, Falcon, and Zavaleta (1997) found that there were large and significant variations of corn yields in U.S. during the past decades. They also found that after reaching the yield ceiling, which was driven by the technological advancement in seeds and farming practice, poor weather were more likely to lead to significant yields loss. Reilly et al. (2003) found an increasing trend in yield variability from 1950 to 1994. The results from Kucharik and Ramankutty (2005) agreed with the increasing variability since 1950, although they found that county level yield variability has once again decreased since the mid-1980s. Challinor et al. (2014) pointed out that global crop yields variations are likely to increase in the near future. Leng (2017) found that the climate variability dominates the change of corn yield variability in the U.S. Midwest Corn

[1]RMA aggregates their individual farm data to the county level to construct a county yield series rather than use NASS yield data.

Belt from 1981-2010, and the yield variation was decreasing for some areas while increasing for others. All these findings are implying there are variations more than the first moment of yield distribution over time.

Meanwhile, GM technology have widely adopted by the farmers since 1996. The Genetically Modified crops (GMCs, GM crops, or biotech crops) are plants used in agriculture, the DNA of which has been modified using genetic engineering methods. In most cases, the aim is to introduce a new trait to the plant, which does not occur naturally in the species. Examples in food crops include resistance to certain pests, diseases, or environmental conditions, reduction of spoilage, or resistance to chemical treatments. In the US, by 2014, 94% of the planted area of soybeans, 96% of cotton and 93% of corn were genetically modified varieties (USDA, 2016).

A meta-analysis (Klümper and Qaim, 2014) considered all publications of the agronomic and economic impacts between 1995 and March 2014 for three major GM crops: corn, soybean and cotton. The study found that herbicide-tolerant crops have lower production costs, while for insect-resistant crops the reduced pesticide use was offset by higher seed prices, leaving overall production costs about the same. For example, corn and some other crops have been engineered to express genes encoding for insecticidal proteins from Bacillus thuringiensis (Bt). The introduction of Bt crops during the period between 1996 and 2005 has been estimated to have reduced the total volume of insecticide active ingredient use in the United States by over 100 thousand tons. This represents a 19.4% reduction in insecticide use (Naranjo et al., 2008). In the United States, corn with stacked Bt traits were planted on 140 million hectares in 29 different countries by 2010 (Barrows, Sexton, and Zilberman, 2014). The advantage of Bt corn over conventional seed has became more and more obvious over the time, providing a 2.3% increase in net returns on average and even more when the pest pressure is high (Fernandez-Cornejo et al., 2014).

With these advancements of seeds technology, combining with the general improvements in farming management and climate changes, it is hardly convincing that the conditional yield densities of major crops are stay invariable, even after some technical adjustment. The changes in the higher moments of yield DGPs through time draws attention since it may have significant impacts on the results of related research topics such as crop insurance, climate-yield relationship and so on. It is not surprising as technological advances shift probability mass in a variety of ways, not simply uniformly upwards, thereby suggesting that not all historical yield data should be used in empirical analyses. For example, with respect to crop insurance applications, it is hard to imagine that losses from the 1950s and 1960s can inform losses today to the anywhere near the same extent as losses in the 2000s.

The purpose of this manuscript is to consider the question of how much historical yield data should be used in empirical analyses. The answer is obviously dependent on the empirical application, crop-region combination, econometric methodology, and chosen loss function. Nonetheless, we attempt to tackle this question in three ways using county-level yield data for corn, soybean, and winter wheat. First we use distributional tests to assess if and when the adjusted yield data may result from different DGPs. Second, we consider the application to crop insurance by using an out-of-sample rating game -- commonly employed in the literature -- to compare rates from the full versus restricted data sets. Third, we estimate flexible time-varying DGPs and then simulate to quantify the additional error when the identically distribution assumption is imposed.

A secondary purpose of this manuscript is to investigate methodologies that can re-incorporate any discarded historical data while: (i) explicitly acknowledging the unknown DGPs are different across time periods; and (ii) not requiring knowledge about the extent or form of those differences. Recall, T (the length of the data) is already quite small of which some may be discarded. Therefore, we investigate methodologies that may re-incorporate this data in alternative ways such that efficiency is increased relative to discarding. We consider three methods: incorporating the discarded data to reduce bias; incorporating the discarded data to reduce variance smoothing; and incorporating the discarded data to reduce variance via Bayesian model averaging. Our results suggest gains in efficiency can be realized by using these methodologies. While our results are most applicable to the crop insurance literature, we certainly feel they suggest proceeding with caution when using historical yield data in other applications as well.

## Literature Review

In the agricultural economics literature, county crop yield data from USDA-NASS database have been used to evaluate alternative rating methodologies for crop insurance, to model the climate-yield relationship, to forecast the effects of a changing climate on yields and land-use, and, to assess productivity efficiencies. In the past twenty years in excess of 100 *American Journal of Agricultural Economics* (*AJAE*) articles make reference to the NASS data set and at least thirty fundamentally use it in their empirical analyses. For most county-crop combinations, this data set is available dating back to 1955 and thus provides a relatively rich set of data to do empirical analyses and test hypotheses. Given the lack of alternatives, we expect this data will continue to be extensively used. In most applications, the data generating process of yields is assumed to change over time only by its location (mean) and scale (variance).

With respect to rating methodologies for crop insurance, the NASS data have been used by Goodwin and Ker (1998); Ker and Goodwin (2000); Ramirez, Misra, and Field (2003); Ker and Coble (2003); Norwood, Roberts, and Lusk (2004); Harri et al. (2011); Woodard and Sherrick (2011); Claassen and Just (2011); Koundouri and Kourogenis (2011); Annan et al. (2013); Tolhurst and Ker (2015); Goodwin and Hungerford (2015); Ker, Tolhurst, and Liu (2016); Yvette Zhang (2017). In these articles, the general treatment of NASS yield data is first to detrend the data then adjust for heteroskedasticity if necessary. After the adjustment, the data are assumed to be identically distributed. In trend estimation, various trend functions are utilized, such as linear (Woodard and Sherrick, 2011), ARIMA (Goodwin and Ker, 1998), polynomial (Ramirez, Misra, and Field, 2003), one-knot or two-knot spline (Ker and Coble, 2003; Harri et al., 2011), noparametric local regression (Claassen and Just, 2011; Goodwin and Hungerford, 2015). The adjustment measures of the second moment (heteroskedasticity adjustment) generally assume constant variance or variance being proportional to the time. Some assume homoskedasticity (Woodard and Sherrick, 2011) and some ignore the issue (Koundouri and Kourogenis, 2011; Goodwin and Hungerford, 2015). Except for Tolhurst and Ker (2015); Ker, Tolhurst, and Liu (2016), few researchers take into account the changing variations on the higher moments.

With respect to other issues of crop insurance, the NASS data has been used by Goodwin, Vandeveer, and Deal (2004); Deng, Barnett, and Vedenov (2007); Woodard et al. (2012); Woodard and Verteramo-Chiu (2017); Claassen, Langpap, and Wu (2017). In these articles, yield data are used as dependent variable or independent variable in regression models, which generally contain time trend. Although in the regression setting, the main interest lies on the conditional mean, therefore heteroskedasticity may not strongly influence the estimation results, but it may affect the standard error in the estimation and has impact on some simulation results such as generating pseudo yield from the assumed distribution.

Moreover, NASS yield data set was applied to modeling the climate-yield relationship Ortiz-Bobea and Just (2013); Roberts, Schlenker, and Eyer (2013); Miao, Khanna, and Huang (2016); Cooper, Nam Tran, and Wallander (2017), agricultural policy analysis Goodwin and Mishra (2006); Goodwin (2009), land using Wu et al. (2004); Claassen, Hellerstein, and Kim (2013), supply function analysis Hendricks, Smith, and Sumner (2014) and commodity pricing Cooper (2010).

The majority of these research follow the common practice with detrending and adjusting heteroskedasticity if necessary. Others embed time trend in regression setting to deal with the changing first moment. Few researchers considered the yields variation of higher moment more than the first two. However, Tack, Harri, and Coble (2012), Tolhurst and Ker (2015), and Ker, Tolhurst, and Liu (2016) found higher moments in yield distributions are also changing through time. This is not surprising as technological advances shift probability mass in a variety of ways, not simply uniformly upwards, thereby suggesting that not all historical yield data should be used in empirical analyses. For example, with respect to crop insurance applications, it is hard to imagine that losses from the 1950s and 1960s can inform losses today to the anywhere near the same extent as losses in the 2000s. The objective of this manuscript is to provide some guidance as to how far back one should use the historical data in empirical analyses. We also consider some of the methodologies that borrow information from like (not identical) data generating processes (DGP) to increase estimation efficiency. These have been used in the crop insurance literature to borrow information across space from like DGP and we adapt to include them as borrowing information across time.

## Tests of Distributional Changes

Before assuming the yield data are independently and identically distributed (i.i.d.) draws from the underlying distribution, common practice in the literature is to adjust the yield data to account for the technological advances and other variations over time. The adjustment approach is first to estimate a trend function, linear or non-linear, which explains the variations of the first moment of yield distribution over time. After the trend estimation, the detrended data are adjusted for prevailing heteroskedasticity, which account for the variations of the second moment of yield distribution over time. After these two steps, detrended and heteroskedasticity adjusted data are employed as i.i.d.s from the constant distribution without further inquiry. Question arises that if the two-step approach is fully account for the complete impacts of technological advances on different moments of yield distribution. After all, it is hard to believe the intricate technological changes over the past half century only shift the first two moments of the yield DGP. A solution to clarify the above concern is to test if there are still structural or distributional changes remaining in the data after the adjustment of first two moments. However, these structural or distributional tests are generally of low power for testing changes in higher moments. That is, if there are still structural or distributional changes in the higher moments, these tests may not be able to reject the false null correctly, committing type II errors. One may argue that these changes in the higher moments have trivial impact on the research result. However, our study shows the opposite. Especially, for research related to crop insurance, the main interest lies on the lower tail of yield distribution. To demonstrate these higher moment (more than the first two) changes have strong impact on the premium rate of crop insurance, we design a simulation study to show that small changes in the higher moments could have significant impacts on the premium rate while keeping undetected by the conventional structural and distributional tests. The details of this simulation study are shown in the Appendix.

The crop yield data we use in this manuscript are chosen for corn, soybean and winter wheat. Specifically, we use county level yield data of Illinois and Iowa for corn and soybean; Kansas, Oklahoma for winter wheat. In total, we have data from 171, 181 and 71 counties for corn, soybean and winter wheat respectively. These county level yield data are chosen on the basis that they are from the major crop-producing states of each individual crop in U.S.. All historical yield data, with length of 61 years from 1955 to 2015, are obtained from NASS website.

Before testing the distributional changes in historical yield data over time, the yield data were first detrended with four different trend functions including linear regression (L2), median regression (L1), nonparametric local smoothing and RMA methodology [2]. Although there are other detrending methodologies used in the literature, these four methods cover the majority of usages in the literature. After detrending, the yields are adjusted for heteroskedasticity following Harri et al. (2011), which is common used in the literature.

We choose the Kolmogorov-Smirnov (KS) test [3] for the structural changes in the detrended and heteroskedasticity adjusted yield by splitting the sample into two segments with different length of historical time. The KS test is a nonparametric test of the equality of continuous, one-dimensional probability distributions. Its statistic quantifies a distance between the empirical distribution functions of two samples. The test statistic for two-sample test is

[2]The RMA detrending methodology is mainly one or two-knot linear spline based on AIC or BIC.

[3]The Chow test, Bai-Perron breaking points test, etc are not appropriate for what we are testing, because the distributional assumptions of these tests are Normal and don't hold for our study.

$$D_{n,m} = \sup_x \mid F_{1,n}(x) - F_{2,m}(x) \mid, \tag{1}$$

where $F_{1,n}$ and $F_{2,m}$ are the empirical distribution functions of the first and the second sample respectively, and sup is the supremum function. The null is rejected at level $\alpha$ when

$$D_{n,m} > c(\alpha)\sqrt{\frac{n+m}{nm}}, \tag{2}$$

where $c(\alpha)$ is calculated from the Kolmogorov distribution.

The test results of split adjusted residuals from different detrending methodologies are showing in Table 1. Although the KS test is fairly low power (under-reject) in samples with size in our test, it still shows notable structural changes in our tests with multiple splits. The results in Table 1 are rather consistent across different detrending methodologies. For corn, there are sizable rejections around 25 to 40 years splits through all four detrending methods. With RMA detrending method, corn shows the most structural changes with 81 of all 171 counties; soybean shows changes with 51 of all 181 counties and winter wheat with 24 of all 71 counties, indicating non-negligible structural changes remaining in the adjusted historical crop yields.

To investigate how these embedded structural changes would impact the premium rate, and further affect the viability of area-type crop insurance program, we conduct an out-of-sample rating game, where the two players using different length of historical yield data, to examine the effect. The out-of-sample rating game we conducted was first proposed by Ker and McGowan (2000) and used by Ker and Coble (2003), Racine and Ker (2006), Harri et al. (2011), Annan et al. (2013), Tolhurst and Ker (2015) and Ker, Tolhurst, and Liu (2016).

The game is inspired by the design of U.S. crop insurance program that the RMA sets the premium rate for each insurance contract, while the contracts are delivered by the private insurance companies to the crop producers. In this design, RMA sets the premium rates and the private companies sell the policies, conduct claims, and asymmetrically share in the underwriting gains and losses with RMA. Standard Reinsurance Agreement (SRA) dictates how underwriting gains and losses are shared between funds and lays out the reimbursement, and private insurance companies must sell all available policies in a state. In the rating game, the companies could make their own decisions to keep the contract or cede it to the government. An underwriting gain could be achieved when the indemnities are less than the premiums paid by the producers. On the other hand, a loss could occur when the indemnities are greater than the paid premiums. An interesting feature of this design is that if the private companies have a different rating methodology, therefore resulting different premium rates from the government's, the companies may achieve more gains and suffer less losses by the first-move advantage of keeping or ceding the contracts. For instance, for a specific private company, if its own pre-calculated premium rate is higher than the government's, it implies that its anticipated future indemnity is higher its income (premium set by the government and paid by the producer). The reasonable choice is to cede the contract to the government to avoid the anticipated

Table 1. Rejections of KS Test Results for The Adjusted Residuals

| | *Corn* | | *Soybean* | | *Winter Wheat* | |
|---|---|---|---|---|---|---|
| | Illinois | Iowa | Illinois | Iowa | Kansas | Oklahoma |
| $n$ County | 76 | 95 | 85 | 96 | 50 | 21 |
| *RMA Detrending* | | | | | | |
| 10 vs 50 | 2 | 6 | 0 | 6 | 5 | 9 |
| 15 vs 45 | 2 | 3 | 2 | 9 | 4 | 2 |
| 20 vs 40 | 4 | 10 | 5 | 5 | 2 | 1 |
| 25 vs 35 | 6 | 10 | 2 | 0 | 2 | 6 |
| 30 vs 30 | 11 | 9 | 7 | 0 | 1 | 7 |
| 35 vs 25 | 11 | 9 | 8 | 3 | 0 | 5 |
| 40 vs 20 | 10 | 7 | 2 | 3 | 3 | 3 |
| 45 vs 15 | 13 | 7 | 0 | 4 | 7 | 7 |
| 50 vs 10 | 11 | 31 | 6 | 11 | 4 | 2 |
| Total Rejections | 29 | 52 | 24 | 27 | 14 | 10 |
| *L2 Detrending* | | | | | | |
| 10 vs 50 | 2 | 5 | 0 | 6 | 4 | 11 |
| 15 vs 45 | 1 | 5 | 1 | 9 | 3 | 1 |
| 20 vs 40 | 3 | 11 | 4 | 5 | 0 | 0 |
| 25 vs 35 | 5 | 10 | 2 | 0 | 1 | 1 |
| 30 vs 30 | 11 | 14 | 7 | 0 | 0 | 0 |
| 35 vs 25 | 12 | 14 | 9 | 3 | 0 | 1 |
| 40 vs 20 | 10 | 13 | 3 | 3 | 4 | 2 |
| 45 vs 15 | 11 | 11 | 1 | 4 | 16 | 1 |
| 50 vs 10 | 8 | 21 | 7 | 12 | 11 | 2 |
| Total Rejections | 28 | 53 | 23 | 27 | 20 | 11 |
| *L1 Detrending* | | | | | | |
| 10 vs 50 | 3 | 6 | 0 | 7 | 6 | 12 |
| 15 vs 45 | 3 | 5 | 2 | 7 | 7 | 2 |
| 20 vs 40 | 10 | 10 | 5 | 1 | 1 | 0 |
| 25 vs 35 | 7 | 8 | 2 | 1 | 0 | 1 |
| 30 vs 30 | 15 | 11 | 2 | 1 | 0 | 0 |
| 35 vs 25 | 10 | 11 | 3 | 1 | 2 | 0 |
| 40 vs 20 | 7 | 6 | 1 | 1 | 6 | 1 |
| 45 vs 15 | 4 | 1 | 4 | 1 | 14 | 1 |
| 50 vs 10 | 2 | 12 | 4 | 10 | 10 | 1 |
| Total Rejections | 27 | 35 | 13 | 21 | 22 | 12 |
| *Nonparametric Detrending* | | | | | | |
| 10 vs 50 | 1 | 3 | 0 | 3 | 1 | 2 |
| 15 vs 45 | 1 | 2 | 0 | 5 | 0 | 0 |
| 20 vs 40 | 4 | 10 | 3 | 4 | 0 | 0 |
| 25 vs 35 | 3 | 7 | 2 | 0 | 0 | 0 |
| 30 vs 30 | 5 | 3 | 2 | 0 | 0 | 0 |
| 35 vs 25 | 1 | 3 | 2 | 2 | 0 | 0 |
| 40 vs 20 | 1 | 2 | 0 | 1 | 0 | 0 |
| 45 vs 15 | 7 | 2 | 0 | 2 | 5 | 0 |
| 50 vs 10 | 6 | 24 | 4 | 8 | 1 | 0 |
| Total Rejections | 20 | 43 | 13 | 20 | 6 | 2 |

loss. If its own rate is less than the governments, keeping the contract would induce more income since its anticipated future indemnity is less than the income.

The rating game starts from using all or partial of the yield data from 1955-1995 to estimate the premium rate of 1996. We repeated this process to estimate the premium rates of 1997, ... , 2015 within 20-year game length by using the all or partial of the yield data from 1955-1996, ... , 1955-2014. The decision rule for using all or partial of the yield data is based on the KS test results. The historical yield data are first detrended with RMA methodology then adjusted for the heteroskedasticity following Harri et al. (2011). For detrended and adjusted yield data using in each year of the game, we first run the KS test following the above mentioned split method and identify the time-point where the rejection happens. If there is no rejection, the current contract would be ignored. If there are multiple rejections, we choose the time-point where the most recent one occurs. For all the contracts with identified time-point of structural changes, the RMA calculates the premium rates using all the yield starts from 1955, while the private company decides the rates using the data starts from the identified time-point. For example, if the KS test identifies that the most recent structural change happens at 1981, the private uses yield data from 1981-1995 to calculate the premium rate of 1996 while the RMA using all the data from 1955-1995. For both players, the 90% premium rate are calculated by estimate the yield density non-parametrically, where the smoothing parameter $h$ is chosen by likelihood cross-validation (LCV).

Table 2. Out-Of-Sample Rating Game, Cut-off from KS Rejection

| Crop-State | Retained by Private (%) | Loss Ratio Government | Loss Ratio Private | $p$-value 1 | $p$-value 2 |
|---|---|---|---|---|---|
| *Corn* | | | | | |
| Illinois (202) | 39.1 | 1.972 | 0.812 | 0.0040 | 0.0059 |
| Iowa (390) | 67.9 | 0.825 | 0.368 | 0.0032 | 0.0059 |
| *Soybean* | | | | | |
| Illinois (89) | 41.6 | 2.805 | 1.090 | 0.0258 | 0.0577 |
| Iowa (106) | 49.1 | 1.119 | 0.452 | 0.0530 | 0.0013 |
| *Winter Wheat* | | | | | |
| Kansas (112) | 53.6 | 1.266 | 0.708 | 0.0774 | 0.0577 |
| Oklahoma (105) | 12.4 | 2.083 | 1.518 | 0.1552 | 0.2517 |

The game results in Table 2 show that the private company who is using partial of the yield data have lower loss ratios in all six crop-state combinations. In Table 2, the total number of contracts with KS rejections is shown in the parentheses after the state name. $p$-value 1 is derived from randomization test. In the test, the same number of contracts as those retained by the private company under its decision rule are randomly selected and the corresponding loss ratio is calculated. This procedure is repeated for 5000 times and the loss ratio under the decision rule are compared with those 5000 loss ratios to derive the $p$-value 1 under the null that the decision rule is equivalent to random selection.

Table 3. Out-Of-Sample Rating Game, Fixed Cut-off vs Full Sample

| Crop-State | Retained by Private (%) | Loss Ratio Government | Loss Ratio Private | $p$-value 1 | $p$-value 2 |
|---|---|---|---|---|---|
| 20-year Cut-off | | | | | |
| *Corn* | | | | | |
| Illinois (76 × 20) | 51.4 | 0.906 | 0.825 | 0.2878 | 0.0207 |
| Iowa (95 × 20) | 49.5 | 0.551 | 0.399 | 0.0286 | 0.2517 |
| *Soybean* | | | | | |
| Illinois (85 × 20) | 48.2 | 1.885 | 0.835 | 0.0000 | 0.0207 |
| Iowa (96 × 20) | 33.3 | 1.078 | 0.458 | 0.0000 | 0.1316 |
| *Winter Wheat* | | | | | |
| Kansas (50 × 20) | 19.8 | 1.364 | 1.134 | 0.1176 | 0.7483 |
| Oklahoma (21 × 20) | 17.9 | 1.811 | 1.468 | 0.1230 | 0.1316 |
| 25-year Cut-off | | | | | |
| *Corn* | | | | | |
| Illinois (76 × 20) | 44.1 | 1.281 | 0.488 | 0.0000 | 0.0002 |
| Iowa (95 × 20) | 39.8 | 0.542 | 0.370 | 0.0114 | 0.7483 |
| *Soybean* | | | | | |
| Illinois (85 × 20) | 35.4 | 1.793 | 0.682 | 0.0000 | 0.0207 |
| Iowa (96 × 20) | 30.4 | 0.947 | 0.606 | 0.0018 | 0.2517 |
| *Winter Wheat* | | | | | |
| Kansas (50 × 20) | 25.1 | 1.447 | 1.009 | 0.0074 | 0.2517 |
| Oklahoma (21 × 20) | 32.1 | 1.954 | 1.343 | 0.0072 | 0.2517 |
| 30-year Cut-off | | | | | |
| *Corn* | | | | | |
| Illinois (76 × 20) | 41.8 | 1.214 | 0.518 | 0.0000 | 0.0000 |
| Iowa (95 × 20) | 44.6 | 0.568 | 0.364 | 0.0042 | 0.7483 |
| *Soybean* | | | | | |
| Illinois (85 × 20) | 28.2 | 1.720 | 0.584 | 0.0000 | 0.0002 |
| Iowa (96 × 20) | 35.3 | 1.021 | 0.516 | 0.0000 | 0.1316 |
| *Winter Wheat* | | | | | |
| Kansas (50 × 20) | 37.9 | 1.596 | 0.973 | 0.0002 | 0.5881 |
| Oklahoma (21 × 20) | 49.0 | 2.003 | 1.501 | 0.0200 | 0.5881 |
| 35-year Cut-off | | | | | |
| *Corn* | | | | | |
| Illinois (76 × 20) | 49.3 | 1.208 | 0.591 | 0.0000 | 0.0013 |
| Iowa (95 × 20) | 54.1 | 0.798 | 0.257 | 0.0000 | 0.1316 |
| *Soybean* | | | | | |
| Illinois (85 × 20) | 27.9 | 1.580 | 0.820 | 0.0000 | 0.0002 |
| Iowa (96 × 20) | 46.6 | 0.946 | 0.691 | 0.0078 | 0.2517 |
| *Winter Wheat* | | | | | |
| Kansas (50 × 20) | 44.7 | 1.502 | 1.139 | 0.0220 | 0.4119 |
| Oklahoma (21 × 20) | 59.5 | 1.884 | 1.656 | 0.1826 | 0.5881 |

As mentioned above, the private insurance company could have a first-move advantage that it chooses to retain or cede the contract first. To cancel out this advantage and present the real efficacy of its decision rule, we derive $p$-value 2 from an efficacy test developed by Ker, Tolhurst, and Liu (2016), in which the null is that both methodologies from the RMA and the private insurance company are equally efficient.

Aside from examining the contracts with KS rejections, we also examine all the contracts available by running the games assuming the private insurance company uses only recent yield data with fixed cut-off versus the RMA uses the full set of yield data. We choose fixed time cut-offs as recent 20, 25, 30 and 35 years. The results are shown in the Table 3.

In Table 3, the values of $p$-value 1 suggest that economically and statistically significant monies can be made restricting the use of historical data for 5 out of 6 state-crop combinations. The values of $p$-value 2 suggest that restricting the data leads to more accurate rates for 5 out of 6 state-crop combinations with the exception of Oklahoma winter wheat. In Table 3, all values of $p$-value 1 from 25 and 30 years cutoffs are statistically significant. Overall, results are fairly consistent between 20, 25, 30, and 35 years cut-offs for $p$-value 1, less so for $p$-value 2.

Although there is no clear sign to show how long the recent yield data should be used from these games, above results show evidences that using less and more recent yield data could greatly improve the efficiency of rating crop insurance contract. The results are consistent with our conjecture that there are significant structural changes in the historical crop yield caused by the technological advances in seeds and farming management.

## Estimating and Simulating DGPs of Yields Overtime

Since common structural and distributional tests do not perform well on detecting the possible underlying structural changes in higher moments of yield density, we propose a different approach to examine the impacts of these changes by estimating the yield distribution with a flexible and accountable modeling method as a start. We propose a Normal mixture estimator with time-varying parameters to model the county level crop yield, which has the following distributional form

$$y_t \sim \sum_{k=1}^{n} \lambda_{tk} N(\alpha_k + \beta_k t,\ \gamma_k + \delta_k t), \tag{3}$$

where $y_t$ is the yield of year $t$, $n$ is the number of Normal components, $\lambda_{tk}$ is weight parameter of Normal component $k$ of year $t$, satisfying $\sum_{k=1}^{n} \lambda_{tk} = 1$; $\alpha_k + \beta_k t$ is time-varying mean parameter and $\gamma_k + \delta_k t$ being time-varying variance parameter for each Normal component of year $t$. Although the choice of the number of components $n$ could be arbitrary, we find that two component ($n = 2$) mixture model is the optimal choice for all crop yields we choose by the measurement of Bayesian Information Criterion (BIC). Throughout this manuscript, we fix our model choice as two components Normal mixture.

Normal mixture model can approximate many distributional structures associated with yield densities, such as asymmetric, skewed, bimodal and long-tailed. In fact, the Normal mixture can be used to approximate any continuous distribution to an arbitrarily small difference (Everitt and Hand, 1981). With extra time-varying settings on the mean and variance parameters, this estimator can model crop yield in a more realistic and reasonable way with the relaxed restrictions on the first two moment parameters over time. Figure 1 shows examples of real and simulated county yields of three crops. Visual inspection shows that the red dots, representing simulated yields, match the real yields in blue pretty well.

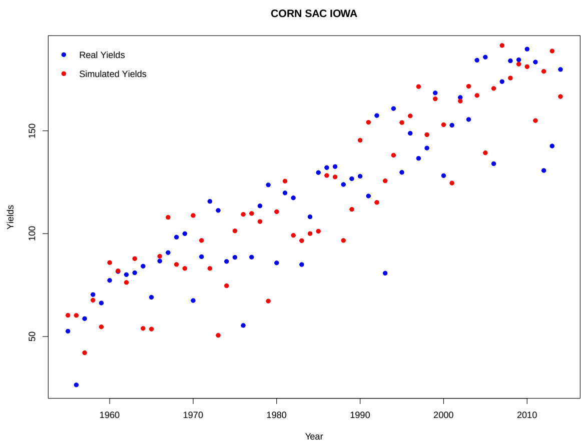
The advantages of employing this time-varying Normal mixture model are in two folds. First, it is flexible on the changes in mean and variance of yield distribution over time, which captures the typical manner of real world crop yields. Simulation results show that it can mimic the real world yield pretty well and capture its characteristics such as increasing mean and variance over time. Second, and more important, using this Normal mixture estimator with time-varying parameters enables us to establish the "true" data generating process (DGP) for each crop-producing county, which allows us to be able to evaluate the performance of different estimators for crop yield densities using different efficiency criteria, and test related hypothesis while avoiding likely type II errors. Admittedly, the legitimacy of this paper is conditional on the assumption that crop yield data are truly generated from above Normal mixture model. Since there is no way to really identify the "true" yield density, our simulation approach provides a practical way to investigate the crop yield DGPs while avoiding the above mentioned type II errors.

Table 4

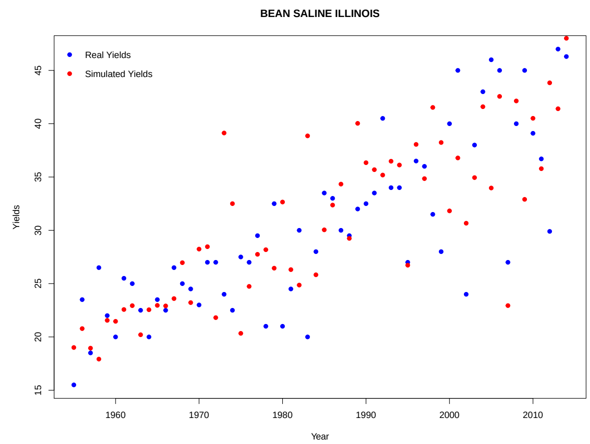
| | Mean Parameters of Mixture Model by Crop | | |
|---|---|---|---|
| | Corn | Soybean | Winter Wheat |
| n Counties | 171 | 181 | 71 |
| $\bar{\lambda}$ | 0.48 | 0.49 | 0.52 |
| $a_1$ | 0.48 | 0.49 | 0.47 |
| $b_1$ | 0.00 | 0.00 | 0.00 |
| $\alpha_1$ | 55.67 | 22.81 | 22.19 |
| $\beta_1$ | 1.68 | 0.40 | 0.22 |
| $\gamma_1$ | 146.51 | 10.62 | 23.48 |
| $\delta_1$ | 11.27 | 0.51 | 1.05 |
| $\alpha_2$ | 64.62 | 24.99 | 26.73 |
| $\beta_2$ | 1.95 | 0.47 | 0.27 |
| $\gamma_2$ | 44.50 | 3.73 | 28.17 |
| $\delta_2$ | 1.74 | 0.10 | 0.39 |

*Note:* Counties with incomplete yield histories are excluded. Unit is bushel per acre.

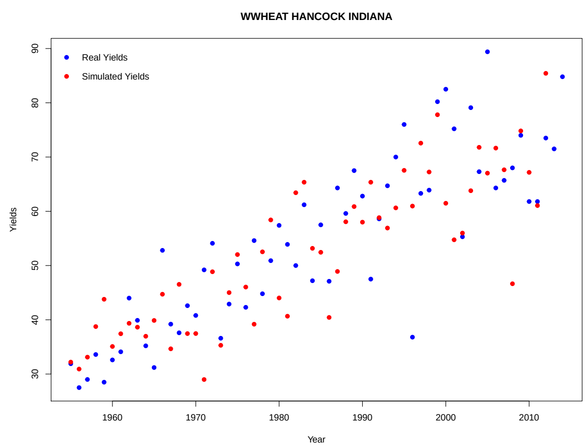
Assuming county level yield data are generated from time-conditional Distribution (3), the parameters of each individual crop producing county yield distribution can be estimated by using EM algorithm. The details of this EM algorithm are referred to Tolhurst and Ker (2015). Table 4 presents the mean values




(A) Corn




(B) Soybean



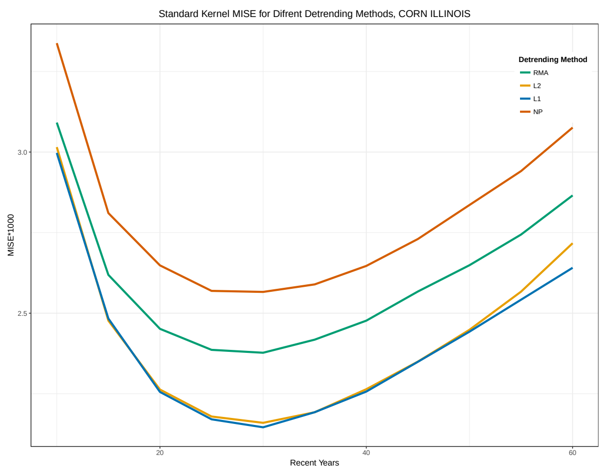

(C) Winter Wheat

FIGURE 1. Real & Simulated Crop Yields

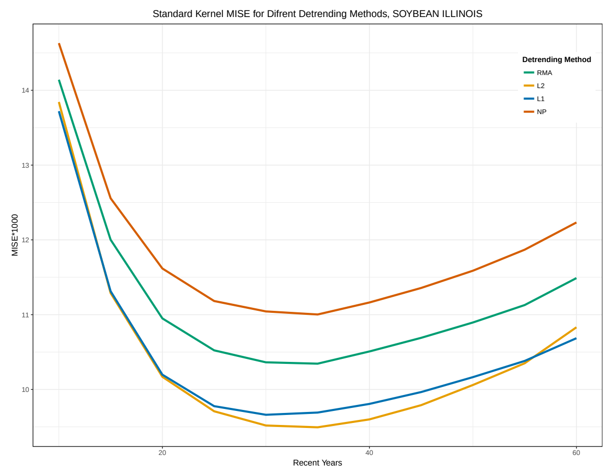
of parameters from proposed two-component Normal mixture model for three crops. With these time-conditional parameters produced by EM algorithm, the simulated yield samples can be drawn from the estimated distribution with a "true" yield DGP in hand. With enough simulated yield data, we can examine and compare the performances of chosen estimators by using simulated yields with different time length, which equivalently answers our research questions that how long we should use the historical data optimally.

To examine the performance of the estimator we proposed under the proposed mixture model, we draw 500 samples with size of 61 (representing historical yield of 61 years) from each crop-county combination. For each sample, we detrend the sample data with former mentioned four different trend functions, then adjust the detrended data for the heteroskedasticity. The estimation starts with utilizing only the recent 10 years data, then incrementally expanding the data utilized with 5 more years sample dating back in each step, until using the full data set. This strategy is applied to all estimators using in this manuscript. As mentioned previously, our argument is that technological advances may have changed the higher moments of yield distribution, therefore the i.i.d. assumption may not hold even if we have adjusted the first two moments of the distribution. The implication is that the performance of estimator may not necessarily achieve the best efficiency by using the full dataset, since the adjusted data from early stage may have different higher moments from the recent ones. Intuitively, the changes in the higher moments would have negative impact on the efficiency of estimator when using such "contaminated" dataset.

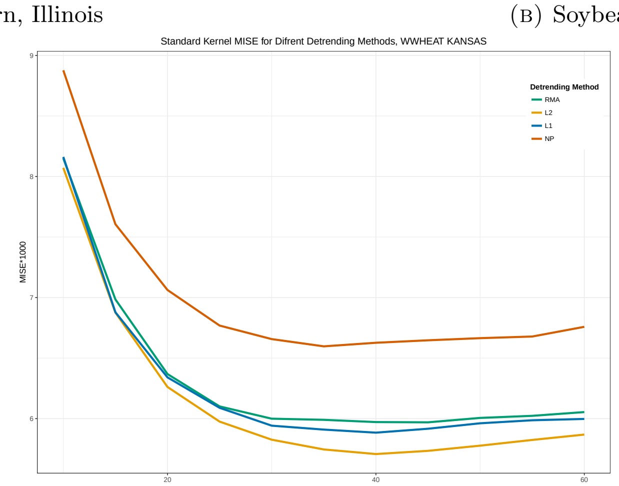
For each estimate, we calculate the mean integrated squared error (MISE) by comparing the estimated density with the "true" density, through which we can observe the overall performance of the estimator used. To specifically examine the lower tail performance, which is crucial for crop insurance, we also recover the premium rate of 90% coverage level for the estimator. These premium rates are compared with the "true" premium rates with root-mean-square error (RMSE) criterion.




(A) Corn, Illinois




(B) Soybean, Illinois




(C) Winter Wheat, Kansas

FIGURE 2. MISE, $h$ by LCV

We first examine the performance of standard kernel estimator under different detrending methods, where the smoothing parameter $h$ of is chosen by the criterion of maximizing the likelihood of cross-validation (LCV) [4]. Figure 2 presents the average MISE across state-wide for three crop-state combinations. As the sample size

[4]The results of using $h$ chosen by the criterion of minimizing the integrated squared error (ISE) are provided in the Appendix.

increasing, the MISE decreases drastically, which is as expected that more sample implies more information, therefore reduces the estimation error. However, what is interesting that draws our attention is that the MISE starts to increase or flatten out as the sample size increases. In Figure 3, the average RMSE of 90% premium rate across state for three crop-state combination also show the similar first decreasing them increasing or flattening out situations. These $U$ shaped or alike graphs evidently show the implication that more data does not necessarily lead to more efficiency for yield density estimation in each crop-state combination. On the contrary, the more "outdated" data used, the less or no efficiency gains in the estimation after a certain historical time point. At this moment, for the application purpose, it seems the reasonably choice is to discard the data older than the year that the minimum MISE hits and using the left "recent" data for the estimation. Nevertheless, discarding data is not a pleasant choice for practitioner since there is still more or less useful information within the outdated data. To exploit this usefulness requires treatments that are more sophisticated. There is a balance need to strike between using all but possibly contaminated data and less but more likely following i.i.d. assumption data. To solve this dilemma, we propose an approach that goes beyond simply discarding data by using estimators that can borrow information from like (but not the same) DGPs. The borrowing-information estimators are introduced in the following section.

## Incorporating Data from Like DGPs

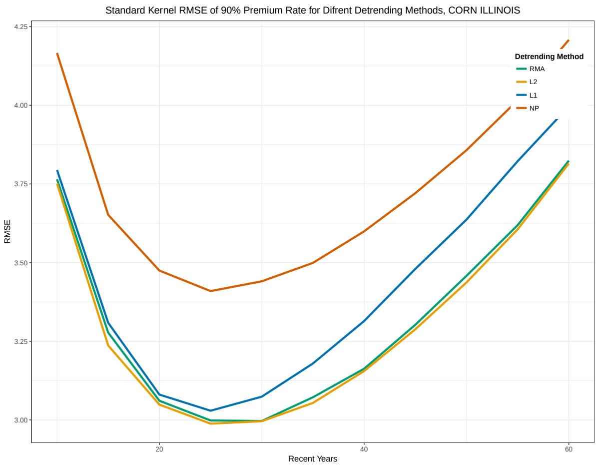
The purpose of proposed methodologies is that we can use data from possibly like DGPs to improve estimation efficiency so we do not necessarily discard historical yield data. We consider three borrowing-information methods: Possibly Similar method, Bayesian Model Averaging (BMA) method, Li and Racine method. These borrowing information estimators can make use of the extraneous data if they are sufficiently like, while not significantly compromised if they are not sufficiently like. Here, the definition of likeness is purposely vague, that is, we make no assumptions as to the degree and form of likeness. These estimators are only trivially compromised by correlated data with an unknown correlation structure.

Possibly Similar method is a non-parametric estimator that uses extraneous data to reduce estimation bias. When estimating a set of densities thought to be similar it may be more efficient to pool the data and estimate a single estimate [5] and then non-parametrically correct it for each individual density estimate. The Possibly Similar estimator has the form

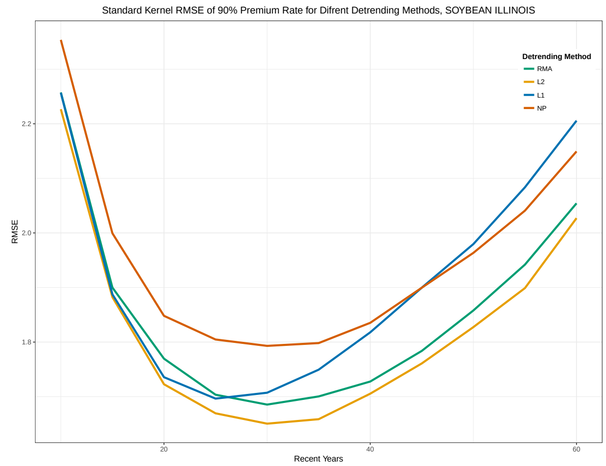
$$\tilde{f}_i(x) = \hat{g}(x)\hat{r}_i(x) = \sum_{j=1}^{n}(1/n)K_h(x - X_{ij})\frac{\hat{g}(x)}{\hat{g}(X_{ij})}, \tag{4}$$

where $\hat{g}$ is the start estimate, which in our study is the kernel estimate based on the full historical data and requires smoothing parameter $h_p$; $h$ is the smoothing parameter for the individual correction function, which corresponds to the partial data of recent; $X_{ij}$ is the sample from density $i$. The intuition is to reduce the global curvature of the underlying function being estimated thereby reducing bias. The correction factor

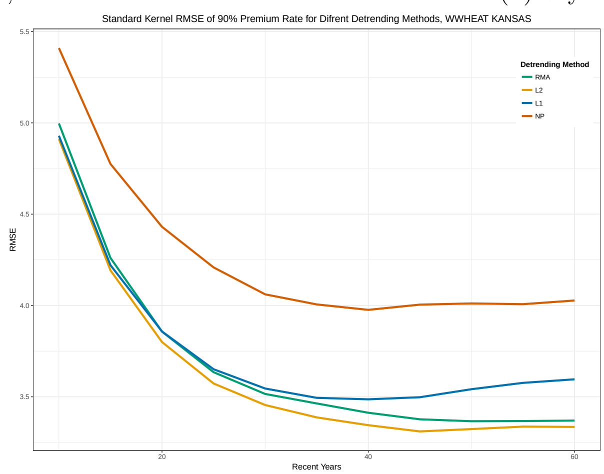
[5] The efficient estimator if the densities considered were identical would be the one to pool the data.




(A) Corn, Illinois




(B) Soybean, Illinois




(C) Winter Wheat, Kansas

Figure 3. RMSE of 90% Premium Rate , $h$ by LCV

will have less global curvature if the start estimate is sufficiently close to the unknown density of interest. The estimator is consistent and does not require assumptions about the form or extent of likeness between DGPs. It performs well in small samples for like densities and no worse for unlike densities (e.g., test densities from Marron and Wand (1992)). It's weakness is that it assumes independence, therefore correlation effects the pooled estimate through $h_p$. Also, it may cause over or under smoothing of the start estimate. A possible solution is to correct the start by block cross-validation. Overall, Possibly Similar estimator is naive (ignorant) in that it treats all other extraneous data equally. Since bias is a function of the curvature, we can often reduce curvature using a better start than a random pick. In this paper, we use the historical data as our start for the current period. We refer to Hjort and Glad (1995) and Ker (2016) for more details.

Bayesian Model Averaging has been widely used in the literature to deal with model uncertainty. For a comprehensive introduction, we refer to Hoeting et al. (1999). However, we use BMA in a novel and different way in this manuscript. Rather than assuming the same data set are generated from different candidate models, we construct the set of candidate models from the set of densities estimates based on extraneous data. The BMA estimator has the form

$$\tilde{f}_i = \sum_{j=1}^{Q} \omega_j^i \hat{f}_j, \tag{5}$$

where

$$\omega_j^i = \frac{\exp\Big\{-\frac{1}{2}\mathrm{BIC}_j^i\Big\}}{\sum_{q=1}^{Q}\exp\Big\{-\frac{1}{2}\mathrm{BIC}_q^i\Big\}}, \tag{6}$$

$\mathrm{BIC}_j^i$ is the Bayesian information criterion (BIC) of model $\hat{f}_j$ with data from experimental unit $i$, $\hat{f}_j$ are the set of density estimates across experimental units. The BMA estimator assumes uniform prior across models and a mixture of Normals for $f_j$ for above form. The weight $\omega_j^i$ assigned to extraneous data (i.e. estimated models) is determined by likelihood the data of interest comes from that estimated model, where the likelihood is represented by the $\mathrm{BIC}_j^i$. In this manuscript, we estimate $\hat{f}_j$ using standard kernel method, and obtain BIC value by interpreting the standard kernel estimator as mixture of Normals. Details are referred to Ker and Liu (2016). Naturally, own unit receives maximum weight in parametric framework, but may not necessarily be true in nonparametric case. For a parametric framework of this BMA estimator, we refer to Ker, Tolhurst, and Liu (2016). The advantage of this estimator is that it converges to individual kernel estimate and does not require assumptions about form or extent of similarity between densities. It is not naive in that it does not necessarily treat all other extraneous data equally. It performs well in small samples for like densities and no worse for unlike densities. Nonetheless, its estimates are based on the independent data and then mixing may be impacted by correlation.

Li-Racine method is to smooth across mixed data-types. Its continuous component is the random variable of interest and the discrete component is the different time periods. It has the following form

$$\hat{f} = (nh_C)^{-1}\left[\underbrace{\sum_{l=1}^{n_1}(1-\lambda)K\left(\frac{x-X_l^{C_j}}{h_C}\right)}_{\text{Current data}} + \underbrace{\sum_{l=1}^{n_2}\lambda K\left(\frac{x-X_l^{C_{-j}}}{h_C}\right)}_{\text{Historical data}}\right]. \tag{7}$$

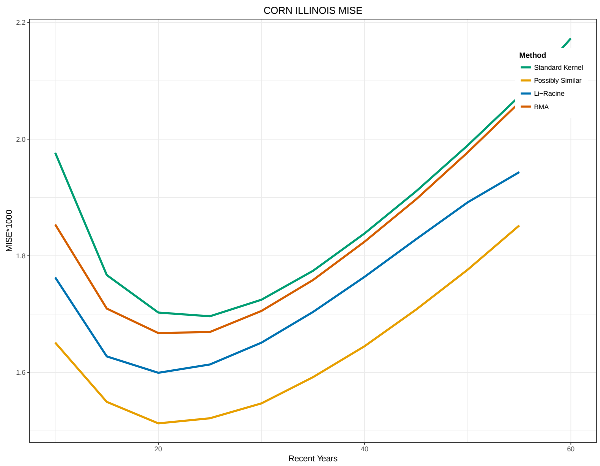
Essentially, Li-Racine estimator smooths across all data but weights the data in the current period higher than that from the historical data. It reduces variance while increases bias for the estimate. The two smoothing parameters, $h_C$ and $\lambda$ in above equation could be chosen by cross-validation. The estimator is consistent and requires no assumptions about form or extent of similarity between DGPs. Correlation may affect smoothing parameter $\lambda$. The estimator is naive in a sense that it treats all other extraneous data equally and could change with additional smoothing parameters. More details about Li-Racine estimator can be obtained from Li and Racine (2003); Racine and Li (2004); Li, Simar, and Zelenyuk (2016).

### Games and Estimations Revisited

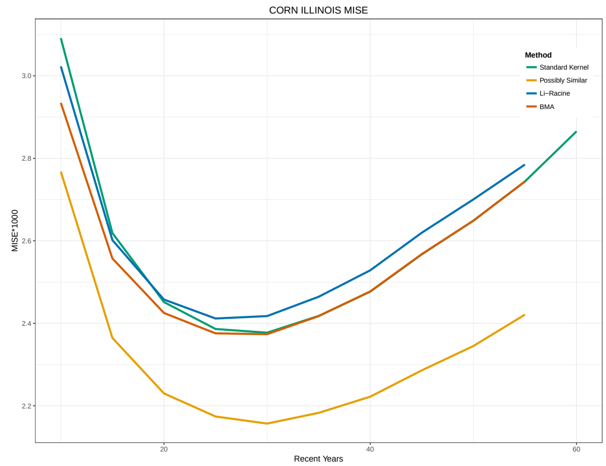
We use these three borrowing-information estimators to re-run the contracting games, assuming the private insurance company employs these estimators for rating the premium rates while the RMA still rates the contracts using the standard kernel estimator with full data set. For the following games, we use the standard RMA detrending methodology because the results are rather consistent through different detrending methodologies. For borrowing estimators, the private insurance company uses the same cut-off time as previously identified for standard kernel estimator , but would treat the adjusted yields earlier than the cut-off time as the reference data to "borrow" information from.

TABLE 5. Results of Out-of-Sample Rating Game, Cut-off from KS Rejection

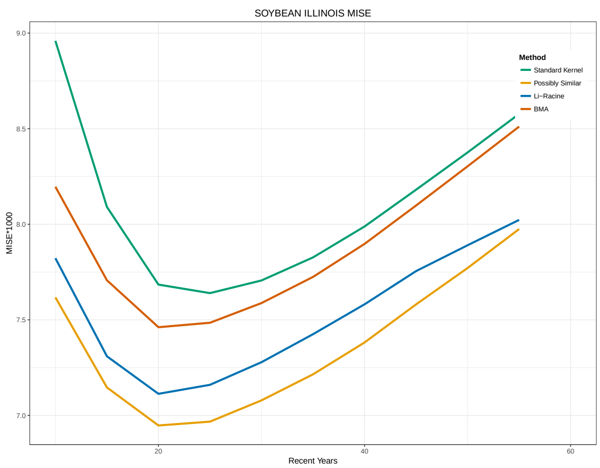
| Crop-State | Retained by Private (%) | Loss Ratio Government | Loss Ratio Private | $p$-value 1 | $p$-value 2 |
|---|---|---|---|---|---|
| *Corn Illinois (202)* | | | | | |
| Standar Kernel | 39.1 | 1.972 | 0.812 | 0.0040 | 0.0059 |
| Possibly Similar | 79.7 | 1.667 | 1.409 | 0.2844 | 0.0207 |
| Li-Racine | 39.1 | 1.953 | 0.829 | 0.0036 | 0.0059 |
| BMA | 39.1 | 1.972 | 0.812 | 0.0022 | 0.0059 |
| *Corn Iowa (390)* | | | | | |
| Standar Kernel | 67.9 | 0.825 | 0.368 | 0.0032 | 0.0059 |
| Possibly Similar | 85.1 | 0.834 | 0.442 | 0.0190 | 0.0002 |
| Li-Racine | 54.6 | 0.633 | 0.416 | 0.0806 | 0.0013 |
| BMA | 66.9 | 0.800 | 0.373 | 0.0024 | 0.0059 |
| *Soybean Illinois (89)* | | | | | |
| Standard Kernel | 41.6 | 2.805 | 1.090 | 0.0258 | 0.0577 |
| Possibly Similar | 73.0 | 3.733 | 1.467 | 0.0246 | 0.0207 |
| Li-Racine | 40.4 | 2.798 | 1.038 | 0.0194 | 0.0577 |
| BMA | 40.4 | 2.685 | 1.146 | 0.0362 | 0.0577 |
| *Soybean Iowa (106)* | | | | | |
| Standard Kernel | 49.1 | 1.119 | 0.452 | 0.0530 | 0.0013 |
| Possibly Similar | 86.8 | 1.737 | 0.601 | 0.0268 | 0.0000 |
| Li-Racine | 35.8 | 0.887 | 0.571 | 0.2322 | 0.0013 |
| BMA | 48.1 | 1.112 | 0.448 | 0.0504 | 0.0013 |
| *Wheat Kansas (112)* | | | | | |
| Standard Kernel | 53.6 | 1.266 | 0.708 | 0.0774 | 0.0577 |
| Possibly Similar | 90.2 | 1.400 | 0.891 | 0.1698 | 0.1316 |
| Li-Racine | 50.0 | 1.247 | 0.689 | 0.0746 | 0.0577 |
| BMA | 53.6 | 1.266 | 0.708 | 0.0736 | 0.0577 |
| *Wheat Oklahoma (105)* | | | | | |
| Standard Kernel | 12.4 | 2.083 | 1.518 | 0.1552 | 0.2517 |
| Possibly Similar | 32.4 | 2.075 | 1.881 | 0.3054 | 0.1316 |
| Li-Racine | 13.3 | 2.088 | 1.519 | 0.1438 | 0.0577 |
| BMA | 12.4 | 2.083 | 1.518 | 0.1598 | 0.2517 |



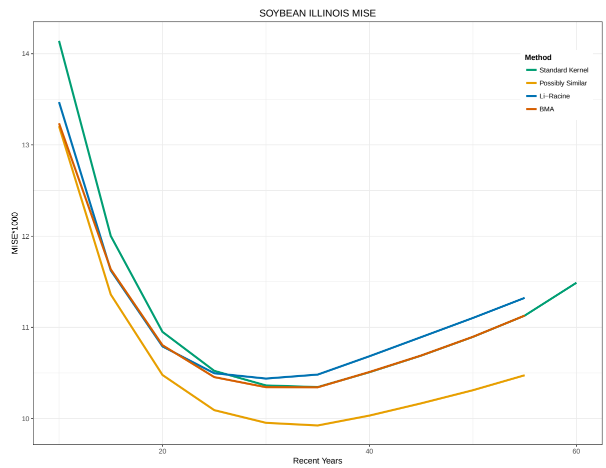

(A) Corn, Optimal $h$



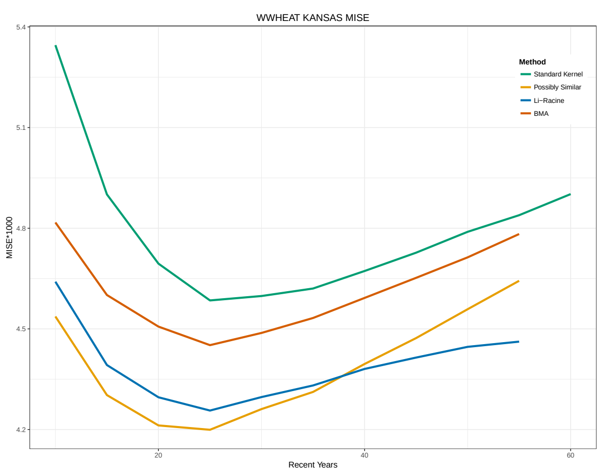

(B) Corn, $h$ by LCV



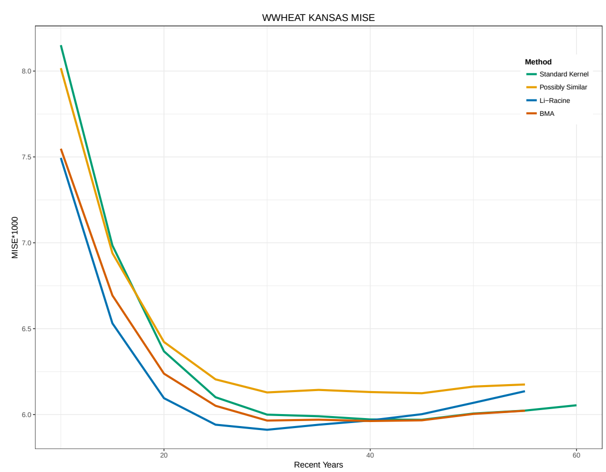

(C) Soybean, Optimal $h$




(D) Soybean, $h$ by LCV

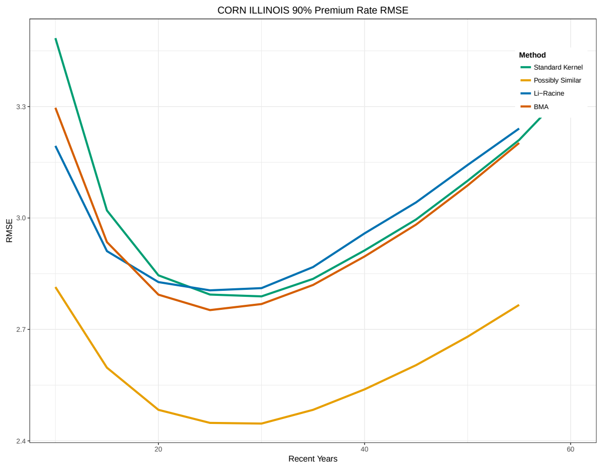



(E) Winter Wheat, Optimal $h$

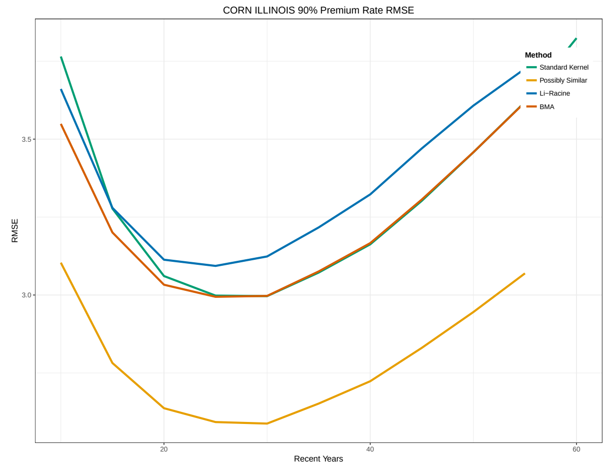



(F) Winter Wheat, $h$ by LCV

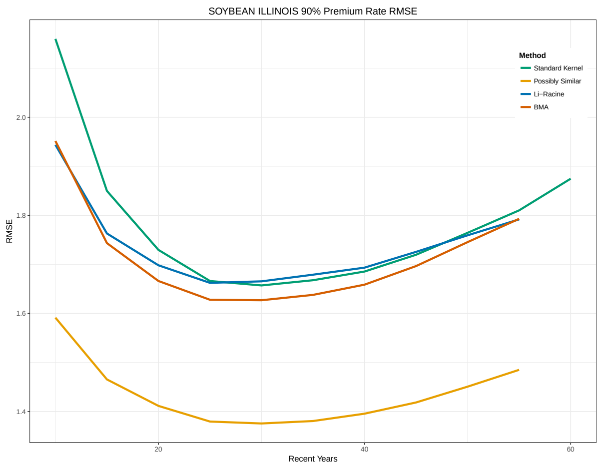
FIGURE 4. MISE, RMA Detrending



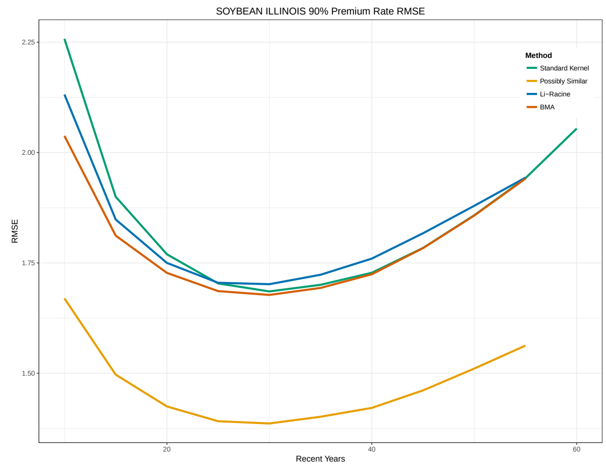

(A) Corn, Optimal $h$



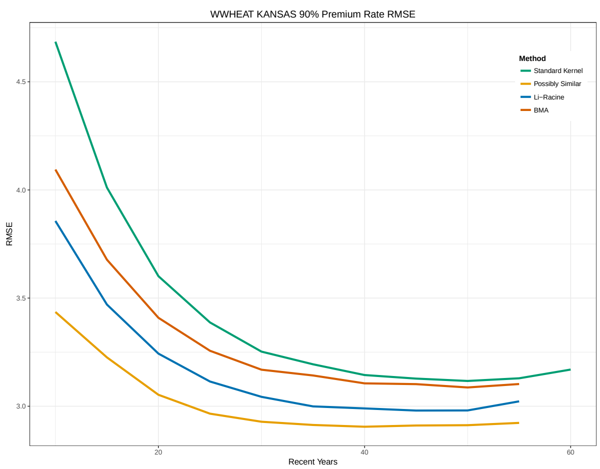

(B) Corn, $h$ by LCV



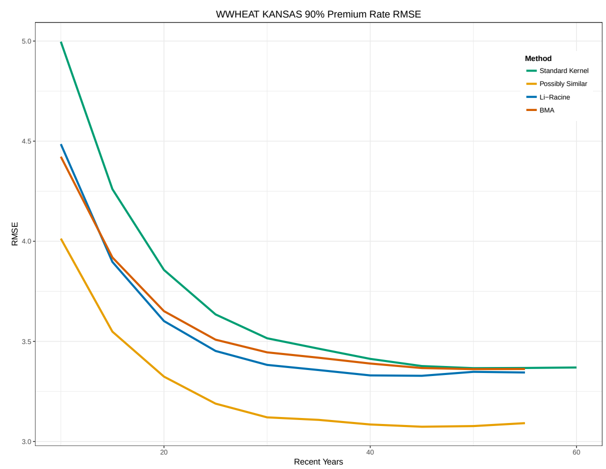

(C) Soybean, Optimal $h$




(D) Soybean, $h$ by LCV




(E) Winter Wheat, Optimal $h$




(F) Winter Wheat, $h$ by LCV

FIGURE 5. RMSE of 90% Premium Rate, RMA Detrending

From the results in Table 5, $p$-value 1 suggest that economically and statistically significant monies can be made using alternative means to incorporate some of the historical data for all state-crop combinations. $p$-value 2 suggest that using alternative means to incorporate some of the data leads to more accurate rates for all crops and regions. Specifically, results suggest that the Possibly Similar estimator leads to statistically significant more accurate rates in 3 of the 6 state-crop combinations, while the rest 3 cases are ambiguous. For the Li-Racine estimator, it produces statistically significant more accurate rates in 1 of the 6 state-crop combinations. The other 5 cases are ambiguous. The BMA estimator leads to statistically significant more accurate rates in 2 of the 6 state-crop combinations. The other 2 cases are ambiguous.

Now we revisit the performance of estimators we proposed, the standard kernel and three borrowing-information estimators under the mixture model from the Distribution 3. The graphs of MISE for under each estimator using RMA detrending methodology are shown in Figure 4. The figures of RMSE of 90% premium rate are shown in Figure 5. The MISE and RMSE using other detrending methodologies are shown in the Appendix. Overall, MISE and RMSE results are fairly consistent across detrending methodologies. For the same estimator, the MISE and RMSE are minimized below full data set. Estimators with smoothing parameter $h$ known (optimal $h$ by minimizing integrated squared errors) use less historical data to reach their minimum MISE and RMSE values. The borrowing estimators minimize error with less historical data and perform better than standard kernel estimator. Specifically, possibly similar method tends to perform best at minimizing MISE. For all three crops, corn uses least historical data followed by soybeans and winter wheat, which is consistent with technological advances. Comparing MISE and RMSE, there are more historical data used when estimating tails.

## Conclusions

County crop yield data from NASS of USDA has been the major data source for extensive literatures. Commonly used methodology in the literature adjusts only the first and second moments of the yields, then assumes the adjusted data are i.i.d.s without further investigation. Meanwhile, there have been significant innovations in farm management and seed technologies. Particular, the widely planted GM crops have fundamentally change the landscape of U.S. agriculture. It is reasonable to conjecture that these technological changes may alter higher moments of the yield DGPs more than just the first two. On the other hand, the changes of higher moments is difficult to detect by popular structural tests since they are of low power to test these changes, suffering likely type II errors.

In this manuscript, we attempt to answer the question how much historical yield data should be used. Obviously the answer is dependent on the empirical question at hand, loss function, and methodology. Our focus is on rating crop insurance contracts. Meanwhile, we propose methodologies that can use data from possibly like DGPs to improve estimation efficiency so we do not necessarily discard historical yield data. Results suggest that for corn and soybeans the identically distributed assumption does not likely hold, and significant efficiency gains (MISE and MSE) from using estimators that can incorporate historical data in alternative ways. Our results coincide with plant scientists that suggest much greater technological change in corn followed by soybeans. The RMA might wish to restrict the amount of historical data used in the rating of their area-based products (similar to their shallow-loss products) for crop and possibly region specific. Our future study following this direction could extend to borrowing information across space and time to increase efficiency. Intuitively, neighboring counties would have similar yield DGPs.

## References


Annan, F., J. Tack, A. Harri, and K. Coble. 2013. "Spatial Pattern of Yield Distributions: Implications for Crop Insurance." *American Journal of Agricultural Economics*, pp. aat085.

Assefa, Y., P. Prasad, P. Carter, M. Hinds, G. Bhalla, R. Schon, M. Jeschke, S. Paszkiewicz, and I.A. Ciampitti. 2017. "A New Insight into Corn Yield: Trends from 1987 through 2015." *Crop Science*, pp. .

Barrows, G., S. Sexton, and D. Zilberman. 2014. "Agricultural biotechnology: The promise and prospects of genetically modified crops." *The Journal of Economic Perspectives* 28:99–119.

Challinor, A.J., J. Watson, D. Lobell, S. Howden, D. Smith, and N. Chhetri. 2014. "A meta-analysis of crop yield under climate change and adaptation." *Nature Climate Change* 4:287–291.

Claassen, R., D. Hellerstein, and S.G. Kim. 2013. "Using mixed logit in land use models: can expectation-maximization (EM) algorithms facilitate estimation?" *American journal of agricultural economics* 95:419–425.

Claassen, R., and R.E. Just. 2011. "Heterogeneity and distributional form of farm-level yields." *American journal of agricultural economics* 93:144–160.

Claassen, R., C. Langpap, and J. Wu. 2017. "Impacts of Federal Crop Insurance on Land Use and Environmental Quality." *American Journal of Agricultural Economics* 99:592–613.

Cooper, J., A. Nam Tran, and S. Wallander. 2017. "Testing for Specification Bias with a Flexible Fourier Transform Model for Crop Yields." *American Journal of Agricultural Economics* 99:800–817.

Cooper, J.C. 2010. "Average crop revenue election: A revenue-based alternative to price-based commodity payment programs." *American Journal of Agricultural Economics*, pp. aaq048.

Deng, X., B.J. Barnett, and D.V. Vedenov. 2007. "Is there a viable market for area-based crop insurance?" *American Journal of Agricultural Economics* 89:508–519.

Duvick, D.N. 2005. "The contribution of breeding to yield advances in maize (Zea mays L.)." *Advances in agronomy* 86:83–145.

Egli, D. 2008. "Comparison of corn and soybean yields in the United States: historical trends and future prospects." *Agronomy journal* 100:S–79.

Egli, D.B. 2017. *Seed biology and yield of grain crops*. CABI.

Everitt, B.S., and D.J. Hand. 1981. *Finite Mixture Distributions*. Springer.

Fernandez-Cornejo, J., S. Wechsler, M. Livingston, and L. Mitchell. 2014. "Genetically engineered crops in the United States.", pp. .

Fernandez-Cornejo, J., et al. 2004. "The seed industry in US agriculture: An exploration of data and information on crop seed markets, regulation, industry structure, and research and development." Working paper, United States Department of Agriculture, Economic Research Service.

Goodwin, B.K. 2009. "Payment limitations and acreage decisions under risk aversion: a simulation approach." *American Journal of Agricultural Economics* 91:19–41.

Goodwin, B.K., and A. Hungerford. 2015. "Copula-based models of systemic risk in US agriculture: implications for crop insurance and reinsurance contracts." *American Journal of Agricultural Economics* 97:879–896.

Goodwin, B.K., and A.P. Ker. 1998. "Nonparametric estimation of crop yield distributions: implications for rating group-risk crop insurance contracts." *American Journal of Agricultural Economics* 80:139–153.

Goodwin, B.K., and A.K. Mishra. 2006. "Are decoupled farm program payments really decoupled? An empirical evaluation." *American Journal of Agricultural Economics* 88:73–89.

Goodwin, B.K., M.L. Vandeveer, and J.L. Deal. 2004. "An empirical analysis of acreage effects of participation in the federal crop insurance program." *American Journal of Agricultural Economics* 86:1058–1077.

Harri, A., K.H. Coble, A.P. Ker, and B.J. Goodwin. 2011. "Relaxing heteroscedasticity assumptions in area-yield crop insurance rating." *American Journal of Agricultural Economics*, pp. aar009.

Hendricks, N.P., A. Smith, and D.A. Sumner. 2014. "Crop supply dynamics and the illusion of partial adjustment." *American Journal of Agricultural Economics*, pp. aau024.

Hjort, N.L., and I.K. Glad. 1995. "Nonparametric density estimation with a parametric start." *The Annals of Statistics* 23(3):882–904.

Hoeting, J.A., D. Madigan, A.E. Raftery, and C.T. Volinsky. 1999. "Bayesian model averaging: A tutorial." *Statistical science* 14(4):382–401.

Ker, A.P. 2016. "Nonparametric estimation of possibly similar densities." *Statistics & Probability Letters* 117:23–30.

Ker, A.P., and K. Coble. 2003. "Modeling conditional yield densities." *American Journal of Agricultural Economics* 85:291–304.

Ker, A.P., and B.K. Goodwin. 2000. "Nonparametric estimation of crop insurance rates revisited." *American Journal of Agricultural Economics* 82(2):463–478.

Ker, A.P., and Y. Liu. 2016. "Bayesian model averaging of possibly similar nonparametric densities." *Computational Statistics*, pp. 1–17.

Ker, A.P., and P. McGowan. 2000. "Weather-based adverse selection and the US crop insurance program: The private insurance company perspective." *Journal of Agricultural and Resource Economics*, pp. 386–410.

Ker, A.P., T.N. Tolhurst, and Y. Liu. 2016. "Bayesian estimation of possibly similar yield densities: implications for rating crop insurance contracts." *American Journal of Agricultural Economics* 98:360–382.

Klümper, W., and M. Qaim. 2014. "A meta-analysis of the impacts of genetically modified crops." *PloS one* 9:e111629.

Koundouri, P., and N. Kourogenis. 2011. "On the distribution of crop yields: does the central limit theorem apply?" *American Journal of Agricultural Economics* 93:1341–1357.

Kucharik, C.J., and N. Ramankutty. 2005. "Trends and variability in US corn yields over the twentieth century." *Earth Interactions* 9:1–29.

Leng, G. 2017. "Recent changes in county-level corn yield variability in the United States from observations and crop models." *Science of The Total Environment* 607:683–690.

Li, D., L. Simar, and V. Zelenyuk. 2016. "Generalized nonparametric smoothing with mixed discrete and continuous data." *Computational Statistics & Data Analysis* 100:424–444.

Li, Q., and J. Racine. 2003. "Nonparametric estimation of distributions with categorical and continuous data." *journal of multivariate analysis* 86:266–292.

Marron, J.S., and M.P. Wand. 1992. "Exact mean integrated squared error." *The Annals of Statistics* 20(2):712–736.

Miao, R., M. Khanna, and H. Huang. 2016. "Responsiveness of crop yield and acreage to prices and climate." *American Journal of Agricultural Economics* 98:191–211.

Naranjo, S.E., J.R. Ruberson, H.C. Sharma, L. Wilson, and K. Wu. 2008. "The present and future role of insect-resistant genetically modified cotton in IPM." In *Integration of insect-resistant genetically modified crops within IPM programs*. Springer, pp. 159–194.

Naylor, R., W. Falcon, and E. Zavaleta. 1997. "Variability and growth in grain yields, 1950-94: does the record point to greater instability?" *Population and Development Review*, pp. 41–58.

Norwood, B., M.C. Roberts, and J.L. Lusk. 2004. "Ranking crop yield models using out-of-sample likelihood functions." *American Journal of Agricultural Economics* 86:1032–1043.

Ortiz-Bobea, A., and R.E. Just. 2013. "Modeling the structure of adaptation in climate change impact assessment." *American Journal of Agricultural Economics* 95:244–251.

Racine, J., and A.P. Ker. 2006. "Rating crop insurance policies with efficient nonparametric estimators that admit mixed data types." *Journal of Agricultural and Resource Economics* 31(1):27–39.

Racine, J., and Q. Li. 2004. "Nonparametric estimation of regression functions with both categorical and continuous data." *Journal of Econometrics* 119:99–130.

Ramirez, O.A., S. Misra, and J. Field. 2003. "Crop-yield distributions revisited." *American Journal of Agricultural Economics* 85:108–120.

Reilly, J., F. Tubiello, B. McCarl, D. Abler, R. Darwin, K. Fuglie, S. Hollinger, C. Izaurralde, S. Jagtap, J. Jones, et al. 2003. "US agriculture and climate change: new results." *Climatic Change* 57:43–67.

Reilly, J.M., and K.O. Fuglie. 1998. "Future yield growth in field crops: what evidence exists?" *Soil and Tillage Research* 47:275–290.

Roberts, M.J., W. Schlenker, and J. Eyer. 2013. "Agronomic weather measures in econometric models of crop yield with implications for climate change." *American Journal of Agricultural Economics* 95:236–243.

Tack, J., A. Harri, and K. Coble. 2012. "More than mean effects: modeling the effect of climate on the higher order moments of crop yields." *American Journal of Agricultural Economics* 94:1037–1054.

Tolhurst, T.N., and A.P. Ker. 2015. "On technological change in crop yields." *American Journal of Agricultural Economics* 97:137–158.

USDA. 2016. "Adoption of genetically engineered crops in the US." Working paper, United States Department of Agriculture, Economic Research Service.

Woodard, J.D., A.D. Pavlista, G.D. Schnitkey, P.A. Burgener, and K.A. Ward. 2012. "Government insurance program design, incentive effects, and technology adoption: the case of skip-row crop insurance." *American Journal of Agricultural Economics* 94:823–837.

Woodard, J.D., and B.J. Sherrick. 2011. "Estimation of mixture models using cross-validation optimization: Implications for crop yield distribution modeling." *American Journal of Agricultural Economics* 93:968–982.

Woodard, J.D., and L.J. Verteramo-Chiu. 2017. "Efficiency Impacts of Utilizing Soil Data in the Pricing of the Federal Crop Insurance Program." *American Journal of Agricultural Economics* 99:757–772.

Wu, J., R.M. Adams, C.L. Kling, and K. Tanaka. 2004. "From microlevel decisions to landscape changes: an assessment of agricultural conservation policies." *American Journal of Agricultural Economics* 86:26–41.

Yvette Zhang, Y. 2017. "A Density-Ratio Model of Crop Yield Distributions." *American Journal of Agricultural Economics*, pp. .