

**On Technological Change and Volatility
in Canadian Agriculture**

**by
Horlick Ng**

A Thesis
presented to
The University of Guelph

In partial fulfillment of requirements
for the degree of
Master of Science
in
Food, Agricultural and Resource Economics

Guelph, Ontario, Canada
© Horlick Ng, April, 2019

ABSTRACT

ON THE TECHNOLOGICAL CHANGE AND VOLATILITY IN CANADIAN AGRICULTURE

Horlick Ng

University of Guelph, 2019

Advisor:

Professor Alan P. Ker

Feeding nine billion people by 2050 is arguably the most pressing challenge in global agriculture. Increasing crop productivity, reducing animal-based protein consumption and reducing food waste have been forwarded as possible solutions. However, reducing the latter two is unlikely with rising global wealth; therefore, our ability to meet future food demand will, as in the past, be dictated by technological change. Technological change not only increases average crop yield, it may also increase or decrease yield volatility. Oftentimes, yield volatility does not change identically between upper and lower tails. The lower tail, which represents the downside yield risk, has significant economic consequences. Increases in downside yield risk contributes to food insecurity and subsequently economic insecurity for agricultural-based economies (i.e. developing countries). Conversely, in developed countries, significant public funds are funnelled to the farmers in order to mitigate downside yield risk. This thesis focuses on the effects of technological change on yield volatility for the major crops in Canada. Specifically, historical county-level yield data in Ontario and Saskatchewan were analyzed for barley, canola, corn, oats, soybean and wheat. The overall results suggest that all province-crop pairs generally exhibit a higher yield volatility over time in different magnitudes. The implications are of particular interest to agricultural risk management policy as well as public policy on agricultural research and development.

ACKNOWLEDGEMENT

First and foremost, I would like to thank my thesis advisor Dr. Alan Ker. The door to Alan's office was always open whenever I had a question about my research. He has shown me, by his example, what a good researcher should be. Besides my advisor, I would also like to thank my thesis committee member, Dr. Yu Na Lee, for her valuable insight into my thesis. Thank you Dr. Getu Hailu for being my external examiner, and also the incredible experience to be your teaching assistant for the Futures and Options class. Thank you Dr. Alfons Weersink for encouraging me to pursue this program. I would also like to thank the Ontario Ministry of Agriculture, Food and Rural Affairs, and Institute for the Advanced Study of Food and Agricultural Policy for funding this project.

I would like to thank my cohort in alphabetical order: Addy, Anna, Dani, Emma, Izzy, Jamie, Jamieson, Joe, Keping, and Yi. This two-year experience was truly amazing. I sincerely wish you all the best in your future endeavors.

Finally but importantly, I would like to express my gratitude to my parents and to my girlfriend, Kay, for providing me with continuous encouragement and support throughout my years of study.

Contents

Abstract	ii
Acknowledgement	iii
Contents	iv
List of Figures	v
List of Tables	vi
1 Introduction	1
2 Data	6
3 Empirical Methods	11
3.1 Literature Review	11
3.2 Mixture Model	14
3.3 Model Selection and Estimation Strategy	15
4 Empirical Results and Implications	27
4.1 Estimation Results	27
4.2 $H_0^1 : \beta_\ell = \beta_u$	36
4.3 $H_0^2 : \lambda_t = \lambda$	38
5 Economic Implications for Crop Insurance	42
6 Conclusions	46
Bibliography	49
7 Appendix	55
7.1 Summary Statistics of Estimated Parameters, by Province and Crop	55

List of Figures

1	Middlesex County, Ontario, Corn Yields, 1949 - 2016.	2
2	Ontario Counties, by Crop	8
3	Saskatchewan Counties, by Crop	9
4	Representative Conditional Quantile Regression Estimates for Each Crop . .	10
5	EM Algorithm and Penalized Log-likelihood Estimates, Prince Edward County, Ontario, Corn Yields, 1949 - 2016.	19
6	Crossing and Non-crossing Technological Trends, Brant County, Ontario, Corn Yields, 1949 - 2016.	20
7	AIC Estimates of Each Model in Box Plots, by Crop in Ontario	22
8	AIC Estimates of Each Model in Box Plots, by Crop in Saskatchewan	23
9	Hoodoo, Saskatchewan Barley	28
9	Wolseley, Saskatchewan Canola	29
9	Peel, Ontario Corn	30
9	Middlesex, Ontario Soybean	31
9	Frenchman Butte, Saskatchewan Oats	32
9	Abernethy, Saskatchewan Wheat	33
10	Ratio of Expected Yield Loss at 90% Coverage Level	45

List of Tables

1	Summary of Data	7
2	Overview of Models for Selection	16
3	Median of AIC Estimates of Each Model, by Crop and Province	22
4	Number of Counties that Obtained the Lowest AIC Estimate, by Crop and Province	23
5	Restriction and Rejection of Restricted Model, by Crop and Province	25
6	Median of Estimated Parameters, by Crop and Province	35
7	Hypothesis Test One and Two: Rejection Counts	37
8	Summary Statistics for Ratio of $\hat{\beta}_u/\hat{\beta}_\ell$	37
9	Summary Statistics for the Regression Coefficient of $\hat{\lambda}_t$	40
10	Summary Statistics of Estimated Parameters, by Province and Crop, Part 1 of 2	55
11	Summary Statistics of Estimated Parameters, by Province and Crop, Part 2 of 2	56

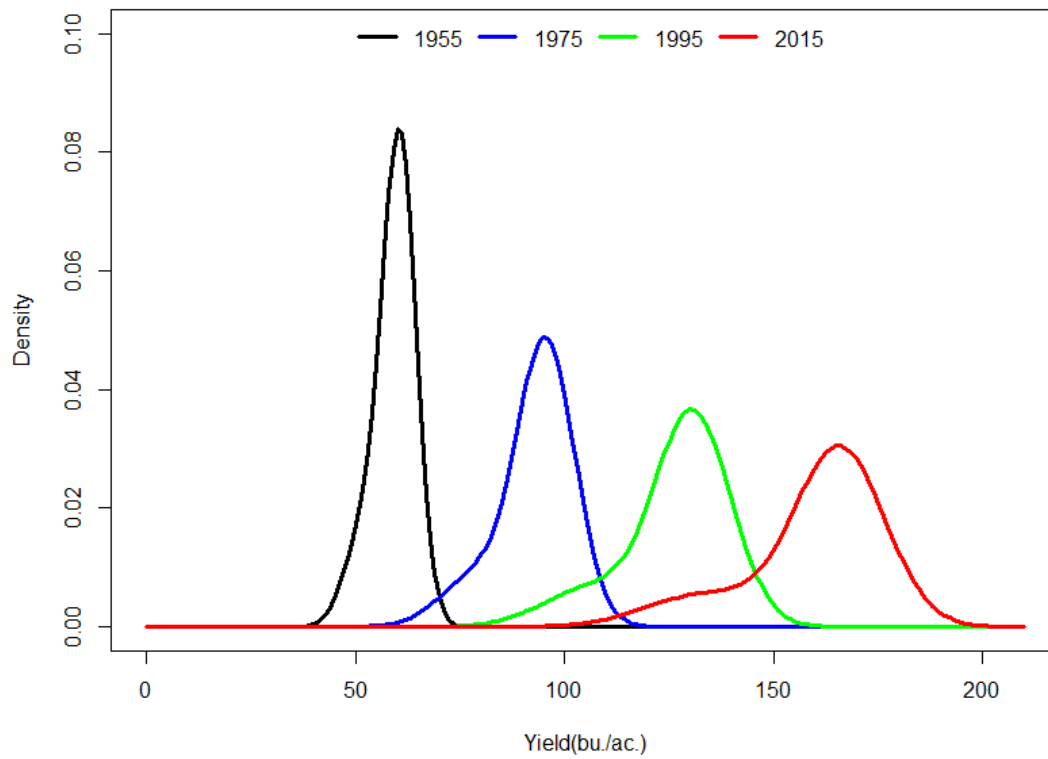
1 Introduction

Meeting the expected food demand by the mid-21st century is one of the most serious challenges in global agriculture (Pretty et al., 2010; Conforti et al., 2011; Searchinger et al., 2014; McKenzie and Williams, 2015). Technological change has played a dominant role in feeding the rapidly-growing population by increasing average crop yields through innovations in fertilizer, farming equipment, pesticides and seed genomics.¹ In addition to technological change, other possible solutions include shifting diets by reducing animal-based protein consumption (White, 2000; Pimentel and Pimentel, 2003; De Boer and Aiking, 2011; Ranganathan et al., 2016) and reducing food waste (Parfitt, Barthel, and Macnaughton, 2010; Kummu et al., 2012; Gustafsson et al., 2013; Lipinski et al., 2013; Grafton, Daugbjerg, and Qureshi, 2015). However, as more people in developing countries earn higher disposable income, the per capita meat consumption can be expected to continue to increase (Machovina, Feeley, and Ripple, 2015). Moreover, improved food affordability for these consumers diminishes their incentives to avoid food waste (Godfray et al., 2010). As a result, reducing either animal-based protein consumption and the amount of food waste are particularly challenging. Therefore, our ability to meet food demand has been, and will likely continue to be, dictated by technological change. Historically, technological change has led to a sizable increase in average crop yields. For example, the average crop yield in the United States for corn has increased by more than five times and soybean has more than doubled since 1940 (U.S. Department of Agriculture, 2018).²

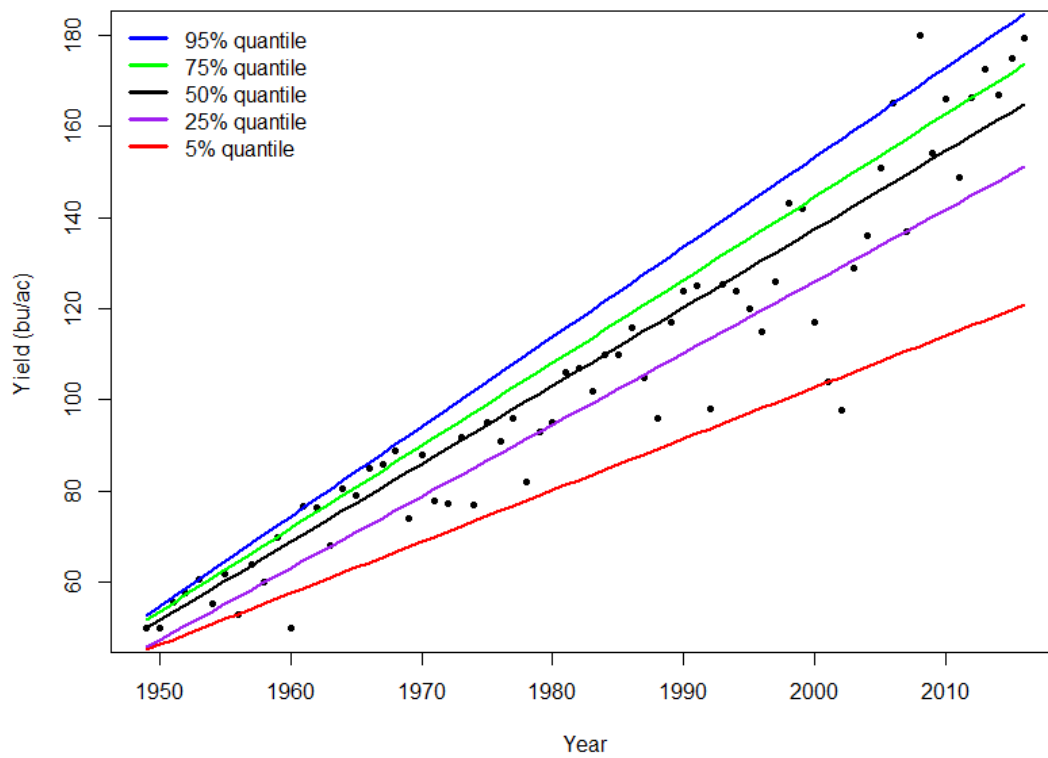
Although technological change increases the biological limit of crop yields over time, it also alters yield volatility. To illustrate this relationship and provide some intuition, Figure 1(a) presents the estimated conditional yield distribution in 1955, 1975, 1995 and 2015

¹Technological change refers to the adoption of innovation that results in an increase in average crop yield. Average crop yield refers to crop yield per unit area of agricultural land, measured in bushels per acre. It is typically used to measure crop productivity in literature, as a result of factors such as technological change, climate change, soil quality and policy change.

²The United States is currently the world’s largest producer of corn and soybean.



(a) Estimated Conditional Yield Distribution



(b) Conditional Quantile Regression

Figure 1: Middlesex County, Ontario, Corn Yields, 1949 - 2016.

for corn yields from Middlesex County, Ontario.³ In 1955, the yield distribution appears to be relatively symmetrical. The following patterns are evident over the course of time: (i) technological change increases average crop yield as the yield distribution shifts to the right; (ii) yield distribution becomes more dispersed, which suggests an increase in overall yield volatility; and (iii) yield distribution develops a heavier lower tail, which indicates a higher probability of low yield (hereafter, downside yield risk). Figure 1(b) presents the corresponding conditional quantile regression at 5%, 25%, 50%, 75% and 95% (Koenker and Bassett Jr, 1978). It clearly shows that the rate of technological change does not increase identically across the quantiles of the distribution, which suggests that technological change appears to impact the middle and upper tails of the yield distribution at a greater rate than in the lower tail. Both Figure 1(a) and 1(b) clearly illustrate that the implementation of technological change does not simply shift the entire distribution to the right (i.e. it does not only shift at the mean), but also affects the higher moments of the distribution.

Analogously, a yield distribution can be seen as a mixture of various components, and technological change contributes at a heterogeneous rate to these components. For instance, innovative seed varieties improve average crop yields over time under ideal growing conditions, but may not outperform their predecessors under catastrophic events. In this case, technological change pushes the upper tail of the yield distribution to the right while the lower tail lags behind and, therefore, the yield distribution exhibits a different rate of technological change across different components. While there exists literature modelling technological change in U.S. yields (i.e. Skees and Reed (1986); Kaylen and Koroma (1991); Ramirez (1997); Goodwin and Ker (1998); Just and Weninger (1999); Sherrick et al. (2004); Tolhurst and Ker (2015)), there is no corresponding literature in regards to Canadian crop yields. This thesis aims to fill the gap. The objectives of this thesis are: (i) to investigate whether crop yields have different rates of technological change across different components

³Conditional yield distribution refers to the estimated distribution of average crop yield in a given year. Middlesex county is used in this example as the county produced the most corn in 2017 in the province of Ontario.

of the yield distribution; (ii) to estimate the effect of technological change on yield volatility; (iii) to examine whether the trends vary across crops and provinces; and (iv) to estimate whether crop insurance payouts (which is related to downside yield risk) will vary in the next five and ten years. To this end, this thesis uses county-level crop yield data for six field crops in the provinces of Ontario and Saskatchewan, which account for 9% and 46.8% of the Canadian field crop area, respectively (Statistics Canada, 2016).⁴ Tolhurst and Ker (2015) proposed modeling crop yields using finite mixture of normals (hereafter, mixture model) with embedded trend functions for potentially different rates of technological change across components. Their approach assumes component variances to be constant over time. This thesis generalizes their approach by allowing the component variances to vary over time, and tests whether there is a potential efficiency gain in estimating conditional yield distributions as opposed to the approach from Tolhurst and Ker (2015). These estimates will provide answers to the following questions: (i) what is the optimal number of components to model yields; (ii) does the rate of technological change in one component outperform another; (iii) are the probabilities of components changing over time; and (iv) are yield volatility and downside yield risk changing over time? The answers to these questions will be used to accomplish the thesis objectives.

The above-mentioned questions are worthwhile to investigate because they have economic significance; changes in yield volatility and downside yield risk have important implications to areas such as farm income variability, crop insurance and food security. Given the fact that crop yields are closely linked to climate variation (Lobell and Asner, 2003; Lobell and Field, 2007; Challinor et al., 2014), agricultural producers have no control over the impacts of weather on their yields. As the vast majority of crops get harvested only once a year in Canada, the inflexibility of farm income is detrimental because poor outcomes can be fatal to the year-to-year survival of farm businesses. Consequently, significant public funds

⁴Field crop area is used rather than the quantity of crop production in the comparison because major crops differ across provinces. For instance, western Canada mainly produces barley, canola, oats, and spring wheat, whereas southern Canada mainly produces corn, soybean, and winter wheat.

(\$3.23 billion between 2013 and 2018 under Business Risk Management programs and \$3 billion between 2018 and 2023 under Canadian Agricultural Partnership) are funnelled to assist farmers in mitigating production risks, partly by subsidizing crop insurance premiums, which are cost-shared among the federal government, provincial governments and producers at a 36:24:40 ratio, respectively (Agriculture and Agri-Food Canada, 2017). Downside yield risk also contributes to food insecurity, notably in developing countries, which is a major impediment to their economic development. On account of the interrelationship between average crop yields and crop prices, rising food costs due to poor yields causes inadequate access to food for vulnerable households, which pushes them further into food insecurity. In Canada – a developed country – the concern about food insecurity is by no means trivial as well. In 2014, one in eight households experienced food insecurity in Canada and more than one in six children lived in these affected households (Tarasuk, Mitchell, and Dachner, 2016). Although developed countries tend to have a higher capacity to alleviate the impacts of poor yield on food insecurity through food imports, food banks and food subsidies, the rate of food insecurity surprisingly did not improve over the years in most part of Canada since 2005. Thus, on the whole, more attention should be paid to understand better the relationship between technological change and yield volatility.

The thesis proceeds as follows. The next chapter describes the data used for the empirical analyses. The following chapter outlines the current approach to modelling crop yields, details the empirical methods used in this thesis, and explains the estimation strategy. This is followed by the estimation and hypothesis test results. The second last chapter discusses the economic implications of the empirical results to expected yield loss (crop insurance premium). The final chapter presents the summary of the thesis.

2 Data

Ideally, farm-level yield data would be used to empirically investigate the effects of technological change such that the data could accurately reflect each producer’s adoption of innovation. Unfortunately, such data does not exist as there is always a trade-off between data availability and disaggregation of data. Thus, county-level data were used in this thesis, which are the least aggregated data that are available to the public with sufficient number of counties and length of data period. Analyzing crop yield data at a more aggregated level could be a concern because it averages out the heterogeneity of less aggregated data, particularly with respect to the higher moments. While aggregating farm-level yields to the county-level will mute the effects of technological change, it should not be completely mitigated. To the extent that yields are spatially correlated within a county, the mitigation effects should be lessened. In addition, with the exception of a few studies, modelling crop yields in literature typically use county-level yield data. Nonetheless, the effects this analysis empirically finds should be conservative relative to the effects at the farm level since the variation at the farm level is generally considered doubled or more than that at the county level (Coble, Dismukes, and Thomas, 2007; Cooper et al., 2009; Claassen and Just, 2011). The entire data set includes six field crops in the provinces of Ontario and Saskatchewan, which consists of a total of 810 county-crop pairs. Note, only counties that have complete yield histories throughout the data period were included in the analysis. Alberta and Manitoba yield data were considered; however, only 40 years and 25 years of yield data, respectively, are available to the public. The analysis was done for both provinces, but is not reported in this thesis to avoid ambiguous conclusions. Table 1 summarizes the data used in the analysis. A detailed data description for each province is as follows.

For Ontario, county-level corn, soybean and winter wheat yield data from the period 1949 to 2016 were used. Yield data were collected from the annual Agricultural Statistics Reports published by the Ontario Ministry of Agriculture, Food, and Rural Affairs (OMAFRA). Ontario is a major producer of corn, soybean and winter wheat, which accounted for 62%,

Table 1: Summary of Data

Province	Observations		Summary Statistics of Crop Yield (bu/ac)				
Crop	Counties	Period (Years)	Min.	Mean	Median	Max.	Std. Dev.
Ontario							
Corn	32	1949 – 2016 (68)	20.2	92.9	85.0	192.9	33.3
Soybean	6	1949 – 2016 (68)	16.0	33.0	32.0	56.4	9.0
Wheat	26	1949 – 2016 (68)	22.0	50.8	46.3	109.5	17.1
Saskatchewan							
Barley	204	1938 – 2016 (79)	1.0	37.8	38.0	102.7	15.5
Canola	144	1970 – 2016 (47)	1.5	23.0	22.2	59.5	8.2
Oats	131	1938 – 2016 (79)	1.0	48.7	48	155.8	20.6
Wheat	267	1938 – 2016 (79)	1.0	24.9	25.0	70.3	9.9

49.2% and 76.5% of national production in 2017, respectively (Statistics Canada, 2018). Figure 2 depicts the data for each crop. As shown in Figure 2(a), there are data for 32 counties available for corn (shaded in purple). Unfortunately, there are only data for six soybean counties with completed yield histories available as most of the soybeans were grown in the southwestern Ontario back in 1970s. These six counties are shaded in brown in Figure 2(b). For winter wheat, the 26 counties are highlighted in green in Figure 2(c).

For Saskatchewan, county-level barley, oats and spring wheat yield data from the period 1938 to 2016 were used. Due to limited data availability, county-level canola yield data are only included from 1970 to 2016. Yield data were collected from the Ministry of Agriculture. Saskatchewan is a major producer of barley, canola, oats and spring wheat, which accounted for 39.7%, 52.4%, 53.3% and 39.4% of national production in 2017, respectively (Statistics Canada, 2018). Similar to the Ontario crops, Figure 3 depicts the Saskatchewan data by each crop. For Saskatchewan barley, Figure 3(a) highlights the 204 counties in yellow. Interestingly, for canola, no data are available in the southwestern part of Saskatchewan; the 144 counties are shaded in red in Figure 3(b). For Saskatchewan oats, the 131 counties are shaded in blue in Figure 3(c). Finally, for Saskatchewan spring wheat, the 267 counties are highlighted in pink in Figure 3(d).

Figure 4 illustrates the representative conditional quantile regression estimates for each of the six crops. It clearly shows that technological change has remarkably increased average crop yield over time for all the crops. What is more interesting is that technological change has generally impacted the middle and upper quantiles at a greater rate than in the lower quantile, causing an increase in yield volatility over time. This phenomenon can be seen particularly in Ontario corn and Ontario soybean, but less so for the Saskatchewan crops.

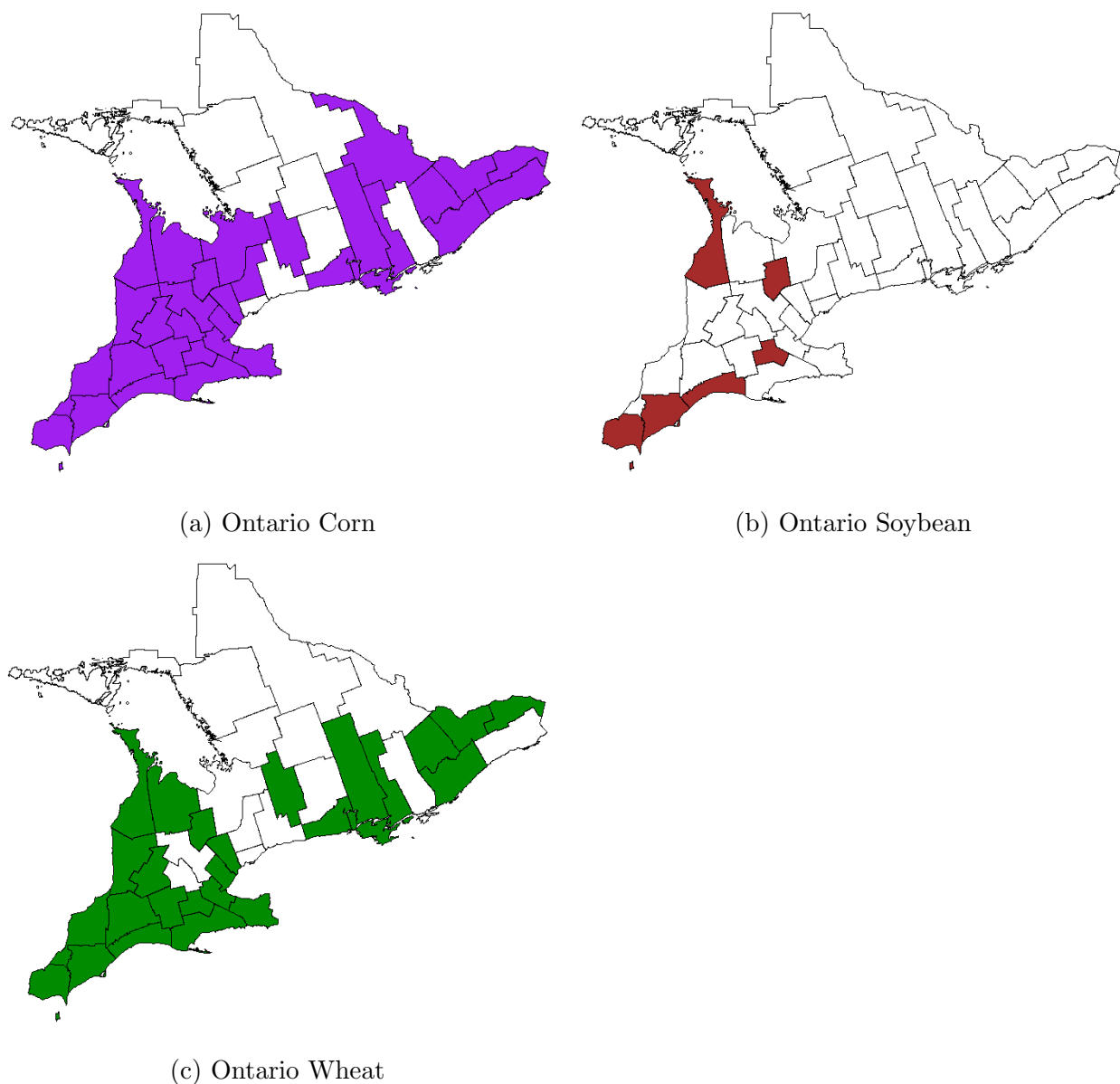
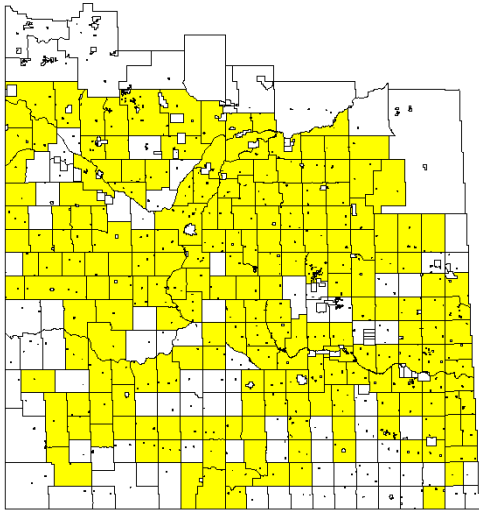
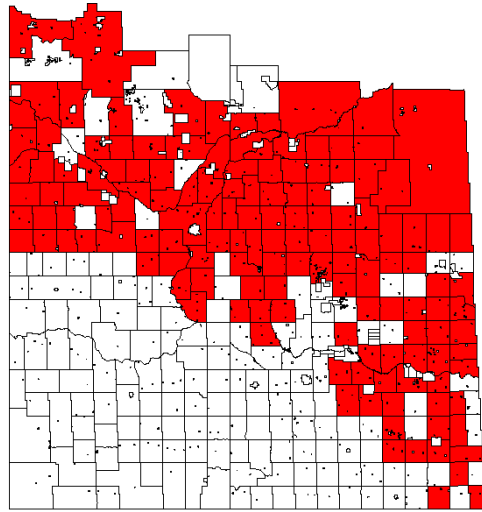


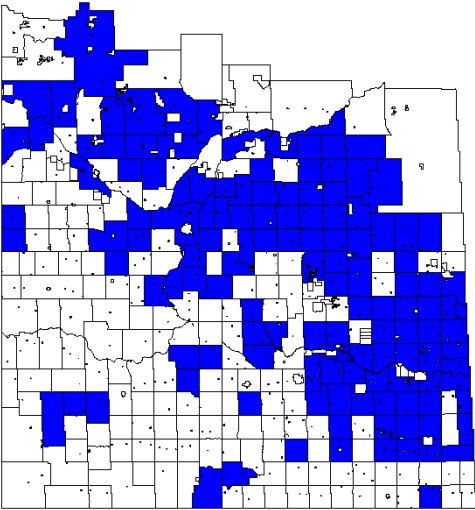
Figure 2: Ontario Counties, by Crop



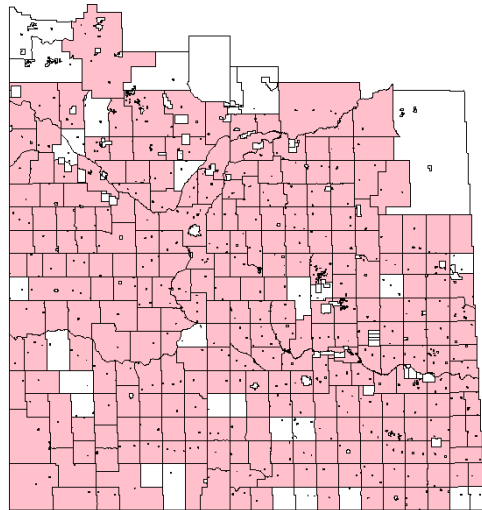
(a) Saskatchewan Barley



(b) Saskatchewan Canola

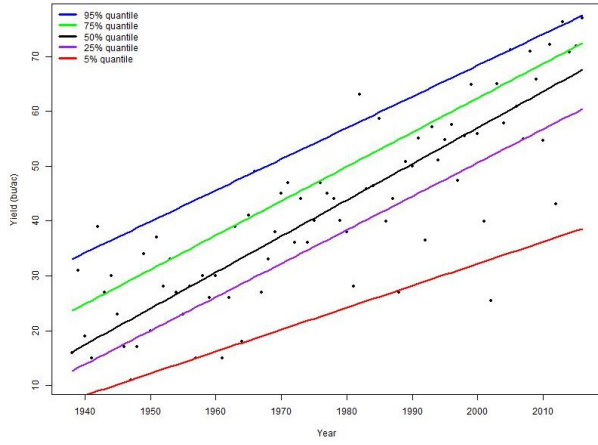


(c) Saskatchewan Oats

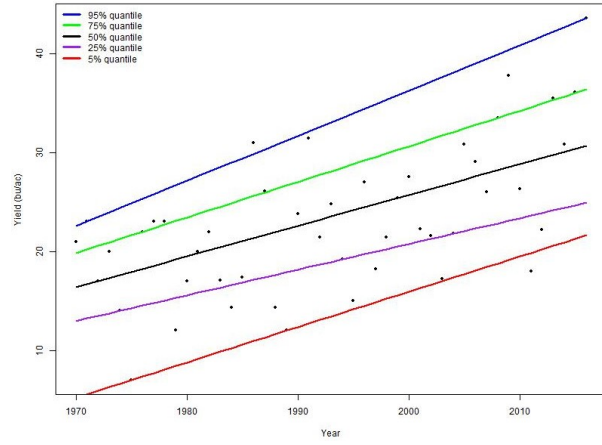


(d) Saskatchewan Wheat

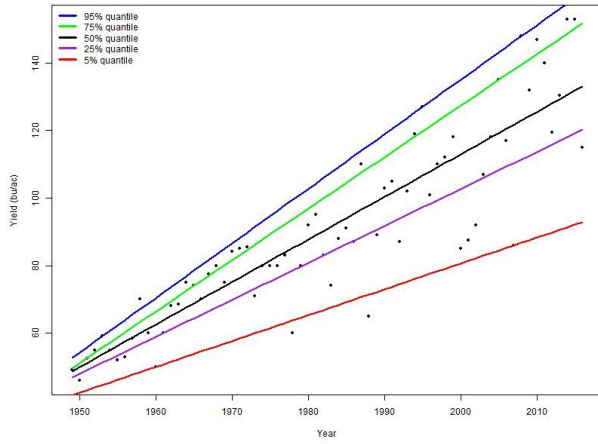
Figure 3: Saskatchewan Counties, by Crop



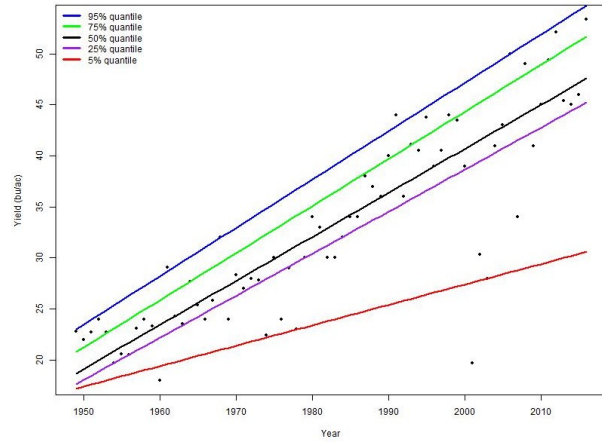
(a) Hoodoo, Saskatchewan Barley



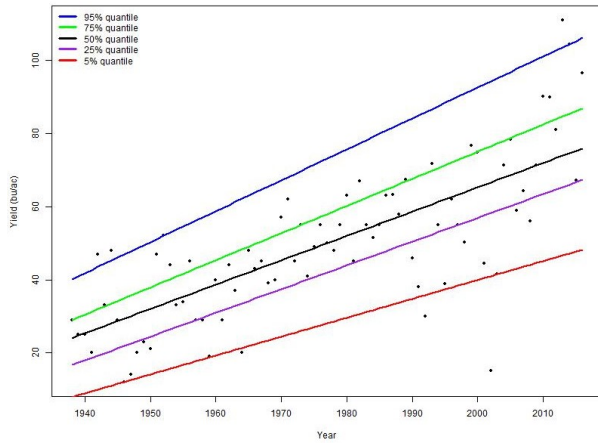
(b) Wolseley, Saskatchewan Canola



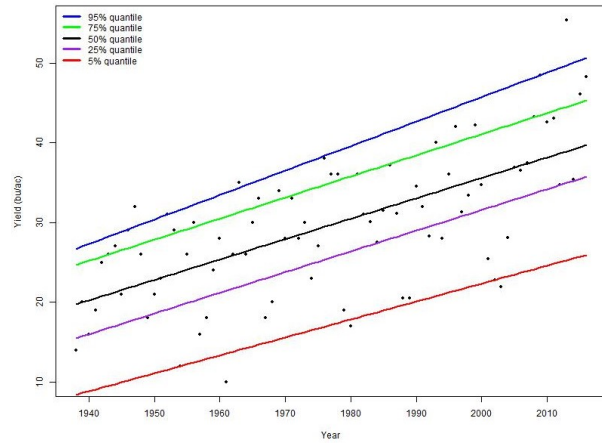
(c) Peel, Ontario Corn



(d) Middlesex, Ontario Soybean



(e) Frenchman Butte, Saskatchewan Oats



(f) Abernethy, Saskatchewan Wheat

Figure 4: Representative Conditional Quantile Regression Estimates for Each Crop

3 Empirical Methods

3.1 Literature Review

Based on existing literature, the typical approach to model crop yields involve three steps: (1) estimate a trend of crop yield over time; (2) estimate the residuals from the first step and adjust for heteroscedasticity if needed, and (3) estimate a yield distribution based on the results from step one and two. A detailed discussion for each step is as follows.

The time-conditional mean of average crop yield, or in other words, the rate of technological change, can be estimated through various approaches ranging from a simple linear trend to other examples including $ARIMA(p, d, q)$ (Goodwin and Ker, 1998), Bayesian hierarchical model (Ozaki and Silva, 2009), Kalman filter (Kaylen and Koroma, 1991), mixture of two normals (Tolhurst and Ker, 2015), polynomial trend (Just and Weninger, 1999), stochastic trend model (Moss and Shonkwiler, 1993) and two-knot linear spline (Skees and Reed, 1986; Harri et al., 2011).

Based on the estimated trend, the residuals can be tested for heteroscedasticity, or in other words, the effects of technological change on the second moment of a yield distribution. Conditioning for heteroscedasticity has received surprisingly little attention in modelling crop yields as previous studies tended to make untested assumptions or simply assume crop yields to be homoscedastic. However, recent contributions have suggested that correctly accounting for heteroscedasticity leads to significant economic consequences. Harri et al. (2011) concluded that the assumption of a specific form of heteroscedasticity limits actuarial soundness in crop insurance premium rate calculations. Ker and Tolhurst (2019) generalized Harri et al. (2011) to incorporate the asymmetric affects of technological change in the third moment of a yield distribution and found that premium rates can have statistically significant difference under asymmetric heteroscedasticity treatments.

Conditional yield distribution, which is commonly used for measuring risk, has been extensively investigated in a wide variety of contributions. In general, methods of modelling

yield distributions can be categorized into parametric, semi-parametric and non-parametric models. Parametric models consist of a finite number of parameters to model yields and, therefore, would require a prior assumption on the functional form of the distribution. Many of the previous studies adopted parametric methods and different types of distributions have been considered and tested, including but not limited to: beta distribution (Nelson and Preckel, 1989; Turvey, Zhao et al., 1999; Ozaki, Goodwin, and Shirota, 2008; Zhu, Goodwin, and Ghosh, 2011), gamma distribution (Gallagher, 1987), inverse hyperbolic sine distribution (Moss and Shonkwiler, 1993), logistic distribution (Atwood, Shaik, and Watts, 2003), log-normal distribution (Day, 1965), maximum entropy distribution (Stochs and LaFrance, 2004; Wu and Zhang, 2012; Tack, Harri, and Coble, 2012), mixture model (Woodard and Sherrick, 2011; Tolhurst and Ker, 2015), normal distribution (Botts and Boles, 1958; Just and Weninger, 1999), reverse lognormal distribution (Claassen and Just, 2011) and Weibull distribution (Chen, Miranda et al., 2004; Sherrick et al., 2004). One concern of the parametric methods is that the initial assumption may lead to a lack of flexibility and accuracy to model yields because the true yield distributions are always unknown.

In contrast to the parametric methods, the non-parametric methods do not require any specifications on the functional form of the distributions such that the shape of distribution is solely derived by the observations. Thus, the use of the non-parametric methods is extremely flexible and prevents any false assumptions. Kernel density estimation, one of the most popular non-parametric methods, has been considered to model yields by Goodwin and Ker (1998), Ker and Goodwin (2000), Goodwin and Mahul (2004), Norwood, Roberts, and Lusk (2004), Racine and Ker (2006), Ker and Tolga Ergun (2007) and Ker, Tolhurst, and Liu (2016). However, Ker and Coble (2003) pointed out that the non-parametric methods tend to be more inefficient than parametric methods in two circumstances: (1) when the prior assumption of the parametric distribution can accurately model yields and (2) a small number of observations. In light of their respective merits, they proposed a semi-parametric estimator to model yields that can account for the benefits of both parametric and non-

parametric methods while mitigating their shortfalls.

Although the "best" method to model crop yields can hardly be determined, many of the reviewed studies concluded that yield distributions tend to be non-normally distributed, meaning that the lower tail and upper tail of the distribution are not symmetrical. Early work includes Day (1965), who argued that yield distributions, in general, are not normally distributed or positively skewed and suggested that the degree of skewness and kurtosis depends on the type of crops and soil nutrient. Indeed, empirical evidence from subsequent literature has supported that the majority of field crops have negatively skewed yield distributions (Gallagher, 1987; Nelson and Preckel, 1989; Goodwin and Ker, 1998; Atwood, Shaik, and Watts, 2003; Ramirez, Misra, and Field, 2003; Tolhurst and Ker, 2015). One exception is cotton, which is well recognized to have positively skewed yield distribution (Day, 1965; Goodwin and Ker, 1998; Ramirez, Misra, and Field, 2003). All of the above-mentioned studies have considered the higher moments of the distribution with the exception of Botts and Boles (1958) and Just and Weninger (1999). The former study suggested that crop insurance premia be calculated based on a normally distributed yield distribution. However, the reason to assume normality was not explained. Since then, modelling yields using normal distribution had been silent until the latter study argued that crop yields are normally distributed because of misspecification and data limitation problems.

Nonetheless, considering the higher moments of yield distribution is crucial because they have significant economic consequences. For instance, Gallagher (1987) found that negatively skewed soybean yield distributions in the United States are associated with a higher chance of low yields. Goodwin and Ker (1998) and Atwood, Shaik, and Watts (2003) found that the assumption of normality, in the majority of cases, may understate the rate of crop insurance premia. The non-normality can be generated by different factors in a crop-specific fashion, including but not limited to, bad weather (Hennessy, 2009), crop rotation (Du, Hennessy, and Yu, 2012), number of overheat days (Du et al., 2015), response to fertilizers (Nelson and Preckel, 1989) and technological change (Tolhurst and Ker, 2015).

3.2 Mixture Model

As noted above, the vast majority of the yield distribution estimation approaches have considered the higher moments of a distribution. To this end, the mixture model is proposed to model yields in this thesis. Although the mixture model is a parametric model, it is able to recognize the higher moments of a distribution due to its exceedingly flexible parametric specifications (Goodwin, Roberts, and Coble, 2000). Such flexibility allows for the use of different numbers of parameters to model any shape on a continuous distribution function to any desired level of bounded error (Everitt and Hand, 1981). From an economic point of view, soft clustering the data corresponds to one of the thesis objectives, which is to investigate whether the rate of technological change differs across different components. The mixture model is widely used in different fields. Specifically, in agricultural economics, it has been used in areas such as agricultural commodity prices (Hall, Brorsen, and Irwin, 1989; Goodwin, Roberts, and Coble, 2000) and crop yields (Woodard and Sherrick, 2011; Tolhurst and Ker, 2015). The mixture model is defined as:

$$y_t \sim \sum_{k=1}^K \lambda_k \mathcal{N}(\mu_k, \sigma_k^2) \quad (1)$$

where y_t is the average crop yield over year t and $t = 1, \dots, T$ and where K is the number of mixture components (i.e. normal distributions). λ_k is the probability of a particular component k (subject to $\lambda_k > 0$ and $\sum_{k=1}^K \lambda_k = 1$) and is obtained by taking the mean of each data point's probability that it belongs to component k in each year (denoted by $w_{k,t}$) such that $\lambda_k = \sum_{t=1}^T w_{k,t}/T$. $\mathcal{N}(\mu_k, \sigma_k^2)$ are the normal distributions with mean μ_k and variance σ_k^2 . Given the yield data y_t , the motivation is to estimate the unknown parameters λ, μ and σ that maximize the log-likelihood function, which is:

$$\ln \mathcal{L}(\lambda, \mu, \sigma \mid y_t) = \sum_{t=1}^T \left(\ln \sum_{k=1}^K \lambda_k \mathcal{N}(\mu_k, \sigma_k^2 \mid y_t) \right) \quad (2)$$

3.3 Model Selection and Estimation Strategy

In order to estimate yield distribution for each county-crop pair by the mixture model, two decisions had to be made: (1) the optimal number of components (K) to include in the model and (2) the number of parameters used to estimate the probability, mean and variance for each component. Theoretically, the likelihood estimate should improve along with the increasing use of (1) and (2), however, this may overfit the data. In view of the overfitting problem, Akaike information criterion (AIC) was employed to balance the trade-off between the complexity of the model and goodness of fit (Akaike, 1974). AIC is defined as:

$$AIC = (-2) \ln L + 2p \quad (3)$$

where L is the maximum likelihood derived from the model and p represents the number of parameters used within the model. The penalized term $2p$ increases along with the number of parameters, which eventually inflates the AIC estimate. The preferred model, among a selection of models used in the comparison, is indicated by the lowest AIC estimate.⁵ Only models with one and two components were considered in the comparison. As suggested by the AIC estimates, including more than two components overfits the data. Table 2 provides an overview of the models that are used in the AIC comparison. A detailed explanation of each model is as follows.

The first model has one component. Note, $K = 1$ represents a conventional normal distribution, which is:

$$y_t \sim \mathcal{N}(\alpha + \beta t, \gamma) \quad (\text{Model 1})$$

where the mean of the component μ was estimated by a linear time trend $\alpha + \beta t$. In other words, it represents the mean rate of technological change over time. The component variance σ^2 is represented by γ . Model 1 assumes γ to be constant over time.

⁵The purpose of AIC is to evaluate the relative efficiency between models. In this thesis, the model with the lowest AIC estimate is the preferred model; however, it is by no means certain that such a model is the best among the selection of models.

Table 2: Overview of Models for Selection

Model	Number of Components	Number of Technological Trends	Restriction	Mean(s)	Variance(s)
1	One	One	No	Temporal	Constant
2	One	One	No	Temporal	Temporal
3	Two	One	No	Temporal	Constant
4	Two	Two	No	Temporal	Constant
4R	Two	Two	Yes	Temporal	Constant
5	Two	Two	No	Temporal	Temporal
5R	Two	Two	Yes	Temporal	Temporal

The next model also has one component. The component mean, which is the same as Model 1, is conditional to time with a linear trend; however, the variance is allowed to change over time for the adjustment of heteroscedasticity, where $\sigma^2 = \gamma + \delta t$. That is:

$$y_t \sim \mathcal{N}(\alpha + \beta t, \gamma + \delta t) \quad (\text{Model 2})$$

After considering one-component models, the following models are two-component (i.e. $K = 2$) such that the estimated yield distribution is a mixture of two normal distributions. Intuitively, the two components can be interpreted as the lower component (i.e. average crop yields under poor growing conditions) and the upper component (i.e. average crop yields under ideal growing conditions), which are denoted by subscript ℓ and u , respectively. The first two-component model is given by:

$$y_t \sim \lambda \mathcal{N}(\alpha + \beta t, \gamma_\ell) + (1 - \lambda) \mathcal{N}(\alpha + \beta t, \gamma_u) \quad (\text{Model 3})$$

where this model assumes both components to have the same rate of technological change over time, which are both represented by $\alpha + \beta t$. However, each component has its unique variance (denoted by γ_ℓ and γ_u) and both variances are constant over time.

In comparison to Model 3, the following two-component model considers each component to have different rates of technological change. Note, this is the approach proposed by

Tolhurst and Ker (2015), that is:

$$y_t \sim \lambda \mathcal{N}(\alpha_\ell + \beta_\ell t, \gamma_\ell) + (1 - \lambda) \mathcal{N}(\alpha_u + \beta_u t, \gamma_u) \quad (\text{Model 4})$$

for $j \in \{\ell, u\}$, the mean of each component μ_j is estimated by $\alpha_j + \beta_j t$, which represents the rate of technological change for the respective component. The variance of each component γ_j is constant over time.

Model 5 builds on Model 4 by incorporating a linear trend into the estimation of variance for both components, that is:

$$y_t \sim \lambda \mathcal{N}(\alpha_\ell + \beta_\ell t, \gamma_\ell + \delta_\ell t) + (1 - \lambda) \mathcal{N}(\alpha_u + \beta_u t, \gamma_u + \delta_u t) \quad (\text{Model 5})$$

for $j \in \{\ell, u\}$, the component's mean μ_j is estimated by $\alpha_j + \beta_j t$, whereas the component's variance σ_j^2 is estimated by $\gamma_j + \delta_j t$.

To this end, the one-component models (Models 1 and 2) were estimated by ordinary least squares regression, whereas the unknown parameters in the two-component models (Models 3, 4 and 5) were estimated using the expectation-maximization (EM) algorithm, under a maximum likelihood framework (Bilmes et al., 1998; Chen and Li, 2009; McLachlan, 2018). In the case of two components, the log-likelihood function is given by:

$$\ln \mathcal{L}(\lambda, \mu, \sigma \mid y_t) = \sum_{t=1}^T \left(\ln \left(\lambda \mathcal{N}(\mu_\ell, \sigma_\ell^2 \mid y_t) + (1 - \lambda) \mathcal{N}(\mu_u, \sigma_u^2 \mid y_t) \right) \right) \quad (4)$$

The EM algorithm involves two steps: E-step and M-step. To begin with, each data point (y_t) is arbitrarily assigned to an initial probability that it belongs to the lower component (denoted by $w_{\ell,t}$), which then forms an initial set of parameters through weighted least squares. The E-step computes $w_{\ell,t}$ for each y_t using the current value of the parameters, whereas the M-step computes a new set of parameters that improve the log-likelihood. The algorithm iterates between E-step and M-step until convergence such that the log-likelihood

function is maximized. However, one of the shortfalls in the EM algorithm is that the convergence may happen at local maxima rather than the global maximum especially when the log-likelihood function has multiple local peaks and is relatively flat (Karlis and Xekalaki, 2003; Tolhurst and Ker, 2015; McLachlan, 2018). As suggested by several studies, multiple starting values can be used to assign an initial probability to $w_{\ell,t}$ in order to avoid the convergence at local maxima (Finch, Mendell, and Thode Jr, 1989; Atwood et al., 1992; Karlis and Xekalaki, 2003). To do so, five different starting values were used in the estimation.⁶ Among the five sets of estimated parameters, the one that gave the highest log-likelihood was chosen to be the final estimates in the EM algorithm.

Even though the convergence is less likely to happen at local maxima by using multiple starting values, another concern of the EM algorithm is that the component probabilities λ_k have a bias towards $1/K$ in small samples. For instance, λ in a two-component mixture model ($K = 2$) is biased towards 0.5. This suggests that the data generating process of average crop yield has half a chance to be drawn from a relatively bad year. To mitigate the bias, a penalty function $P(\lambda - 0)^2$ is proposed to add to the log-likelihood function in equation (4) such that the penalized log-likelihood function exhibits a penalty when λ is moving further away from 0. The new penalized log-likelihood function is given by:

$$\ln \mathcal{L}(\lambda, \mu, \sigma \mid y_t) = \sum_{t=1}^T \left(\ln \left(\lambda \mathcal{N}(\mu_\ell, \sigma_\ell^2 \mid y_t) + (1 - \lambda) \mathcal{N}(\mu_u, \sigma_u^2 \mid y_t) \right) \right) - P(\lambda - 0)^2 \quad (5)$$

The objective of the penalized log-likelihood function is to investigate if there is any set of estimated parameters that can give a higher log-likelihood by searching through a lower value of $\hat{\lambda}$. To do so, parameters that were derived from the EM algorithm were re-estimated by maximizing the penalized log-likelihood function rather than the log-likelihood function, under different penalty values (P) ranging from 1 to 50. Since a higher value of P will subject to a heavier penalty, this would lead to a lower estimate of $\hat{\lambda}$ (it also alters

⁶Yields y_t were regressed at 20%, 35%, 50%, 65% and 80% on time t using conditional quantile regression. For each quantile regression, residuals $\hat{e}_t < 0$ were assigned to the lower component with a probability of 1, otherwise 0.

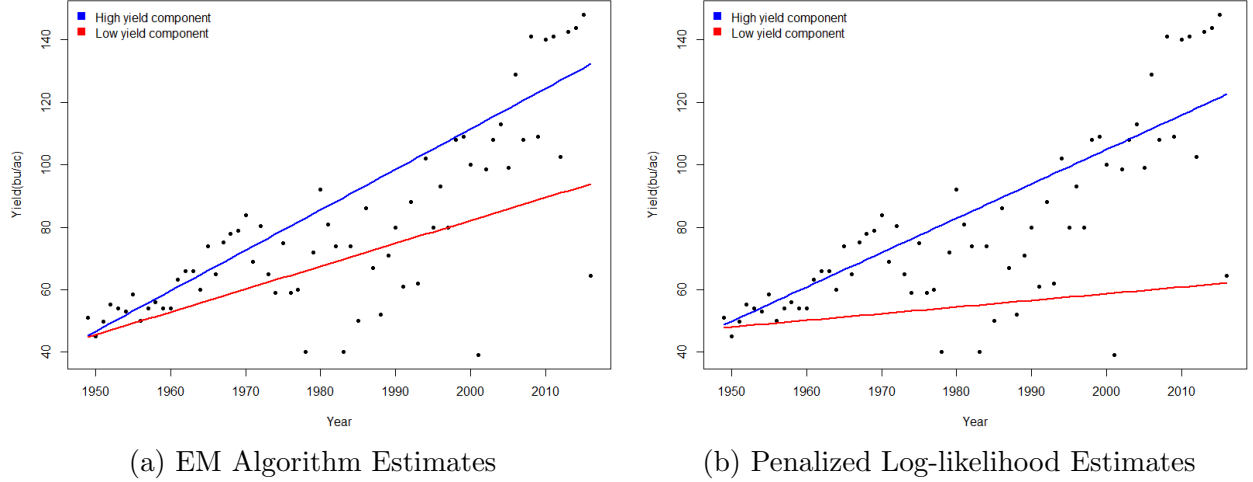


Figure 5: EM Algorithm and Penalized Log-likelihood Estimates, Prince Edward County, Ontario, Corn Yields, 1949 - 2016.

other parameter estimates). Among the different penalty values, the set of parameters that gave the highest log-likelihood in equation (4) was chosen to be the final estimates in the penalized log-likelihood estimation. Eventually, the set of parameters that gave a higher log-likelihood between the penalized log-likelihood estimation and the EM algorithm was chosen to be the final estimates for the corresponding county-crop pair. To demonstrate, Figure 5(a) and 5(b) depict the estimated technological trends from the EM algorithm and the penalized log-likelihood, respectively, for corn yields from Prince Edward County, Ontario. These two technological trends are essentially the temporal means for the two components (for $j \in \{\ell, u\}$, $\alpha_j + \beta_j t$). Figure 5(b) clearly shows that the probability of the low yield component ($\hat{\lambda}$) is less than the probability in Figure 5(a) as the yields have less chance to be drawn from the lower component.

After obtaining the final estimated parameters for each county-crop pair, one of the limitations is that the two technological trends from the set of parameters that maximize the log-likelihood may or may not intersect if no restrictions are imposed. This limitation only happens in Models 4 and 5 as they are the only models that have two unique technological trends. To provide some insight, Figure 6 presents the two technological trends for corn yields from Brant County, Ontario. Figure 6(a) clearly illustrates that the unrestricted two

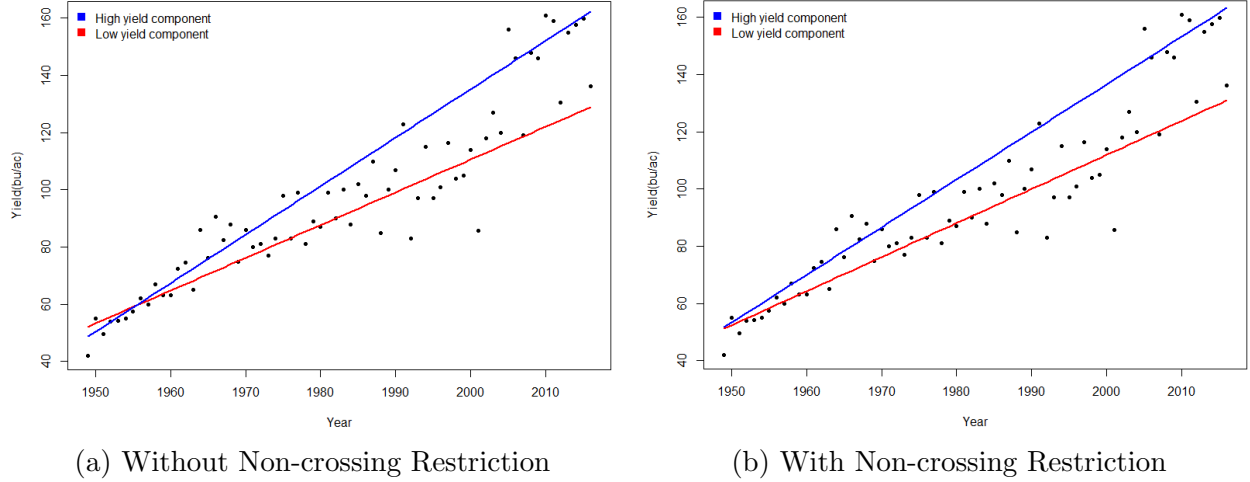


Figure 6: Crossing and Non-crossing Technological Trends, Brant County, Ontario, Corn Yields, 1949 - 2016.

trends are crossing in the year 1960. Since the upper component intuitively should always be higher than the lower component, as illustrated in Figure 6(b), a restriction was imposed to avoid crossing (hereafter, non-crossing restriction).⁷

Other than the non-crossing restriction, another limitation appears when the component variances turn negative if they are estimated by a time trend (for $j \in \{\ell, u\}$, $\gamma_j + \delta_j t$). Recall that one of the thesis objectives is to estimate the crop insurance payouts in the next five and ten years (i.e. year 2023 and 2028). Therefore, a restriction was imposed on the time-varying component variances such that they were prohibited from turning negative before the 2028 (hereafter, non-negative restriction).⁸ Note, this restriction only applies to Model 5 as it is the only model that has estimated time-varying variances turning negative before 2028.

If the unrestricted two trends are initially not crossing, then the non-crossing restriction does not bind. Similarly, if the unrestricted time-varying variances do not turn negative before the year 2028, the non-negative restriction also does not bind. Therefore, in both cases, the estimated parameters are identical with and without restrictions. In the analysis,

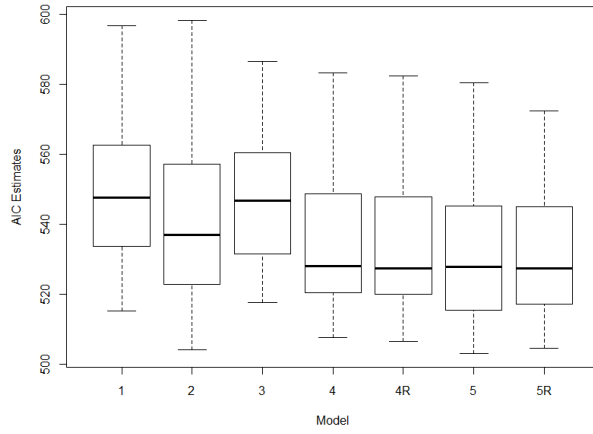
⁷Restriction was set up in two ways: (1) the two trends were restricted to start from the same intercept, and (2) the non-crossing two trends that gave the highest log-likelihood in the convergence process were picked. Estimated parameters that obtained a higher log-likelihood were chosen between the two.

⁸For instance, the year 2028 is the 80th year from 1949 and, therefore, $t = 80$ such that $\hat{\gamma} + 80 \hat{\delta}$ has to be greater than 0.

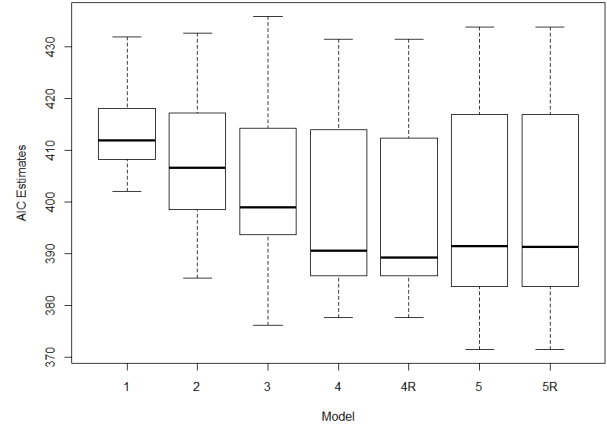
Models 4 and 5 were restricted in all cases. The restricted Models 4 and 5 were labelled as Models 4R and 5R, respectively.

For ease of interpretation, each county-crop pair’s AIC estimate for each model was aggregated in box plots. Figure 7 and Figure 8 illustrate the box plots by crop for Ontario and Saskatchewan, respectively. Table 3 presents the median of each box plot. Consider the medians of the box plots: 3 out of the 7 province-crop pairs obtain the lowest median in Model 4, 2 out of the 7 in Model 4R, 1 out of the 7 in Model 5 and 1 out of the 7 has the same median between Model 4 and 4R. This confirms that using two-component models to model crop yields from the data set achieves better efficiency than using one-component models, which suggests that the consideration of the higher moments better explains the data. This is also consistent with other literature that estimating yield distribution with a single normal distribution (one component) may not be appropriate, even with the adjustment of heteroscedasticity (i.e. Gallagher (1987); Nelson and Preckel (1989); Goodwin and Ker (1998); Atwood, Shaik, and Watts (2003); Ramirez, Misra, and Field (2003); Tolhurst and Ker (2015)). Interestingly, Ontario corn and soybean have notable decreases in their medians using two-component models, but less so for other province-crop pairs. It is also worth noting that Ontario corn and soybean achieve better efficiency by using the restricted model (Model 4R) rather than the unrestricted model (Model 4). This happens when the initial two technological trends cross each other such that they are restricted to start from the same intercept (non-crossing restriction). Consequently, instead of estimating α_ℓ and α_u , only one intercept term needs to be estimated. As a result, the AIC estimate is lower if the reduced log-likelihood is offset by the decreased use of parameters.

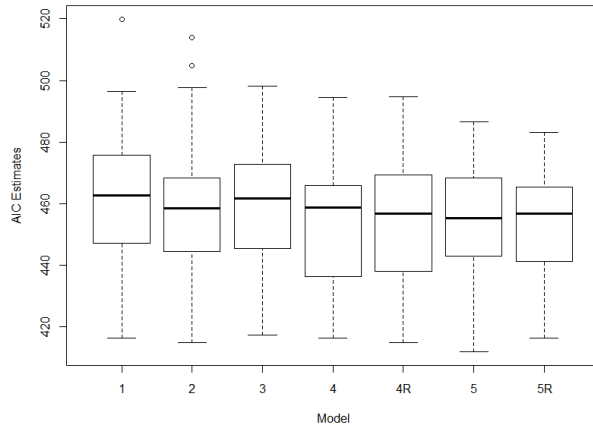
To select the appropriate model for each province-crop pair, Table 4 reports the number of counties that obtained the lowest AIC estimate under each model. Among all of the seven province-crop pairs, Ontario corn is the only pair that dominates in Model 5R: 62.5% of counties achieve the highest efficiency using Model 5R. The remaining six pairs dominate in Model 4R: 66.7% for Ontario soybean, 42.3% for Ontario wheat, 49% for Saskatchewan



(a) Ontario Corn



(b) Ontario Soybean



(c) Ontario Wheat

Figure 7: AIC Estimates of Each Model in Box Plots, by Crop in Ontario

Table 3: Median of AIC Estimates of Each Model, by Crop and Province

Province	Observations			Model					
Crop	Counties	Years	1	2	3	4	4R	5	5R
Ontario									
Corn	32	68	547.54	536.92	546.79	528.09	527.42	527.93	527.48
Soybean	6	68	411.86	406.65	398.99	390.61	389.29	391.40	391.25
Wheat	26	68	462.74	458.45	461.55	458.65	456.76	455.25	456.76
Saskatchewan									
Barley	204	79	597.94	598.76	599.20	592.91	594.64	596.91	597.29
Canola	144	47	303.66	302.11	305.53	297.60	301.62	302.54	304.74
Oats	131	79	641.26	642.04	643.65	636.73	637.55	640.86	642.45
Wheat	267	79	530.10	531.27	531.69	528.38	528.38	530.75	531.48

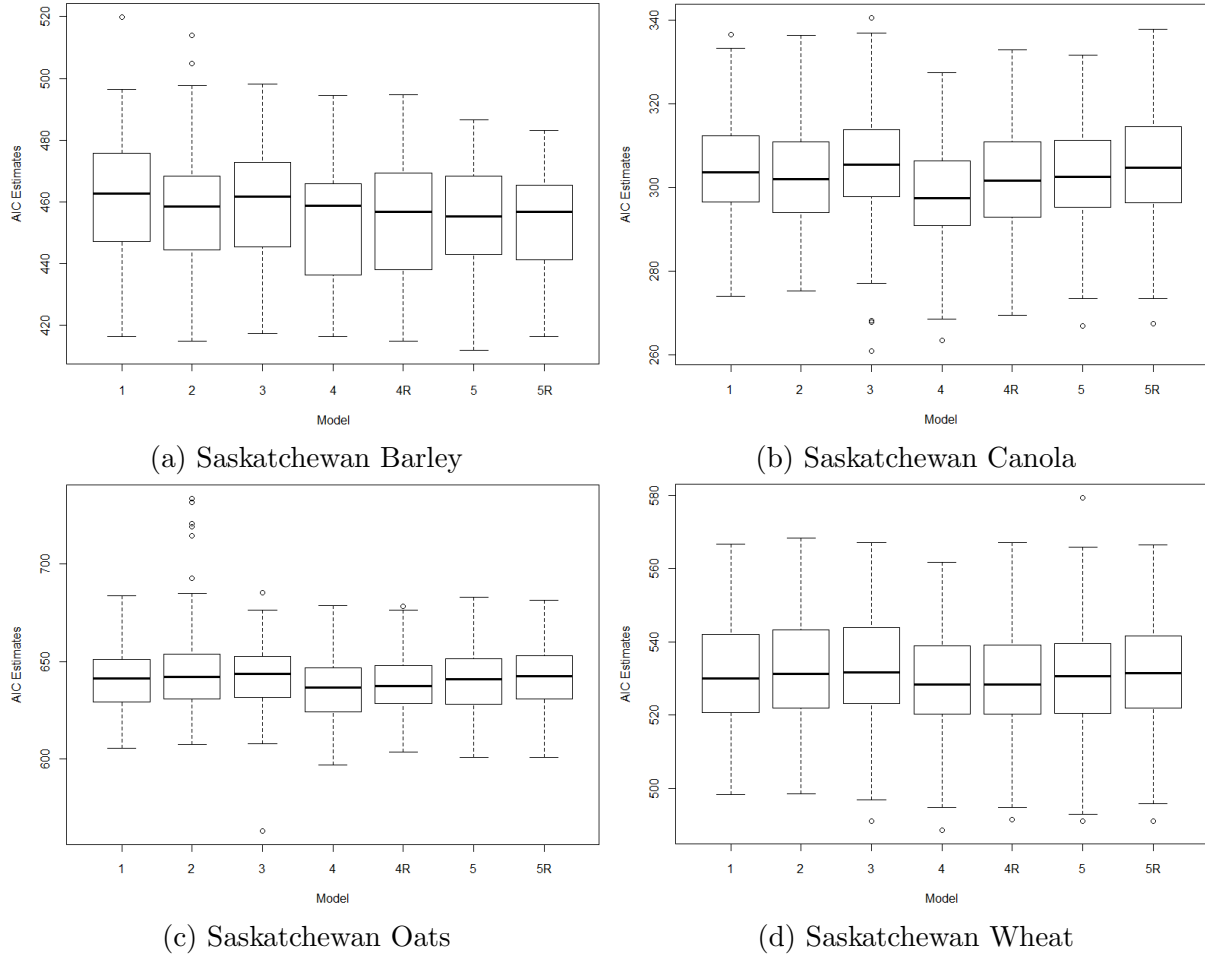


Figure 8: AIC Estimates of Each Model in Box Plots, by Crop in Saskatchewan

Table 4: Number of Counties that Obtained the Lowest AIC Estimate, by Crop and Province

Province	Observations		Model						
	Counties	Years	1	2	3	4	4R	5	5R
Ontario									
Corn	32	68	0	5	0	NA	7	NA	20
Soybean	6	68	0	0	0	NA	4	NA	2
Wheat	26	68	0	6	0	NA	11	NA	9
Saskatchewan									
Barley	204	79	55	17	9	NA	100	NA	23
Canola	144	47	33	35	9	NA	46	NA	21
Oats	131	79	27	11	7	NA	72	NA	14
Wheat	267	79	84	14	26	NA	101	NA	42

Notes : Model 4 and 5 were assigned NAs since models were restricted in all cases.

barley, 31.9% in Saskatchewan canola, 55% in Saskatchewan oats and lastly, 37.8% for Saskatchewan wheat. For Saskatchewan canola, the numbers of counties that give the lowest AIC are not far behind when compare Model 4R (46 counties) to Models 1 and 2 (33 and 35 counties, respectively), which suggests that some of its counties can be sufficiently explained by using one-component models. This is not surprising with 47 years of data (relatively less than other province-crop pairs); the model theoretically should require fewer parameters to fit the data. The overall results in Table 4 suggest that the most efficient model for each county varies within a province-crop pair. Nonetheless, the model with the highest number of counties was employed to model yields for the corresponding province-crop pair. That is, all counties within the province-crop pair were estimated under the same model. As a result, all province-crop pairs, with the exception of Ontario corn, were estimated using Model 4R. Ontario corn is the only province-crop pair that was estimated using Model 5R.

With the exception of Saskatchewan canola, the AIC estimates between the unrestricted and restricted models are quite similar as evidenced by the box plots. As a robustness check, a likelihood ratio test was performed at the five percent significance level to investigate whether the restricted models have statistically significant differences from the unrestricted models. Table 5 reports the number of counties that require the restrictions, and the number of counties that reject the null hypothesis of no difference (between the restricted and the unrestricted model) for each province-crop pair.

Note, a multiple testing problem arises since more than one hypothesis test is performed simultaneously such that the probability of type I error is no longer at five percent.⁹ To mitigate the issue, Table 5 also reports the rejection counts under the Holm-Bonferroni (family-wise) p -value (Holm, 1979). Note, the Holm-Bonferroni adjustment is widely believed to be overly conservative because all the tests are assumed to be independent, however, counties are spatially correlated. The adjustment also has a very low power when the number

⁹For example, given that Ontario corn has 32 counties, the probability of type I error (family-wise error rate) is $1 - 0.95^{32} = 0.806$ if the 32 tests are independent, meaning that the chance of incorrectly rejecting the null hypothesis at least once among the 32 hypotheses is 80.6% without the Holm-Bonferroni adjustment.

Table 5: Restriction and Rejection of Restricted Model, by Crop and Province

Province	Observations		Non-crossing			Non-negative			Both		
Crop	Counties	Years	R	P	F	R	P	F	R	P	F
Ontario											
Corn	32	68	11	4	0	3	0	0	4	1	1
Soybean	6	68	3	0	0	NA	NA	NA	NA	NA	NA
Wheat	26	68	20	5	1	NA	NA	NA	NA	NA	NA
Saskatchewan											
Barley	204	79	57	9	0	NA	NA	NA	NA	NA	NA
Canola	144	47	74	10	2	NA	NA	NA	NA	NA	NA
Oats	131	79	56	32	8	NA	NA	NA	NA	NA	NA
Wheat	267	79	98	44	3	NA	NA	NA	NA	NA	NA

Notes : Statistical significance evaluated at the 5% using a likelihood ratio test. The non-crossing restriction avoids the two technological trends crossing in any given year. The non-negative restriction avoids components variances turning negative before the year 2028. R represents the number of counties that require the above-mentioned restriction(s). P represents the number of counties that reject the restriction(s) using standard p -values. F represents the number of counties that reject the restriction(s) using family-wise p -values.

of simultaneously tested hypotheses are large (especially for Saskatchewan counties in this thesis) (Chen, Feng, and Yi, 2017). Nonetheless, the Holm-Bonferroni adjustment serves to provide a lower bound of these rejections. To interpret the results, consider Ontario corn as an example: 11 out of the 32 counties require non-crossing restriction and 4 of these restricted counties reject the restriction at the standard five percent significance level; however, none of these counties are able to reject the restriction under the family-wise p -value. For the non-negative restriction, 3 out of the 32 counties require the restriction but none of the three reject the restriction. Next, 4 out of the 32 counties require both the non-crossing and non-negative restrictions and 1 of the 4 rejects both restrictions under the standard and family-wise p -value. Recall that the non-negative restriction only applies to Ontario corn due to the time-varying component variances and, therefore, it is not applicable to other province-crop pairs. In general, the percentage of counties that require non-crossing restriction are between 25% to 50% for each province-crop pair, with the exception of Ontario wheat (76.9%). Not surprisingly, the rejection rates of the restricted models are fairly low

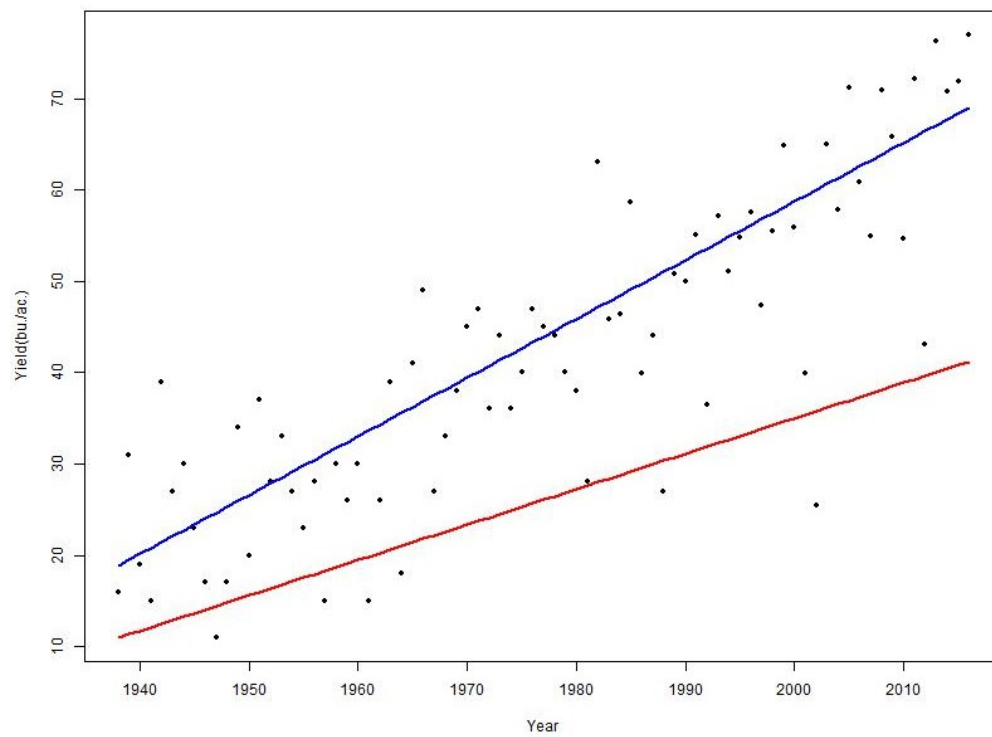
for all province-crop pairs under standard p -value (ranging from 0% to 24.4%), and much lower under family-wise p -value (ranging from 0% to 6.1%). The overall results suggest that the imposed restrictions, in most cases, do not statistically significantly affect the estimation efficiency.

4 Empirical Results and Implications

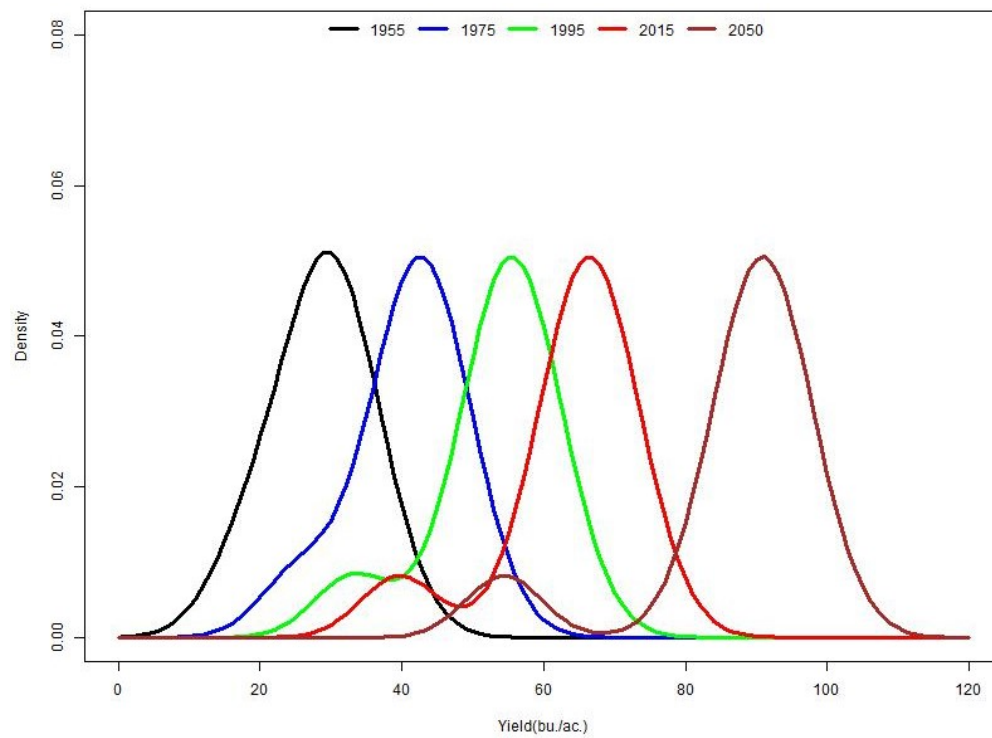
4.1 Estimation Results

Figure 9 illustrates the representative two-component technological trend and yield distribution estimates in pairs for each of the six crop. Recall that Ontario corn counties were estimated in the two-component model with time-varying variances, whereas for other province-crop pairs the component variances were assumed to be constant over time. Therefore, the interpretation of the estimated variance parameters is slightly different between the two models. To demonstrate, Figure 9(e) and 9(k) present the two-component technological trend estimates for Peel, Ontario corn and Abernethy, Saskatchewan wheat. First, for Peel, Ontario corn, the lower rate of technological change is estimated at $\hat{\alpha}_\ell + \hat{\beta}_\ell t$ for a given year t , where $\hat{\alpha}_\ell = 48.18$ and $\hat{\beta}_\ell = 1.47$. The upper technological trend is estimated at $\hat{\alpha}_u + \hat{\beta}_u t$ for a given year t , where $\hat{\alpha}_u = 49.17$ and $\hat{\beta}_u = 1.81$. The slope coefficients of technological change $\hat{\beta}_\ell$ and $\hat{\beta}_u$ represent the annual increase in average crop yield (measured in bushels per acre) for the lower component and upper component, respectively. In this case, $\hat{\beta}_u > \hat{\beta}_\ell$ suggests that the rate of technological change in relatively good years is higher than in relatively bad years, which also indicates a higher year-to-year dispersion of average crop yield over time. As for the component variances, the variance of the lower component is estimated at $\hat{\gamma}_\ell + \hat{\delta}_\ell t$ for a given year t , where $\hat{\gamma}_\ell = 2.31$ and $\hat{\delta}_\ell = 2.9$. The variance of the upper component is estimated at $\hat{\gamma}_u + \hat{\delta}_u t$ for a given year t , where $\hat{\gamma}_u = 31.37$ and $\hat{\delta}_u = 0.96$. The slope coefficients of variance $\hat{\delta}_\ell$ and $\hat{\delta}_u$ represent the annual increase in the variances of average crop yield for the lower component and upper component, respectively. In this case, $\hat{\delta}_\ell > \hat{\delta}_u$ suggests that the yield dispersion in both components are increasing asymmetrically over time, but higher in relatively bad years than in relatively good years, which also indicates a higher downside yield risk over time.

Interpreting the parameter estimates of the estimated yield distribution, which is illustrated at different time frames in Figure 9(f), is rather straightforward. For a given year t , the data generating process of average crop yield has a probability of $\hat{\lambda} = 0.41$ to fall

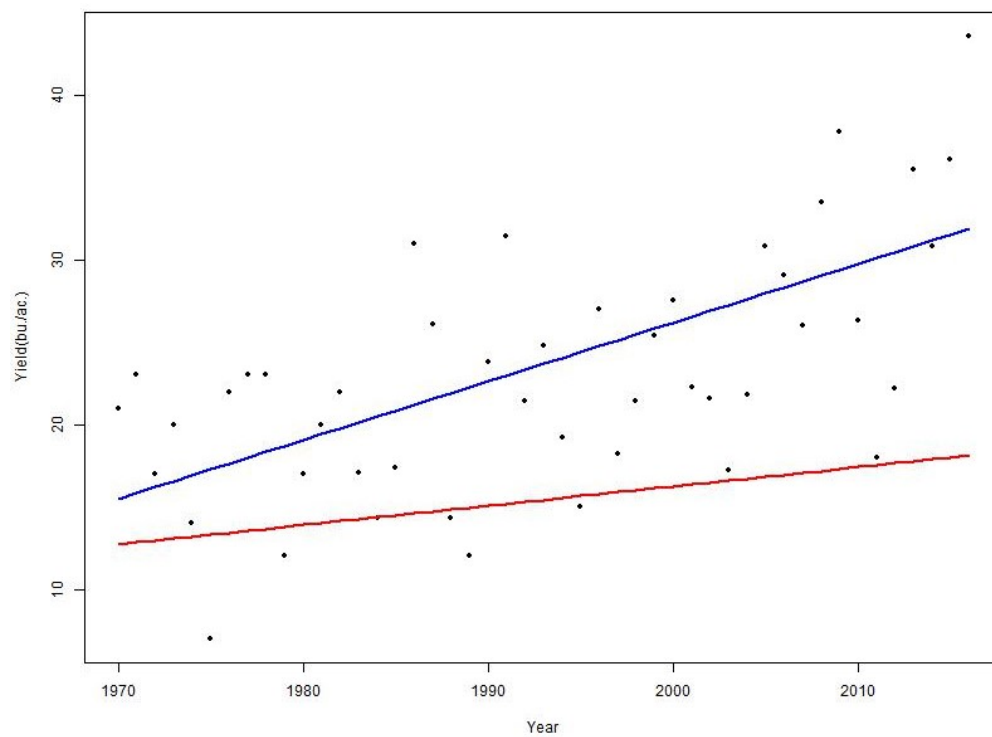


(a)

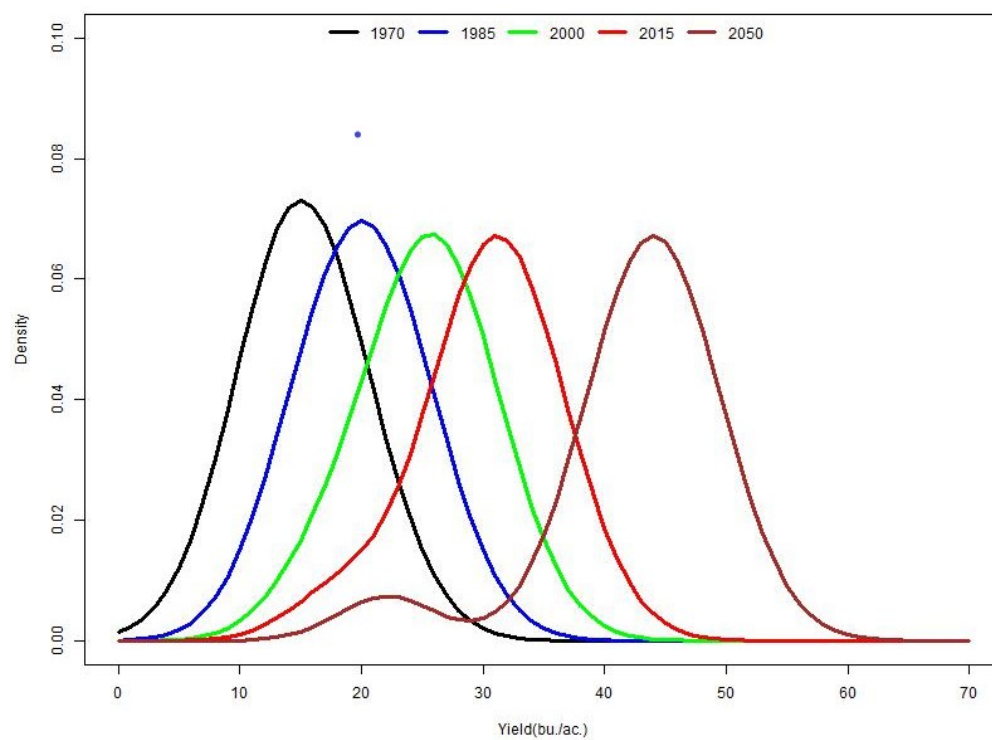


(b)

Figure 9: Hoodoo, Saskatchewan Barley

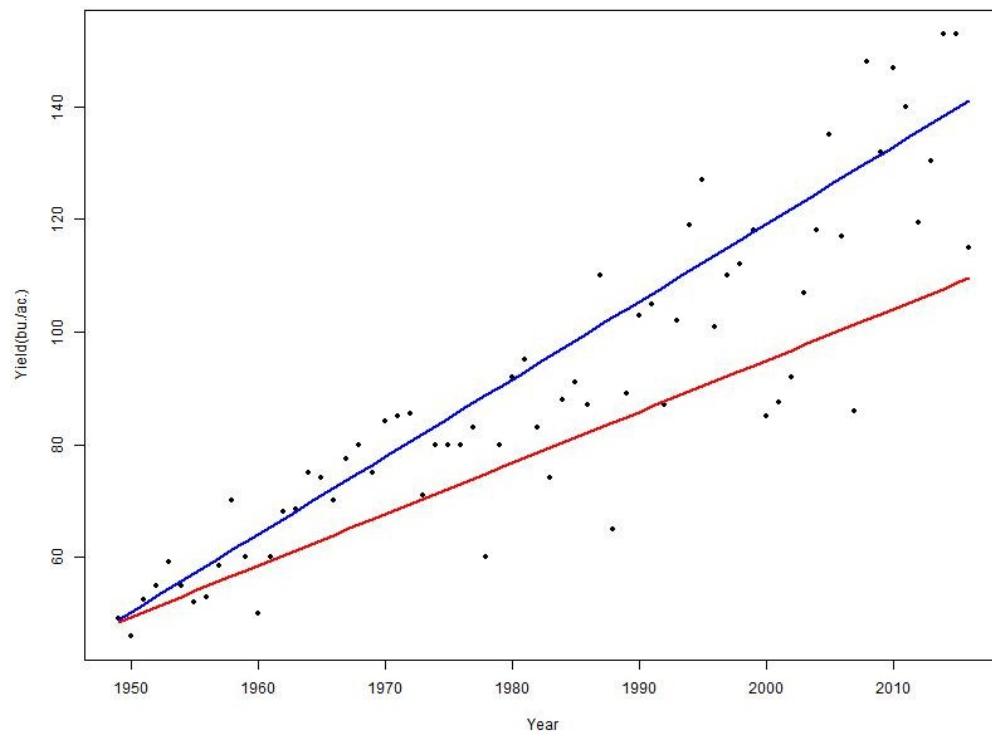


(c)

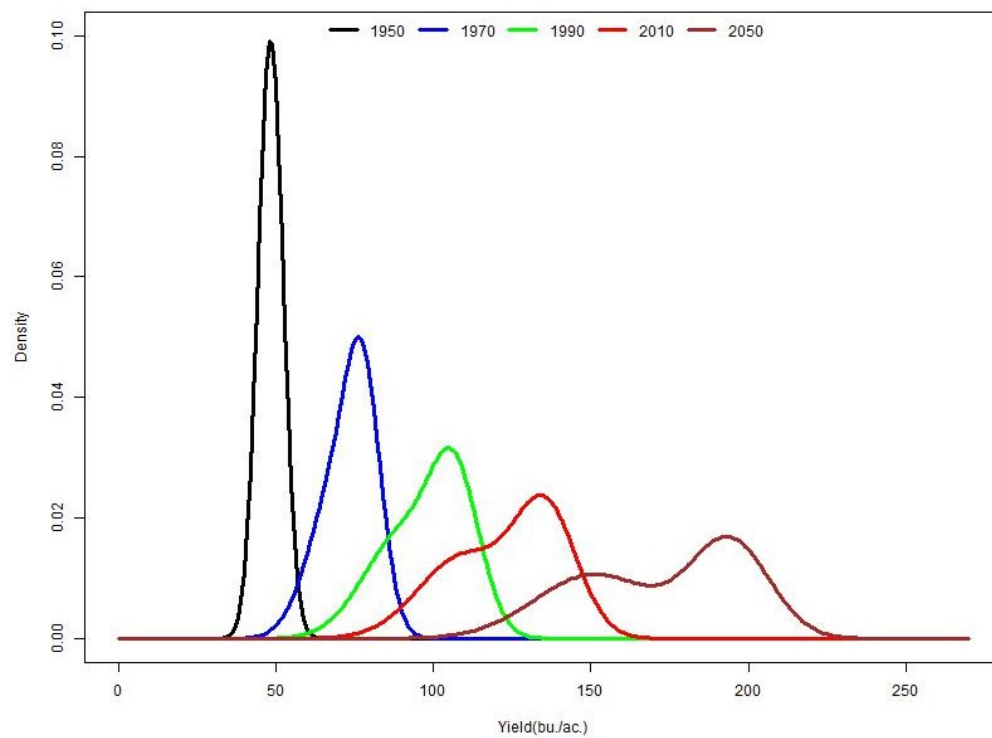


(d)

Figure 9: Wolseley, Saskatchewan Canola

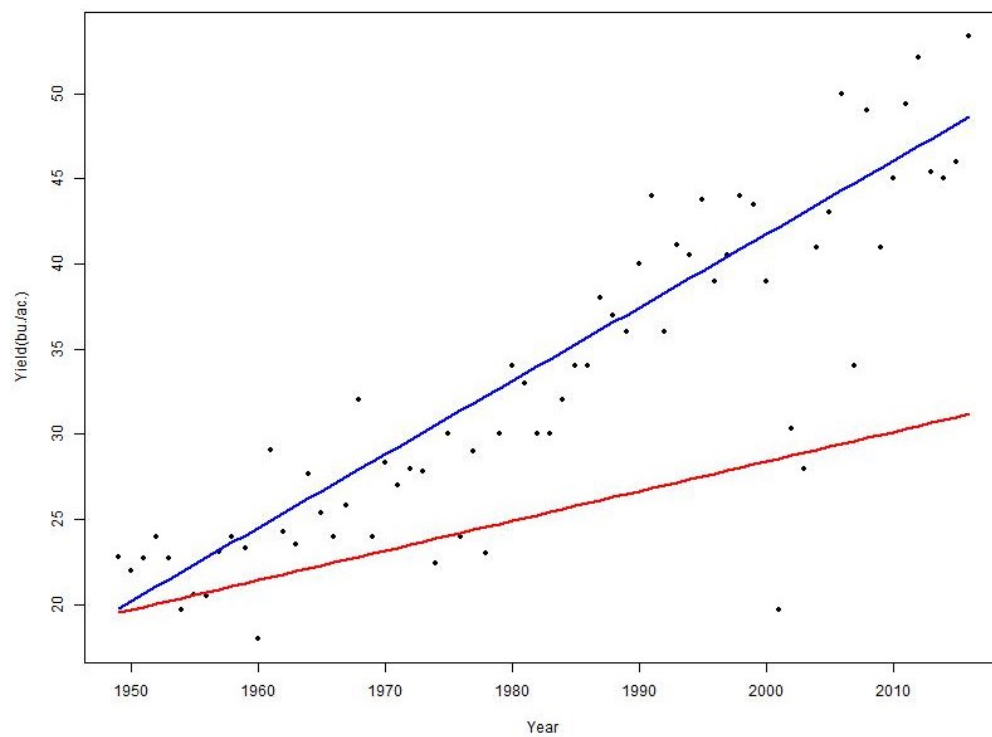


(e)

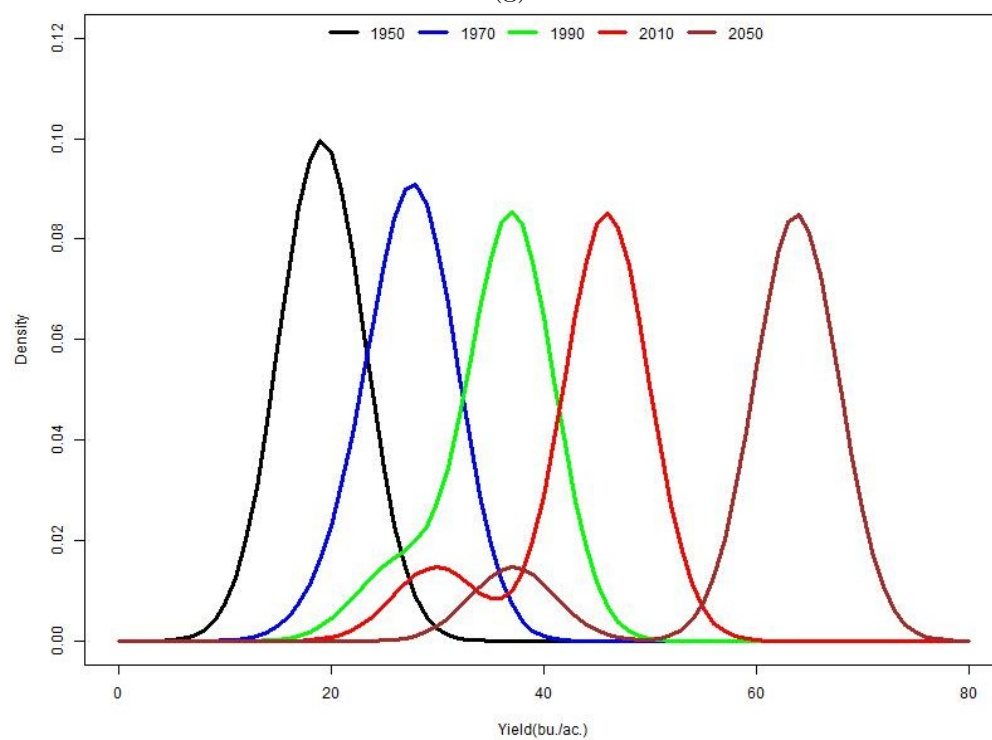


(f)

Figure 9: Peel, Ontario Corn

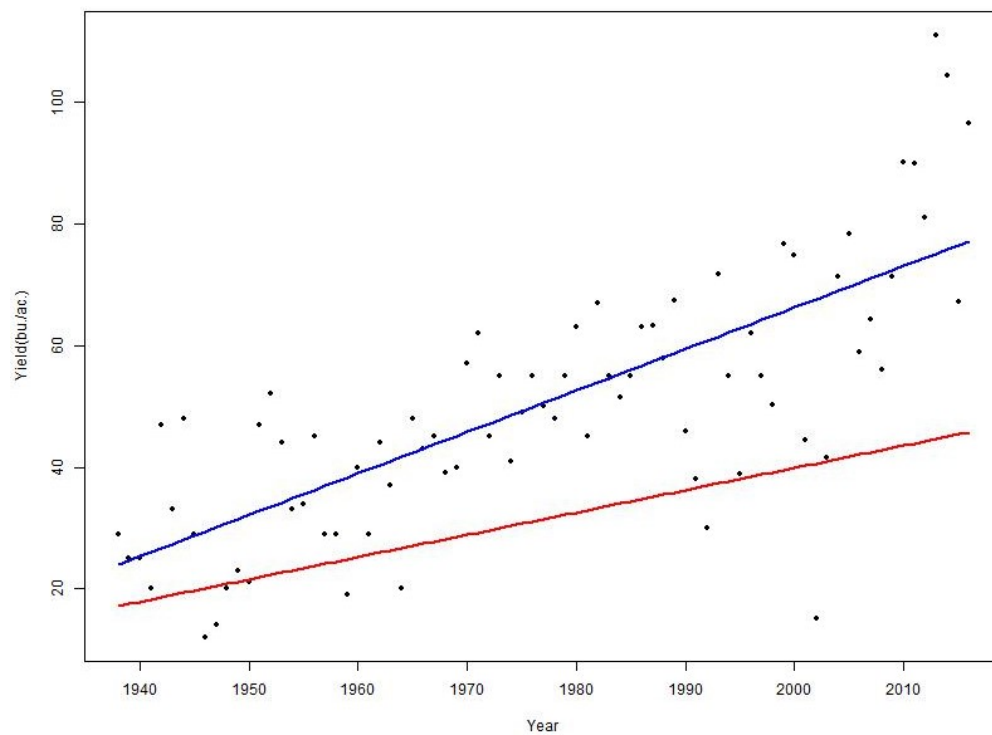


(g)

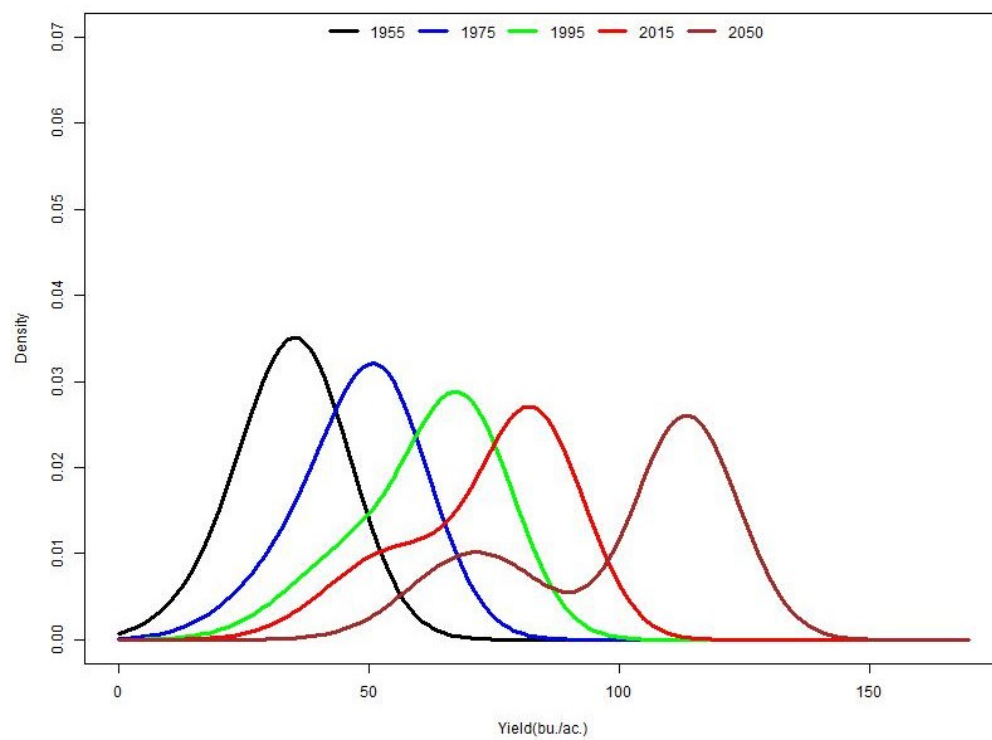


(h)

Figure 9: Middlesex, Ontario Soybean

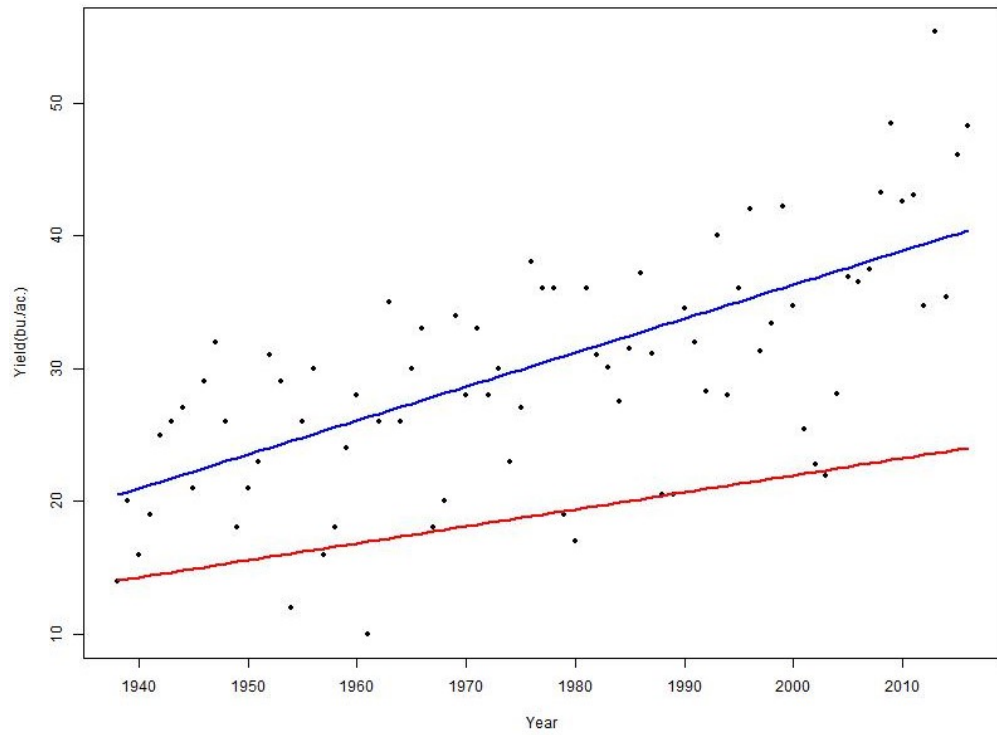


(i)

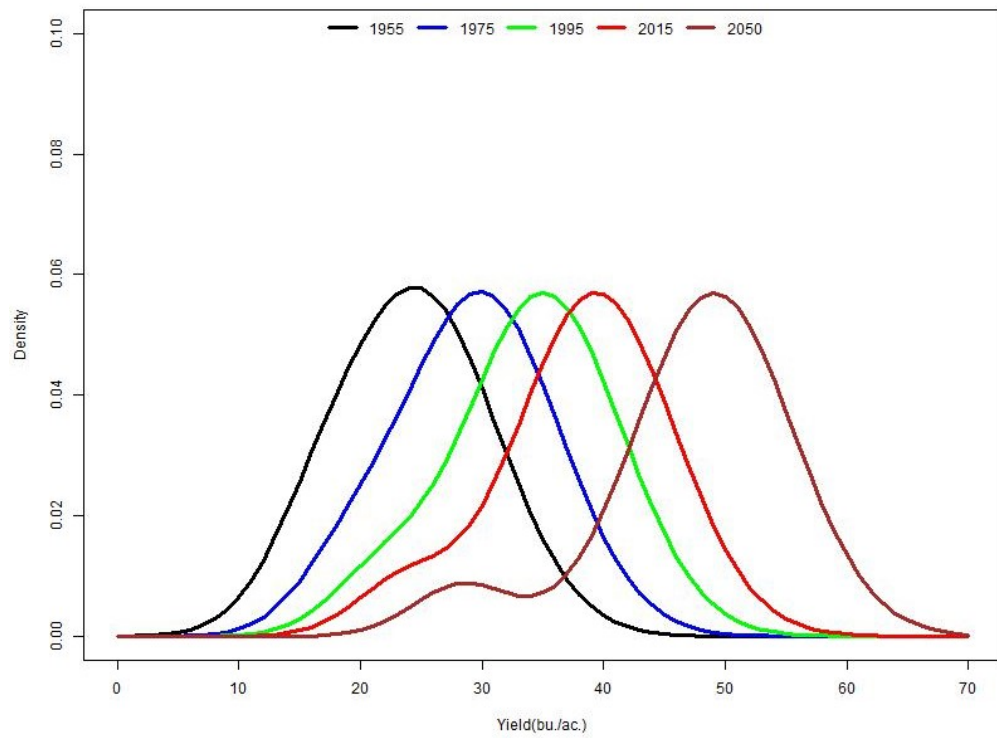


(j)

Figure 9: Frenchman Butte, Saskatchewan Oats



(k)



(l)

Figure 9: Abernethy, Saskatchewan Wheat

within the lower component $\mathcal{N}(48.18 + 1.47t, 2.31 + 2.9t)$ or a probability of $1 - \hat{\lambda} = 0.59$ to fall within the upper component $\mathcal{N}(49.17 + 1.81t, 31.37 + 0.96t)$. Estimating the yield distribution for an exact year can be done by substituting the number of the year into t . For instance, the year 2015 is the 67th year from the starting period year 1949 and, therefore, $t = 67$. As a result, the estimated yield distribution in year 2015 is a mixture of 41% at the lower component $\mathcal{N}(146.67, 196.61)$ and 59% at the upper component $\mathcal{N}(170.44, 95.69)$. If the current trends of technological change and component variances remain unchanged, then the hypothetical yield distribution in year 2050 will be a mixture of 41% at the lower component $\mathcal{N}(198.12, 298.11)$ and 59% at the upper component $\mathcal{N}(233.79, 129.29)$.

Moving forward to Abernethy, Saskatchewan wheat, the interpretation of the variance estimates is slightly different (interpretation of other parameters remains unchanged). Unlike Peel, Ontario corn, the variance of the lower and upper components are represented by $\hat{\gamma}_\ell$ and $\hat{\gamma}_u$, respectively, where $\hat{\gamma}_\ell = 0.41$ and $\hat{\gamma}_u = 41.04$ such that they were both assumed to be constant over time. Interestingly, $\hat{\gamma}_\ell$ is much lower than $\hat{\gamma}_u$, which suggests that the yield dispersion in relatively bad years is much lower than in relatively good years. In Figure 9(k), the lower technological trend is estimated at $13.9 + 0.13t$ and the upper technological trend is estimated at $20.21 + 0.26t$. The average crop yield has a probability of $\hat{\lambda} = 0.09$ to fall within the lower component $\mathcal{N}(13.9 + 0.13t, 0.41)$ and a probability of $1 - \hat{\lambda} = 0.91$ to fall within the upper component $\mathcal{N}(20.21 + 0.26t, 41.04)$. Similar to the previous example, $\hat{\beta}_u > \hat{\beta}_\ell$ suggests that the rate of technological change outperforms in relatively good years; however, in contrast to the last example, this county has a much lower $\hat{\lambda}$. It is not surprising that a lower $\hat{\lambda}$ would be accompanied by a lower $\hat{\gamma}_\ell$, because, if only a few observations have a particularly high weight in the lower component, the variation of these observations also tends to be low. Table 6 summarizes the median of parameter estimates for all the 810 county-crop pairs (see Appendix 7.1 for the detailed summary statistics). There are several patterns that are worthwhile to mention. Consider the medians of $\hat{\beta}_\ell$ and $\hat{\beta}_u$: all the province-crop pairs exhibit a higher $\hat{\beta}_u$ than $\hat{\beta}_\ell$, which suggests that the rate of technological change in relatively

Table 6: Median of Estimated Parameters, by Crop and Province

Province	Observations		Parameters								
Crop	Counties	Years	λ	α_ℓ	β_ℓ	γ_ℓ	δ_ℓ	α_u	β_u	γ_u	δ_u
Ontario											
Corn	32	68	0.42	46.54	1.06	4.13	2.01	47.24	1.54	6.46	1.67
Soybean	6	68	0.19	19.00	0.19	17.74	NA	19.70	0.40	8.03	NA
Wheat	26	68	0.19	26.14	0.38	15.83	NA	26.14	0.77	30.10	NA
Saskatchewan											
Barley	204	79	0.17	14.52	0.31	54.24	NA	20.52	0.53	52.59	NA
Canola	144	47	0.47	15.25	0.17	19.38	NA	17.27	0.42	13.64	NA
Oats	131	79	0.33	19.18	0.44	86.58	NA	28.14	0.65	101.29	NA
Wheat	267	79	0.18	8.85	0.19	16.28	NA	16.36	0.28	28.24	NA

good years generally outperforms the rate in relatively bad years. For the magnitude of $\hat{\beta}_\ell$ and $\hat{\beta}_u$, the annual increase in average crop yield is clearly higher in Ontario corn for both components (median of $\hat{\beta}_\ell = 1.06$ bu./ac., $\hat{\beta}_u = 1.54$ bu./ac.), but less so for other province-crop combinations, particularly for Saskatchewan wheat (median of $\hat{\beta}_\ell = 0.19$ bu./ac., $\hat{\beta}_u = 0.28$ bu./ac.). It is not surprising that corn has a higher yield improvement than other crops over time, due to its early widespread commercialization of hybrid varieties since the early 20th century (Crow, 1998; Edgerton, 2009). In contrast, advances in the variety development for other crops have relatively lagged behind. The significant yield improvement for Ontario corn also leads to an increasing volatility over time in both components, especially in the downside yield risk (as $\hat{\delta}_\ell > \hat{\delta}_u$). The year-to-year dispersion of crop yields in relatively bad years has generally increased by 2.01 bu./ac. annually, which is higher than 1.67 bu./ac. in relatively good years. Finally, considering the median of $\hat{\lambda}$ is also interesting because of the heterogeneity among the province-crop pairs: the data generating process of yields for Ontario soybean, Ontario wheat, Saskatchewan barley, and Saskatchewan wheat have less than a 20% probability of being drawn from a relatively bad year. Conversely, the estimates are much higher for Ontario corn and Saskatchewan canola, which have a probability of 42%

and 47%, respectively.

4.2 $H_0^1 : \beta_\ell = \beta_u$

The estimated parameters also allow further investigation of the above-mentioned tendencies by testing a number of hypotheses. Table 7 (column 3 and 4) summarizes the rejection counts for the first hypothesis: $H_0^1 : \beta_\ell = \beta_u$, and reports under the standard p -value as well as the family-wise p -value. Using a likelihood ratio test for each of the 810 county-crop pairs, the first hypothesis is a two-sided test which looks at whether the upper technological trend has a statistically significant difference from the lower technological trend. Interestingly, this null hypothesis has a higher rejection rate in Ontario than Saskatchewan with a standard p -value. Namely, Ontario corn (25 of the 32 counties), Ontario wheat (15 out of 26), and Ontario soybean (3 out of 6). In total, 297 of the 810 county-crop pairs reject the null. With the family-wise p -value, the rejection rate for Saskatchewan crops is very low (less than 10 counties for each of the four crops). Given that Saskatchewan has many more counties than Ontario, as mentioned in the Methods section, it is not surprising that most of the counties are not statistically significant when a large number of hypotheses are tested simultaneously. Nonetheless, Ontario corn and soybean still maintain a relatively high rejection rate (12 out of 32 and 3 out of 6, respectively). In total, 42 of the 810 county-crop pairs reject the null with the family-wise p -value.

Based on the first hypothesis, a one-sided test: $H_0 : \beta_u \leq \beta_\ell$ was also considered. This hypothesis looks at whether the upper technological trend is statistically significantly larger than the lower technological trend. The reason for focusing on this direction is that an increase in yield volatility is, in most cases, caused by the two diverging technological trends. As expected, most of the counties that reject the first null (297 counties in total) are also able to reject this null (260 counties in total), meaning that counties that have statistically significant differences in the rate of technological change between the two components, are very likely to have a higher rate in the upper component than the lower component. This

Table 7: Hypothesis Test One and Two: Rejection Counts

Province	Observations		$H_0^1 : \beta_\ell = \beta_u$		$H_0^2 : \lambda_t = \lambda$		
	Counties	Years	P	F	P	F	J
Ontario							
Corn	32	68	25	12	4	4	0
Soybean	6	68	3	3	3	3	1
Wheat	26	68	15	3	2	2	1
Saskatchewan							
Barley	204	79	89	5	49	5	9
Canola	144	47	32	6	10	0	0
Oats	131	79	45	4	14	4	0
Wheat	267	79	88	9	85	10	13
All	810	NA	297	42	167	28	24

Notes : Statistical significance evaluated at the 5% using a likelihood ratio test for H_0^1 . Statistical significance evaluated at the 5% using a t -test for H_0^2 . P represents the number of counties that reject the respective null hypothesis using standard p -values. F represents the number of counties that reject the respective null hypothesis using family-wise p -values. J represents the number of counties that reject H_0^2 using jackknife p -values.

Table 8: Summary Statistics for Ratio of $\hat{\beta}_u/\hat{\beta}_\ell$

Province	Observations		Summary Statistics				
	Counties	Years	Minimum	Mean	Median	Maximum	Std. Dev.
Ontario							
Corn	32	68	0.90	1.77	1.43	5.54	1.05
Soybean	6	68	1.01	2.23	2.10	4.28	1.13
Wheat	26	68	1.36	2.02	1.89	3.50	0.54
Saskatchewan							
Barley	204	79	0.13	4.89	1.90	241.60	18.55
Canola	144	47	-11.78	53.03	2.71	5459.43	457.62
Oats	131	79	0.09	3.63	1.62	57.77	7.81
Wheat	267	79	0.19	95.53	1.86	20690.64	1269.79

Notes : The positive extreme values (for example a maximum ratio of 20690.64 for Saskatchewan wheat) and negative extreme values (for example a minimum ratio of -11.78 for Saskatchewan canola) appear because $\hat{\beta}_\ell \rightarrow 0$, which inflates the ratio. $\hat{\beta}_\ell \rightarrow 0$ from the positive side generates positive extreme values whereas $\hat{\beta}_\ell \rightarrow 0$ from the negative side generates negative extreme values. Note, $\hat{\beta}_u > 0$ for all county-crop pairs.

result is also in line with Table 8, which shows that counties generally have diverging technological trends as suggested by the medians of $\hat{\beta}_u/\hat{\beta}_\ell$ ratio. Conversely, for Saskatchewan wheat, only 64 out of the 88 rejections in the first null are able to reject the second null. This indicates that 24 counties do have statistically significantly converging technological trends over time, which is not consistent with the overall results. In total, 45 of the 810 county-crop pairs reject the second null with the family-wise p -value. For Ontario soybean, Ontario wheat and Saskatchewan canola, all counties that reject the first null are also able to reject the second null under both p -values.

Contributing factors to asymmetrical rate of technological change between components are countless. The diverging two technological trends over time can be explained by, but not limited to, the technological advancement in seed genomics. An innovative seed variety may improve crop productivity in ideal growing conditions. However, it still requires a certain level of growing condition to develop and, therefore, has little or no effect in bad growing conditions. Moreover, technological change has increased the number of plants per acre. However, if the demand for water remains fixed over time, the minimum amount of precipitation would theoretically also increase over time. This is not conjectural as literature has suggested that crop productivity is positively correlated to drought sensitivity (DeLucia et al., 2014; Ort and Long, 2014), which consequently results in a higher downside yield risk. Another explanation could be that the Canadian subsidized crop insurance program (60% subsidized by government) has encouraged producers to adopt high-risk and high-return technologies over time, given that these producers have been shielded from the downside yield risk (Ker et al., 2017).

4.3 $H_0^2 : \lambda_t = \lambda$

The next hypothesis test: $H_0^2 : \lambda_t = \lambda$ investigates whether technological change has statistically significantly changed the probability of crop yields drawing from relatively bad

years over time.¹⁰ This hypothesis is of particular interest because technological change may decrease or increase the probability of the lower component: probability decreases when crops have better productivity and tolerance to various stresses, and probability increases when producers adopt high-risk and high-return seed varieties under the protection of the subsidized crop insurance. In the meantime, other factors, such as the outcome of weather, may also increase or decrease the chance of low yield depending on the severeness of the weather and geographical location. Using a t -test for each of the 810 county-crop pairs, Table 7 (column 5-7) reports the rejection counts under the standard p -value, family-wise p -value, and jackknife p -value. Note, the jackknife p -value was not previously employed in this thesis as the jackknife standard error is not applicable to the likelihood ratio test (which were used in all the previous tests). The conventional standard error is downward biased due to three factors: (1) it does not account for the estimated regressor; (2) it does not account for the multiple testing problem; and (3) it does not account for the spatial correlation of the 810 counties. Given this downward bias, for comparison the jackknife standard error is also reported. The jackknife estimate of standard error, following Efron and Tibshirani (1994), is given by:

$$\sqrt{\frac{T-1}{T} \sum_{t=1}^T (\hat{\theta}_{-t} - \bar{\hat{\theta}})^2} \quad (6)$$

where $\hat{\theta}_{-t}$ is the coefficient estimate of regression of probability in the lower component on time excluding year t and $\bar{\hat{\theta}}$ is the sample mean of these estimates. Note, Efron and Stein (1981) and Efron and Tibshirani (1994) both pointed out that the jackknife standard error tends to be biased upwards. In addition, as noted previously, the family-wise p -value is also overly-conservative. Nonetheless, rejection counts are reported under all three p -values.

Consider first the standard p -value: across the province-crop pairs, rejection rates are relatively high for Ontario soybean (3 of the 6 counties), Saskatchewan wheat (85 of the

¹⁰To obtain λ_t , weights in the lower component $w_{\ell,t}$ were regressed on time t .

Table 9: Summary Statistics for the Regression Coefficient of $\hat{\lambda}_t$

Province	Observations		Summary Statistics				
Crop	Counties	Years	Minimum	Mean	Median	Maximum	Std. Dev.
Ontario							
Corn	32	68	−0.00663	−0.00152	−0.00175	0.00477	0.00254
Soybean	6	68	−0.00709	−0.00304	−0.00351	0.00446	0.00405
Wheat	26	68	−0.00845	−0.00356	−0.00424	0.00103	0.00297
Saskatchewan							
Barley	204	79	−0.00678	−0.00155	−0.00176	0.00631	0.00249
Canola	144	47	−0.01039	−0.00081	−0.00060	0.01116	0.00395
Oats	131	79	−0.00654	−0.00109	−0.00110	0.00584	0.00248
Wheat	267	79	−0.00712	−0.00169	−0.00194	0.00627	0.00274

267), and Saskatchewan barley (49 of the 204). In contrast, only 10 of the 144 counties and 2 out of 26 reject this null for Saskatchewan canola and Ontario wheat, respectively. Overall, 167 of the 810 counties reject this null under the standard p -value. Not surprisingly, the rejection counts under the family-wise and jackknife p -values are much lower. Only 28 counties and 24 counties are able to reject the null under the family-wise and jackknife p -values, respectively.

Based on the second hypothesis, this hypothesis: $H_0 : \lambda_t \geq \lambda$ is a one-sided test which looks at whether the probability of a relatively bad year has statistically significantly decreased over time. Interestingly, the rejection counts in this one-sided null (187 out of the 810 counties) is larger than the previous two-sided null (167 counties in total) under the standard p -value. With respect to the family-wise and jackknife p -values, 38 counties and 41 counties are able to reject the one-sided null, respectively. Since there are more rejection counts in the one-sided null, this suggests that for counties have a statistically significant change in the probability of the lower component, the probabilities are most likely heading downward. Table 9 summarizes the regression coefficient for $\hat{\lambda}_t$. Consider the medians of the estimated coefficients: all the province-crop pairs exhibit a negative value, which indicates that the probability of a relatively bad year has generally decreased over time. Interestingly,

the magnitude is quite heterogeneous among the province-crop pairs. For Ontario wheat, the probability of a yield realization being drawn from a relatively bad year has been decreasing by 0.4% annually, whereas the probability for Saskatchewan canola has remained relatively constant over time.

5 Economic Implications for Crop Insurance

The previous section shows that yield volatility has a tendency to increase over time when the rate of technological change in relatively good years outperforms relatively bad years. In light of this, the change in yield volatility should theoretically also alter the expected yield loss, which affects the payout for the risk management programs in Canada. Like most of the developed countries, the Canadian government directs a significant amount of public monies to agricultural producers through risk management programs (\$3 billion to the latest five-year framework). Ker et al. (2017) suggested that risk management has become a popular tool to funnel monies to agricultural producers because publicly subsidized agriculture insurance is compliant with trade agreements and a relatively easy political sell to both producers and general public. Thus, subsidized crop insurance is likely to continue to play a significant role in the future.

Agricultural support programs are by no means new, as the first program was offered to all the Canadian farmers in 1958. For a more thorough history of the Canadian agricultural support programs in Canada, see Barichello (1995) and Ker et al. (2017). The latest policy framework – Canadian Agricultural Partnership (CAP) – is a five-year commitment (2018 to 2023) funded by the Canada’s federal, provincial, and territorial governments that aims to support the agricultural and food sector. The Business Risk Management (BRM) policy under CAP consists of four products: AgriInvest, AgriStability, AgriRecovery and AgriInsurance. They each have different functionality to mitigate production risks.

AgriInvest aims to ease the immediate cash flow problem for producers. Producers can deposit up to 100% of their Allowable Net Sales into the account and will receive a matching contribution from the government for the first 1%. Note, the maximum matching contribution is limited to \$10000 and producers have the right to withdraw the fund at any time. With respect to AgriStability, a producer will get compensated by the program if the net farming income declines more than 70% of the reference margin, which is calculated by the Olympic average (i.e. removing the highest and lowest observations) of the most recent five

years of production margin. For AgriRecovery, this framework is meant to mitigate the impacts of natural disasters. However, there is a lack of clarity on how AgriRecovery will be triggered. In general, a joint investigation is undertaken by the provincial/territorial and federal government if a disaster ever occurs. A payment will be made to the targeted producers based on the investigation. Finally, for AgriInsurance, the crop insurance program insures producers up to a certain percentage of the yield losses incurred. Producers can choose from different coverage levels depending on their farming operations and budgets for insurance premiums. The insurance premium is cost-shared among the federal government, provincial/territorial government and producer at a 36:24:40 ratio, respectively. Specifically, AgriInsurance insures Ontario and Saskatchewan producers up to 90% and 80% of the yield losses incurred, respectively.

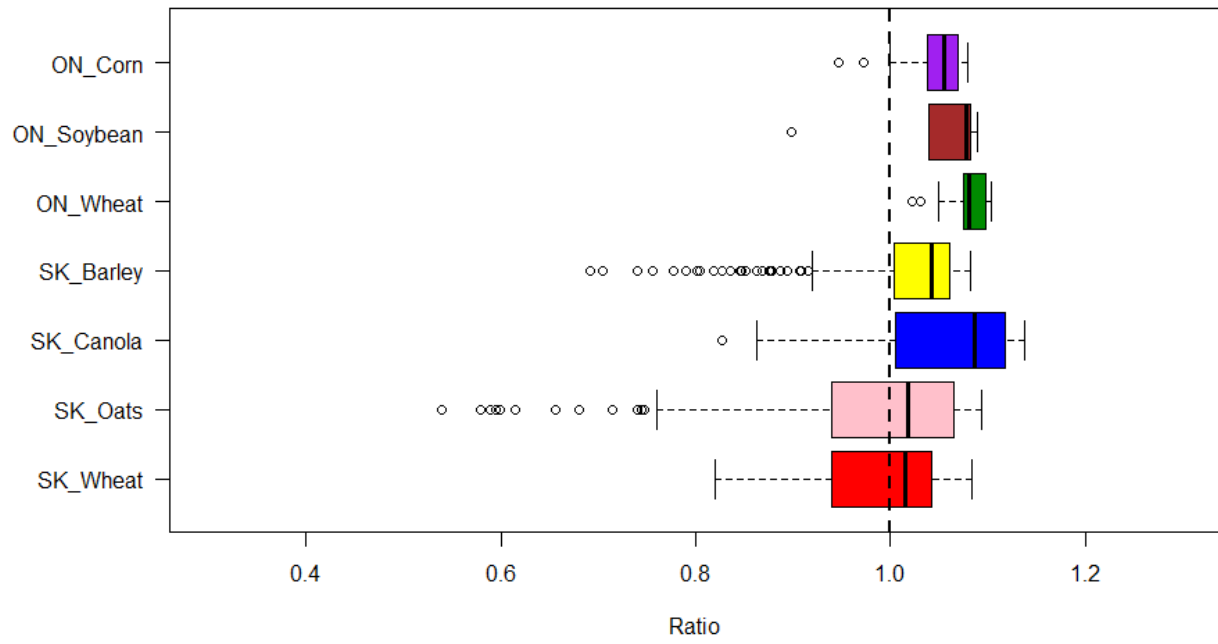
Among the four products, AgriInsurance is most relevant with respect to yield volatility since the payout is directly triggered by the average crop yield. Given that 60% of the insurance premium is subsidized by the federal government and provincial/territorial governments, the change of expected yield loss would be accompanied by the use of taxpayer money. Taken together, a better understanding of yield volatility is important as it has significant implications on the budgeted amount in the support programs. Recall that the data are analyzed at a county level, whereas the Canadian crop insurance premium ratings are evaluated at a farm level. Therefore, it is not surprising that these results are relatively conservative as variation at the farm level is generally higher than at the county level (Coble, Dismukes, and Thomas, 2007; Cooper et al., 2009; Claassen and Just, 2011). Nonetheless, these results serve to motivate the relevance of empirically investigated technological change on the crop insurance premium. To do so, the expected yield loss was compared between 2018 and 2023 (the period for the CAP framework), and between 2018 and 2028. Given that the county-level variation is underestimated, measuring expected yield loss at the highest possible coverage level would be more relevant. Therefore, a 90% coverage level was chosen to be the threshold, as it is the highest coverage level available for producers. The expected

yield loss (denoted by ρ) is given by:

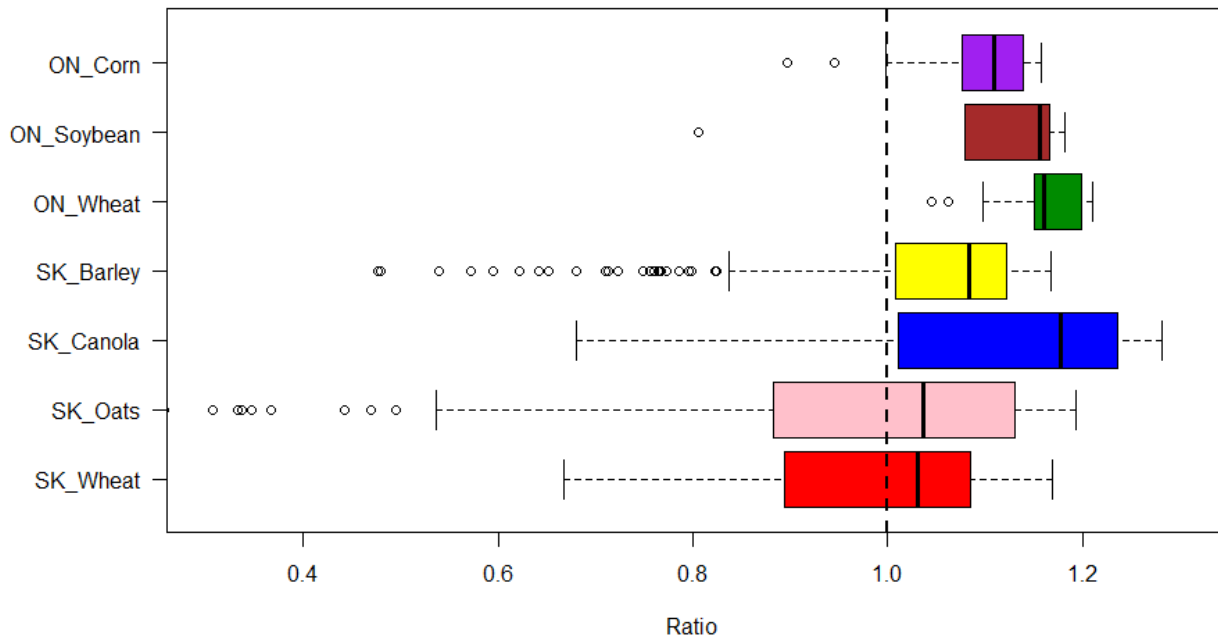
$$\rho = \int_0^{y_{90\%}} (y_{90\%} - y) f y dy \quad (7)$$

where y is the actual crop yield and $y_{90\%}$ is the crop yield at 90% coverage level. The expected yield loss is equivalent to the area under the estimated yield distribution between zero and average crop yield at 90% coverage level.

Figure 10(a) summarizes the ratio of expected yield loss ratio between 2023 and 2018 for all the 810 county-crop pairs in box plots. The ratio is equal to the expected yield loss in 2023 divided by the expected yield loss in 2018, meaning that the expected yield loss in 2023 is greater than in 2018 when the ratio is greater than one. Consider the medians of the box plots: all province-crop pairs are projected to encounter an increase in expected yield loss in the next five years, however, their magnitudes are quite different. For instance, expected yield loss is projected to increase 8.72% for Saskatchewan canola, 8.02% for Ontario wheat, and 7.74% for Ontario soybean. In contrast, the rate is much lower for Saskatchewan wheat (1.54%) and Saskatchewan oats (1.84%). Figure 10(b) presents the expected yield loss ratio between 2028 and 2018. It is not surprising that the results exhibit a wider range if the current trends remain unchanged. For example, expected yield loss in the next decade is projected to increase 17.79% for Saskatchewan canola and 16.07% for Ontario wheat. The overall results suggest that expected yield loss is projected to increase in the next five and ten years. Thus, it is not surprising that more taxpayer money will be funneled into the agricultural support programs in the future.



(a) Ratio of ρ_{2023}/ρ_{2018}



(b) Ratio of ρ_{2028}/ρ_{2018}

Figure 10: Ratio of Expected Yield Loss at 90% Coverage Level

6 Conclusions

The ability to meet future food demand will very likely continue to be dictated by technological change as other possible solutions are hindered by the rapid growth in developing countries. Although technological change increases yields over time, it may also increase or decrease yield volatility. This thesis aimed to achieve four objectives: (1) to investigate whether rate of technological change behaves heterogeneously across different components of a yield distribution; (2) to investigate how rate of technological change has an effect on yield volatility in a Canadian context, which has huge implications on areas such as farm income variability, rating crop insurance and food insecurity; (3) to examine whether these trends vary across crops and provinces; and (4) to estimate whether expected yield loss will vary in the next five and ten years. To this end, the analysis considered county-level yields for the major crops (barley, canola, corn, oats, soybean and wheat) in Ontario and Saskatchewan. Another contribution of this thesis is methodological; specifically, this thesis generalized the approach from Tolhurst and Ker (2015) and developed a procedure to test whether there is any potential efficiency gains to model yields by using different combinations of parameters.

All province-crop pairs, with the exception of Ontario corn, were estimated using a two-component mixture model with constant variances. Ontario corn was the only province-crop pair that was estimated using a two-component mixture model with time-varying variances for potentially different yield dispersion over time within a component. As a result, several interesting conclusions were reached. First, for Ontario corn, the annual increase in the variance of average crop yield for the lower component was generally higher in relatively bad years than in relatively good years, meaning that the downside yield risk has exhibited an increase over time.

Second, the vast majority of counties exhibited a higher rate of technological change in relatively good years than in relatively bad years. The diverging two technological trends also indicate that yield volatility has been increasing over the course of time. In addition, statistical significance was tested using a likelihood ratio test for all the 810 county-crop pairs.

Interestingly, Ontario had a relatively higher percentage of counties than Saskatchewan to exhibit a statistically significant different rate of technological change across the two components under a standard p -value, namely: Ontario corn (78%), Ontario wheat (58%), and Ontario soybean (50%). In a total of 810 counties, 297 counties rejected this null hypothesis under a standard p -value. As expected, most of the counties that had statistically significant differences in the rate of technological change between the two components, were very likely to have a higher rate of growth in the upper component than the lower component. In order to avoid the multiple testing issue, the rejection counts under the overly-conservative family-wise p -value were also reported. In total, 42 of the 810 counties remain to have statistically significant differences in the rate of technological change between the two components under the family-wise p -value. The asymmetrical rate of technological change between the two components can be contributed by countless factors, including but not limited to, technological advancement in seed genomics, increasing demand for the minimum amount of precipitation per plant, and the subsidized crop insurance which encourages producers to adopt high-risk and high-return technologies.

Third, the probability of a relatively bad year has been decreasing over time for the vast majority of counties. The results are of particular interest because factors such as technological change and climate change may increase or decrease the probability of lower component over time. This indicates that the data generating process of average crop yield is less likely to be drawn from a relatively bad year over time. Using a standard t -test for each of the 810 county-crop pairs, 167 out of the 810 counties exhibited a statistically significant increase or decrease probability of the lower component over time. In addition, 28 counties and 24 counties rejected the null under the family-wise p -value and jackknife p -value, respectively. Surprisingly, for counties have a statistically significant change in the probability of the lower component, the probabilities are most likely heading downward. Under a standard p -value, 187 out of the 810 counties exhibited a statistically significant decrease in the probability of the lower component. 38 counties and 41 counties rejected the

null under the family-wise p -value and jackknife p -value, respectively.

Finally, the expected yield loss will likely continue to increase in the future. Perhaps a more relevant economic measure — particularly with respect to the Canadian Agricultural Partnership framework — is the expected change of crop insurance payouts in the next five and ten years. Payouts were estimated to increase over time in different magnitudes. For instance, payout for Ontario wheat was projected to increase 8.02% and 16.07% at the 90% coverage level by 2023 and 2028, respectively. In contrast, payout for Saskatchewan wheat was projected to increase 1.54% and 3.1% at the 90% coverage level by 2023 and 2028, respectively. The overall increase in crop insurance payout suggests that more taxpayer's monies are expected to funnel into the Canadian agricultural support programs, given that 60% of the crop insurance premium is subsidized by the provincial/territorial governments and federal government.

In summary, this thesis provided a compelling evidence that technological change has led to a change in yield volatility over time. Unfortunately, most of the counties exhibited an increase in yield volatility. If these current trends remain changed, the expected yield loss is likely to continue to increase in the future, which consequently also requires a higher amount of taxpayer money to subsidize the crop insurance premium. For future research, several interesting applications that are related to this thesis are suggested: (i) decompose the effects of technological change, climate change, soil quality and policy change on yield volatility; (ii) investigate why certain counties or crops have relatively lower rate of technological change in the lower component; (iii) provide solutions to increase rate of technological change in the lower component, which may be helpful to decrease the overall variation of crop yields, and (iv) investigate how the Canadian crop insurance program would affect agricultural producer's willingness to adopt new technologies.

Bibliography

- Agriculture and Agri-Food Canada. 2017. "Evaluation of AgriStability, AgriInvest, AgriInsurance and the Wildlife Compensation Program." <http://www.agr.gc.ca/eng/about-us/offices-and-locations/office-of-audit-and-evaluation/evaluation-reports/evaluation-of-agristability-agriinvest-agriinsurance-and-the-wildlife-compensation-program/?id=1503612344518>, Accessed: 2018-12-08.
- Akaike, H. 1974. "A new look at the statistical model identification." *IEEE transactions on automatic control* 19:716–723.
- Atwood, J., S. Shaik, and M. Watts. 2003. "Are crop yields normally distributed? A reexamination." *American Journal of Agricultural Economics* 85:888–901.
- Atwood, L., A. Wilson, R. Elston, and J. Bailey-Wilson. 1992. "Computational aspects of fitting a mixture of two normal distributions using maximum likelihood." *Communications in Statistics-Simulation and Computation* 21:769–781.
- Barichello, R.R. 1995. "Overview of Canadian agricultural policy systems." Working paper.
- Bilmes, J.A., et al. 1998. "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models." *International Computer Science Institute* 4:126.
- Botts, R.R., and J.N. Boles. 1958. "Use of normal-curve theory in crop insurance ratemaking." *Journal of Farm Economics* 40:733–740.
- 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.
- Chen, J., and P. Li. 2009. "Hypothesis test for normal mixture models: The EM approach." *The Annals of Statistics* 37(5A):2523–2542.
- Chen, S.L., M. Miranda, et al. 2004. "Modeling multivariate crop yield densities with frequent extreme events." In *Selected paper presented at AAEA meeting*.
- Chen, S.Y., Z. Feng, and X. Yi. 2017. "A general introduction to adjustment for multiple comparisons." *Journal of thoracic disease* 9:1725.
- Claassen, R., and R.E. Just. 2011. "Heterogeneity and distributional form of farm-level yields." *American journal of agricultural economics* 93:144–160.
- Coble, K.H., R. Dismukes, and S.E. Thomas. 2007. "Policy implications of crop yield and revenue variability at differing levels of disaggregation." Working paper.
- Conforti, P., et al. 2011. *Looking ahead in world food and agriculture: perspectives to 2050..* Food and Agriculture Organization of the United Nations (FAO).

- Cooper, J.C., M.R. Langemeier, G.D. Schnitkey, and C.R. Zulauf. 2009. "Constructing farm level yield densities from aggregated data: analysis and comparison of approaches." Working paper.
- Crow, J.F. 1998. "90 years ago: the beginning of hybrid maize." *Genetics* 148:923–928.
- Day, R.H. 1965. "Probability distributions of field crop yields." *Journal of farm Economics* 47:713–741.
- De Boer, J., and H. Aiking. 2011. "On the merits of plant-based proteins for global food security: Marrying macro and micro perspectives." *Ecological Economics* 70:1259–1265.
- DeLucia, E.H., N. Gomez-Casanovas, J.A. Greenberg, T.W. Hudiburg, I.B. Kantola, S.P. Long, A.D. Miller, D.R. Ort, and W.J. Parton. 2014. "The theoretical limit to plant productivity." *Environmental Science & Technology* 48:9471–9477.
- Du, X., D.A. Hennessy, and C.L. Yu. 2012. "Testing Days conjecture that more nitrogen decreases crop yield skewness." *American Journal of Agricultural Economics* 94(1):225–237.
- Du, X., C.L. Yu, D.A. Hennessy, and R. Miao. 2015. "Geography of crop yield skewness." *Agricultural economics* 46:463–473.
- Edgerton, M.D. 2009. "Increasing crop productivity to meet global needs for feed, food, and fuel." *Plant physiology* 149:7–13.
- Efron, B., and C. Stein. 1981. "The jackknife estimate of variance." *The Annals of Statistics*, pp. 586–596.
- Efron, B., and R.J. Tibshirani. 1994. *An Introduction to the Bootstrap*. Boca Raton, FL: Chapman & Hall/CRC Press.
- Everitt, B.S., and D.J. Hand. 1981. *Finite Mixture Distributions*. Springer.
- Finch, S.J., N.R. Mendell, and H.C. Thode Jr. 1989. "Probabilistic measures of adequacy of a numerical search for a global maximum." *Journal of the American Statistical Association* 84:1020–1023.
- Gallagher, P. 1987. "US soybean yields: Estimation and forecasting with nonsymmetric disturbances." *American Journal of Agricultural Economics* 69:796–803.
- Godfray, H.C.J., J.R. Beddington, I.R. Crute, L. Haddad, D. Lawrence, J.F. Muir, J. Pretty, S. Robinson, S.M. Thomas, and C. Toulmin. 2010. "Food security: the challenge of feeding 9 billion people." *science* 327:812–818.
- 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 O. Mahul. 2004. *Risk modeling concepts relating to the design and rating of agricultural insurance contracts*. The World Bank.
- Goodwin, B.K., M.C. Roberts, and K.H. Coble. 2000. "Measurement of price risk in revenue insurance: implications of distributional assumptions." *Journal of Agricultural and Resource Economics*, pp. 195–214.
- Grafton, R.Q., C. Daugbjerg, and M.E. Qureshi. 2015. "Towards food security by 2050." *Food Security* 7:179–183.
- Gustafsson, J., C. Cederberg, U. Sonesson, and A. Emanuelsson. 2013. *The methodology of the FAO study: Global Food Losses and Food Waste-extent, causes and prevention-FAO, 2011*. SIK Institutet för livsmedel och bioteknik.
- Hall, J.A., B.W. Brorsen, and S.H. Irwin. 1989. "The distribution of futures prices: A test of the stable Paretian and mixture of normals hypotheses." *Journal of Financial and Quantitative Analysis* 24:105–116.
- Harri, A., K.H. Coble, A.P. Ker, and B.J. Goodwin. 2011. "Relaxing heteroscedasticity assumptions in area-yield crop insurance rating." *American Journal of Agricultural Economics* 93:707–717.
- Hennessy, D.A. 2009. "Crop yield skewness under law of the minimum technology." *American Journal of Agricultural Economics* 91(1):197–208.
- Holm, S. 1979. "A simple sequentially rejective multiple test procedure." *Scandinavian Journal of Statistics* 6(2):65–70.
- Just, R.E., and Q. Weninger. 1999. "Are crop yields normally distributed?" *American Journal of Agricultural Economics* 81:287–304.
- Karlis, D., and E. Xekalaki. 2003. "Choosing initial values for the EM algorithm for finite mixtures." *Computational Statistics & Data Analysis* 41(3):577–590.
- Kaylen, M.S., and S.S. Koroma. 1991. "Trend, weather variables, and the distribution of US corn yields." *Review of agricultural economics* 13:249–258.
- Ker, A., and A. Tolga Ergun. 2007. "On the revelation of private information in the US crop insurance program." *Journal of Risk and Insurance* 74:761–776.
- Ker, A.P., B. Barnett, D. Jacques, and T. Tolhurst. 2017. "Canadian business risk management: Private firms, crown corporations, and public Institutions." *Canadian Journal of Agricultural Economics* 65(4):591–612.
- 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 T.N. Tolhurst. 2019. "On the Treatment of Heteroscedasticity in Crop Yield Data." *American Journal of Agricultural Economics* (forthcoming).
- 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.
- Koenker, R., and G. Bassett Jr. 1978. "Regression quantiles." *Econometrica: journal of the Econometric Society*, pp. 33–50.
- Kummu, M., H. de Moel, M. Porkka, S. Siebert, O. Varis, and P.J. Ward. 2012. "Lost food, wasted resources: Global food supply chain losses and their impacts on freshwater, cropland, and fertiliser use." *Science of the total environment* 438:477–489.
- Lipinski, B., C. Hanson, J. Lomax, L. Kitinoja, R. Waite, and T. Searchinger. 2013. "Reducing food loss and waste." *World Resources Institute Working Paper, June*, pp. .
- Lobell, D.B., and G.P. Asner. 2003. "Climate and management contributions to recent trends in US agricultural yields." *Science* 299:1032–1032.
- Lobell, D.B., and C.B. Field. 2007. "Global scale climate–crop yield relationships and the impacts of recent warming." *Environmental research letters* 2:014002.
- Machovina, B., K.J. Feeley, and W.J. Ripple. 2015. "Biodiversity conservation: The key is reducing meat consumption." *Science of the Total Environment* 536:419–431.
- McKenzie, F.C., and J. Williams. 2015. "Sustainable food production: constraints, challenges and choices by 2050." *Food Security* 7:221–233.
- McLachlan, G. 2018. "Finite mixture models." *Annual Review of Statistics and Its Application* 5.
- Moss, C.B., and J.S. Shonkwiler. 1993. "Estimating yield distributions with a stochastic trend and non-normal errors." *American Journal of Agricultural Economics* 75(5):1056–1062.
- Nelson, C.H., and P.V. Preckel. 1989. "The conditional beta distribution as a stochastic production function." *American Journal of Agricultural Economics* 71:370–378.
- 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.
- Ort, D.R., and S.P. Long. 2014. "Limits on yields in the corn belt." *Science* 344(6183):484–485.
- Ozaki, V.A., B.K. Goodwin, and R. Shirota. 2008. "Parametric and nonparametric statistical modelling of crop yield: implications for pricing crop insurance contracts." *Applied Economics* 40:1151–1164.

- Ozaki, V.A., and R.S. Silva. 2009. "Bayesian ratemaking procedure of crop insurance contracts with skewed distribution." *Journal of Applied Statistics* 36(4):443–452.
- Parfitt, J., M. Barthel, and S. Macnaughton. 2010. "Food waste within food supply chains: quantification and potential for change to 2050." *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 365:3065–3081.
- Pimentel, D., and M. Pimentel. 2003. "Sustainability of meat-based and plant-based diets and the environment." *The American journal of clinical nutrition* 78:660S–663S.
- Pretty, J., W.J. Sutherland, J. Ashby, J. Auburn, D. Baulcombe, M. Bell, J. Bentley, S. Bickerteth, K. Brown, J. Burke, et al. 2010. "The top 100 questions of importance to the future of global agriculture." *International journal of agricultural sustainability* 8:219–236.
- 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.
- Ramirez, O.A. 1997. "Parametric model for simulating heteroskedastic, correlated, nonnormal random variables: The case of Corn Belt corn, soybean and wheat yields." *American Journal of Agricultural Economics* 79(1):191–205.
- Ramirez, O.A., S. Misra, and J. Field. 2003. "Crop-yield distributions revisited." *American Journal of Agricultural Economics* 85:108–120.
- Ranganathan, J., D. Vennard, R. Waite, P. Dumas, B. Lipinski, T. Searchinger, et al. 2016. "Shifting diets for a sustainable food future." *World Resources Institute*, pp. .
- Searchinger, T., C. Hanson, J. Ranganathan, B. Lipinski, R. Waite, R. Winterbottom, A. Dinshaw, R. Heimlich, M. Boval, P. Chemineau, et al. 2014. "Creating a sustainable food future. A menu of solutions to sustainably feed more than 9 billion people by 2050. World resources report 2013-14: interim findings.", pp. .
- Sherrick, B.J., F.C. Zanini, G.D. Schnitkey, and S.H. Irwin. 2004. "Crop insurance valuation under alternative yield distributions." *American Journal of Agricultural Economics* 86:406–419.
- Skees, J.R., and M.R. Reed. 1986. "Rate making for farm-level crop insurance: implications for adverse selection." *American Journal of Agricultural Economics* 68:653–659.
- Statistics Canada. 2016. "Census of Agriculture (3438)."
- . 2018. "Estimated areas, yield, production, average farm price and total farm value of principal field crops, in metric and imperial units."
- Stochs, S.M., and J. LaFrance. 2004. "A learning rule for inferring local distributions over space and time." In T. B. Fomby and R. C. Hill, eds. *Advances in Econometrics, Volume 18: Spatial and Spatiotemporal Econometrics*. New York: Elsevier Science, pp. 295–331.

- 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(5):1037–1054.
- Tarasuk, V., A. Mitchell, and N. Dachner. 2016. *Household food insecurity in Canada, 2014*. Research to identify policy options to reduce food insecurity.
- Tolhurst, T.N., and A.P. Ker. 2015. "On technological change in crop yields." *American Journal of Agricultural Economics* 97:137–158.
- Turvey, C., J. Zhao, et al. 1999. *Parametric and non-parametric crop yield distributions and their effects on all-risk crop insurance premiums*. Department of Agricultural Economics and Business, University of Guelph.
- U.S. Department of Agriculture. 2018. "Quick Stats." <https://quickstats.nass.usda.gov/>, Accessed: 2018-12-08.
- White, T. 2000. "Diet and the distribution of environmental impact." *Ecological economics* 34:145–153.
- 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.
- Wu, X., and Y.Y. Zhang. 2012. "Nonparametric estimation of crop yield distributions: A panel data approach." AAEE 2012 Annual Meeting, August 12-14, Seattle, WA., No. 124630.
- Zhu, Y., B.K. Goodwin, and S.K. Ghosh. 2011. "Modeling yield risk under technological change: Dynamic yield distributions and the US crop insurance program." *Journal of Agricultural and Resource Economics*, pp. 192–210.

7 Appendix

7.1 Summary Statistics of Estimated Parameters, by Province and Crop

Table 10: Summary Statistics of Estimated Parameters, by Province and Crop, Part 1 of 2

Province	Crop	Parameter	Min.	Mean	Median	Max.	Std. Dev.
ON	Corn	λ	0.03	0.37	0.42	0.96	0.26
		α_ℓ	10.75	42.23	46.54	51.85	10.20
		β_ℓ	0.20	1.06	1.06	1.99	0.40
		γ_ℓ	0.00	19.09	4.13	241.57	46.07
		δ_ℓ	-1.85	2.06	2.01	9.63	2.13
		α_u	38.46	46.63	47.24	54.21	4.02
		β_u	1.10	1.53	1.54	1.91	0.23
		γ_u	0.00	11.75	6.46	47.46	12.94
		δ_u	-0.22	1.99	1.67	5.62	1.65
ON	Soybean	λ	0.06	0.18	0.19	0.26	0.07
		α_ℓ	14.37	18.92	19.00	23.92	3.09
		β_ℓ	0.09	0.22	0.19	0.42	0.11
		γ_ℓ	7.28	22.87	17.74	82.81	4.88
		α_u	18.74	20.81	19.70	23.94	2.47
		β_u	0.34	0.39	0.40	0.45	0.04
		γ_u	4.50	9.13	8.03	16.96	0.57
ON	Wheat	λ	0.02	0.25	0.19	0.72	0.23
		α_ℓ	13.56	25.79	26.14	32.16	3.77
		β_ℓ	0.25	0.40	0.38	0.64	0.10
		γ_ℓ	0.45	10.80	15.83	46.36	3.53
		α_u	18.43	26.49	26.14	32.16	2.73
		β_u	0.54	0.77	0.77	1.09	0.14
		γ_u	8.51	32.04	30.10	158.63	2.97
SK	Barley	λ	0.02	0.31	0.17	0.98	0.30
		α_ℓ	0.22	13.51	14.52	27.23	6.16
		β_ℓ	0.00	0.29	0.31	0.63	0.14
		γ_ℓ	0.58	42.09	54.24	237.64	13.29
		α_u	10.40	21.41	20.52	39.29	5.35
		β_u	0.03	0.53	0.53	1.09	0.14
		γ_u	0.21	47.11	52.59	150.56	6.85

Table 11: Summary Statistics of Estimated Parameters, by Province and Crop, Part 2 of 2

Province	Crop	Parameter	Min.	Mean	Median	Max.	Std. Dev.
SK	Canola	λ	0.03	0.43	0.47	0.98	0.27
		α_ℓ	0.24	14.37	15.25	22.06	4.87
		β_ℓ	-0.03	0.18	0.17	0.55	0.13
		γ_ℓ	0.21	15.75	19.38	151.87	4.27
		α_u	10.97	17.23	17.27	22.07	2.30
		β_u	0.19	0.42	0.42	0.85	0.11
		γ_u	0.05	13.45	13.64	316.27	3.55
SK	Oats	λ	0.01	0.38	0.33	0.98	0.29
		α_ℓ	0.08	17.99	19.18	39.87	8.90
		β_ℓ	0.01	0.41	0.44	0.86	0.19
		γ_ℓ	1.51	73.75	86.58	452.03	17.92
		α_u	14.80	28.56	28.14	56.14	7.38
		β_u	0.04	0.67	0.65	1.71	0.26
		γ_u	0.79	91.05	101.29	447.54	13.96
SK	Wheat	λ	0.02	0.34	0.18	1.00	0.32
		α_ℓ	0.03	9.71	8.85	22.10	5.41
		β_ℓ	0.00	0.18	0.19	0.42	0.11
		γ_ℓ	0.29	17.46	16.28	168.25	6.24
		α_u	4.17	16.17	16.36	28.58	4.32
		β_u	0.05	0.29	0.28	0.66	0.10
		γ_u	0.08	23.74	28.24	88.81	3.35

Codes in R

Model 4 and 4R

```
#Read crop data
crop <- read.csv("Crop_Data.csv")
#Municipality number (-1 is to exclude the year column)
municipal <- colnames(crop[,-1])
#number of municipalities
n_municipal <- length(municipal)
library(optimization)
library(quantreg)
#Model 4 and 4R 2-component Constant Variance
p_final_penalty <- matrix(0, n_municipal, 7)
l_final_penalty <- matrix(0, n_municipal, 1)
#Multiple Starting Value
starting <- c(.2,.35,.5,.65,.8)
for(i in 1:n_municipal){
y <- crop[,i+1]
#penalized log-likelihood
fun<-function(start) {-(sum(log(start[1]*dnorm(y,start[2]+start[3]*t,
start[4])+(1-start[1])*dnorm(y,start[5]+start[6]*t,
start[7]))) - penalty*(start[1]-0)^2)}
#log-likelihood
fun.unr<-function(start)
{-(sum(log(start[1]*dnorm(y,start[2]+start[3]*t,start[4])+(
1-start[1])*dnorm(y,start[5]+start[6]*t,start[7]))))}
tol<-0.001
conv<-0
it.conv<-0
T<-length(y)
t<-seq(1:T)
p_loop <- matrix (0, length(starting), 7)
l_loop <- matrix (0, length(starting), 1)
for(s in 1:length(starting)){
result<-rq(y~t,starting[s])
w<-rep(0,T)
w[result$residuals <0]<-1
yy<-c(y,y)
x1t<-c(t,rep(0,T))
x2t<-c(rep(0,T),t)
x1<-c(rep(1,T),rep(0,T))
x2<-c(rep(0,T),rep(1,T))
yy<-c(y,y)
ww<-c(w,1-w)
```

```

XXR<-cbind(x1,x1t,x2,x2t)
coef.last<-rep(0,7)
temp.a<-10000
j<-1
TT<-T+1
TTT<-2*T
while (conv < 1){
ww<-c(w,1-w)
result<-lm(yy~XXR-1,weights=ww)
ss2 <- (sum((1-w)*(result$residuals [TT:TTT])^2)/(sum(1-w)))^.5
ss1 <- pmax((0.1*ss2),
(sum(w*(result$residuals [1:T])^2)/(sum(w)))^.5)
p1<-dnorm(y,result$fitted.values [1:T],ss1)
p2<-dnorm(y,result$fitted.values [TT:TTT],ss2)
w <- p1/(p1+p2)
mw<-mean(w)
coef.new<-
c(mw,result$coefficients [1:2],ss1,result$coefficients [3:4],ss2)
temp.a <- sum(abs(coef.new-coef.last))
if (j>500) temp.a<-0
if (temp.a<tol) {conv<-1
it.conv<-j}
coef.last<-coef.new
j<-j+1
}
p_loop[s,] <- coef.new
l_loop[s] <- fun.unr(coef.new)
}
coef.new <- p_loop [which.max(l_loop),]
fit1<-coef.new[2]+coef.new[3]*t
fit2<-coef.new[5]+coef.new[6]*t
start<-as.numeric(coef.new)
start2 <- c(pmin(1, start [1]), start [2:7])
ll<-rep(99999999,100)
trial <- matrix(0, 100, 7)
for (k in 1:100){
penalty<-k/2
result.opt<-optim(start2,fun)
coef.opt<-result.opt$par
ll[k]<-fun.unr(coef.opt)
if (coef.opt[1]>1) ll[k]<-9999999999 #avoid lambda > 1
if (min(coef.opt)<0) ll[k]<-9999999999 #avoid parameters < 0
#lower variance has to be above 10% of the upper to avoid collapse
if (coef.opt[4]<0.1*coef.opt [7]) ll[k]<-9999999999
trial[k,] <- result.opt$par

```

```

}
penalty<-which.min(ll)
coef.opt <- trial[penalty,]
ll[which(ll == "NaN")] <- 9999999999
if(min(ll)>fun.unr(coef.new)){coef.opt <- coef.new}
if(min(ll)>5000){coef.opt <- coef.new}
opt1<-coef.opt[2]+coef.opt[3]*t
opt2<-coef.opt[5]+coef.opt[6]*t
plot(t,y)
lines(t,fit1,col="red")
lines(t,fit2,col="red")
lines(t,opt1,col="blue")
lines(t,opt2,col="blue")
fun(coef.new)
fun(coef.opt)
fun.unr(coef.new)
fun.unr(coef.opt)
p_final_penalty[i,] <- c(coef.opt[1], coef.opt[2], coef.opt[3],
coef.opt[4], coef.opt[5], coef.opt[6], coef.opt[7])
l_final_penalty[i] <- -(fun.unr(coef.opt))
}
#Make sure the blue line is the upper trend
uu <- matrix(0, n_municipal, 1)
for(i in 1:n_municipal){
if(mean(p_final_penalty[i,2]+ p_final_penalty[i,3]*t) >
mean(p_final_penalty[i,5]+ p_final_penalty[i,6]*t)){uu[i]<-1}
}
mm<- which(uu==1)
#switch back the parameters
for(i in mm){
store <- p_final_penalty[i,]
p_final_penalty[i,1] <- 1- p_final_penalty[i,1]
p_final_penalty[i,2] <- p_final_penalty[i,5]
p_final_penalty[i,3] <- p_final_penalty[i,6]
p_final_penalty[i,4] <- p_final_penalty[i,7]
p_final_penalty[i,5] <- store[2]
p_final_penalty[i,6] <- store[3]
p_final_penalty[i,7] <- store[4]
}
#Find out crossing trends
oo <- matrix(0, n_municipal, 1)
for(i in 1:n_municipal){
if((p_final_penalty[i,2]>p_final_penalty[i,5])
|
((p_final_penalty[i,2]+p_final_penalty[i,3]*T)>

```

```

(p_final_penalty[i,5]+p_final_penalty[i,6]*T)))
{oo[i,1]<-1}
}
aa <- which(oo == 0)
bb <- which(oo == 1)
#Restrict the two trends not to cross
p_final_restrict_penalty <- matrix (0, n_municipal, 7)
l_final_restrict_penalty <- matrix (0, n_municipal, 1)
for(i in aa){
p_final_restrict_penalty[i,] <- p_final_penalty[i,]
l_final_restrict_penalty[i] <- l_final_penalty[i]
}
#Set counties that cross to start from the same intercept
p_final_restrict_penalty8 <- matrix (0, n_municipal, 7)
l_final_restrict_penalty8 <- matrix (0, n_municipal, 1)
for(i in bb){
#Penalized log-likelihood
fun<-function(start)
{-(sum(log(start[1]*dnorm(y,start[2]+start[3]*t,
start[4])+(1-start[1])*dnorm(y,start[2]+start[5]*t,
start[6])))-penalty*(start[1]-0)^2)}
#log-likelihood
fun.unr<-function(start)
{-(sum(log(start[1]*dnorm(y,start[2]+start[3]*t,
start[4])+(1-start[1])*dnorm(y,start[2]+start[5]*t,start[6]))))}
y<-crop[,i+1]
tol<-0.001
conv<-0
it.conv<-0
T<-length(y)
t<-seq(1:T)
p_loop <- matrix (0, length(starting), 6)
l_loop <- matrix (0, length(starting), 1)
for(s in 1:length(starting)){
result<-rq(y~t,starting[s])
w<-rep(0,T)
w[result$residuals<0]<-1
yy<-c(y,y)
x1t<-c(t,rep(0,T))
x2t<-c(rep(0,T),t)
x1<-c(rep(1,T),rep(1,T))
yy<-c(y,y)
ww<-c(w,1-w)
XXR<-cbind(x1,x1t,x2t)
conv<-0

```

```

coef.last<-rep(0,6)
temp.a<-10000
j<-1
TT<-T+1
TTT<-2*T
while (conv < 1){
ww<-c(w,1-w)
result<-lm(yy~XXR-1,weights=ww)
ss2 <- (sum((1-w)*(result$residuals [TT:TTT])^2)/(sum(1-w)))^.5
ss1 <- pmax((0.1*ss2),
(sum(w*(result$residuals [1:T])^2)/(sum(w)))^.5)
mw<-mean(w)
p1<-dnorm(y,result$fitted.values [1:T],ss1)
p2<-dnorm(y,result$fitted.values [TT:TTT],ss2)
w <- p1/(p1+p2)
coef.new<-c(mw,result$coefficients [1:2],ss1,
result$coefficients [3],ss2)
temp.a <- sum(abs(coef.new-coef.last))
if (j>500) temp.a<-0
if (temp.a<tol) {conv<-1
it.conv<-j}
coef.last<-coef.new
j<-j+1
}
p_loop[s,] <- coef.new
l_loop[s] <- fun.unr(coef.new)
}
coef.new <- p_loop [which.max(l_loop),]
fit1<-coef.new[2]+coef.new[3]*t
fit2<-coef.new[2]+coef.new[5]*t
start<-as.numeric(coef.new)
start2 <- c(pmin(1, start [1]), start [2:6])
ll<-rep(99999999,100)
trial <- matrix(0, 100, 6)
for (k in 1:100){
penalty<-k
result.opt<-optim(start2,fun)
coef.opt<-result.opt$par
ll[k]<-fun.unr(coef.opt)
if (coef.opt[1]>1) ll[k]<-9999999999 #avoid lambda is > 1
if (min(coef.opt)<0) ll[k]<-9999999999 #avoid parameters < 0
#lower variance has to be above 10% of the upper
if (coef.opt[4]<0.1*coef.opt[6]) ll[k]<-9999999999
trial[k,] <- result.opt$par
}

```

```

penalty<-which.min(ll)
coef.opt <- trial[penalty,]
if(min(ll)>fun.unr(coef.new)){coef.opt <- coef.new}
opt1<-coef.opt[2]+coef.opt[3]*t
opt2<-coef.opt[2]+coef.opt[5]*t
plot(t,y)
lines(t,fit1,col="red")
lines(t,fit2,col="red")
lines(t,opt1,col="blue")
lines(t,opt2,col="blue")
fun(coef.new)
fun(coef.opt)
fun.unr(coef.new)
fun.unr(coef.opt)
p_final_restrict_penalty8 [i,] <-
c(coef.opt[1], coef.opt[2], coef.opt[3],
coef.opt[4], coef.opt[2], coef.opt[5], coef.opt[6])
l_final_restrict_penalty8 [i] <- -(fun.unr(coef.opt))
}
#Find the non-crossing parameters with the highest likelihood
p_final_restrict_penalty9 <- matrix (0, n_municipal, 7)
l_final_restrict_penalty9 <- matrix (0, n_municipal, 1)
starting <- c(.2,.35,.5,.65,.8)
for(i in bb){
y <- crop[,i+1]
#penalized log-likelihood
fun<-function(start) {-(sum(log(start[1]*dnorm(y,start[2]+start[3]*t,
start[4])+(1-start[1])*dnorm(y,start[5]+start[6]*t,
start[7])))-penalty*(start[1]-0)^2)}
#log-likelihood
fun.unr<-function(start)
{-(sum(log(start[1]*dnorm(y,start[2]+start[3]*t,
start[4])+(1-start[1])*dnorm(y,start[5]+start[6]*t,start[7]))))}
tol<-0.001
conv<-0
it.conv<-0
T<-length(y)
t<-seq(1:T)
p_loop <- matrix (0, length(starting), 7)
l_loop <- matrix (0, length(starting), 1)
for(s in 1:length(starting)){
result<-rq(y~t,starting[s])
w<-rep(0,T)
w[result$residuals<0]<-1
yy<-c(y,y)

```

```

x1t<-c(t,rep(0,T))
x2t<-c(rep(0,T),t)
x1<-c(rep(1,T),rep(0,T))
x2<-c(rep(0,T),rep(1,T))
yy<-c(y,y)
ww<-c(w,1-w)
XXR<-cbind(x1,x1t,x2,x2t)
coef.last<-rep(0,7)
temp.a<-10000
j<-1
TT<-T+1
TTT<-2*T
while (conv < 1){
ww<-c(w,1-w)
result<-lm(yy~XXR-1,weights=ww)
ss2 <- (sum((1-w)*(result$residuals[TT:TTT])^2)/(sum(1-w)))^ .5
ss1 <- pmax((0.1*ss2), (sum(w*(result$residuals[1:T])^2)/
(sum(w)))^ .5)
p1<-dnorm(y,result$fitted.values[1:T],ss1)
p2<-dnorm(y,result$fitted.values[TT:TTT],ss2)
w <- p1/(p1+p2)
mw<-mean(w)
coef.new<-c(mw,result$coefficients[1:2],ss1,
result$coefficients[3:4],ss2)
temp.a <- sum(abs(coef.new-coef.last))
if (j>500) temp.a<-0
if (temp.a<tol) {conv<-1
it.conv<-j}
coef.last<-coef.new
j<-j+1
}
p_loop[s,] <- coef.new
l_loop[s] <- fun.unr(coef.new)
}
coef.new <- p_loop [which.max(l_loop),]
fit1<-coef.new[2]+coef.new[3]*t
fit2<-coef.new[5]+coef.new[6]*t
start<-as.numeric(coef.new)
start2 <- c(pmin(1, start[1]), start[2:7])
ll<-rep(99999999,100)
trial <- matrix(0, 100, 7)
for (k in 1:100){
penalty<-k/2
result.opt<-optim(start2,fun)
coef.opt<-result.opt$par

```

```

ll[k]<-fun.unr(coef.opt)
if (coef.opt[1]>1) ll[k]<-9999999999 #avoid lambda > 1
if (min(coef.opt)<0) ll[k]<-9999999999 #avoid parameters < 0
#lower variance has to be above 10% of the upper
if (coef.opt[4]<0.1*coef.opt[7]) ll[k]<-9999999999
if ((mean(coef.opt[2]+ coef.opt[3]*t) <
mean(coef.opt[5]+coef.opt[6]*t))
&
((coef.opt[2]+coef.opt[3]*T) > (coef.opt[5]+coef.opt[6]*T)))
{ll[k]<-9999999999}
if ((mean(coef.opt[2]+ coef.opt[3]*t) <
mean(coef.opt[5]+coef.opt[6]*t))&
(coef.opt[2]>coef.opt[5])){ll[k]<-9999999999}
if ((mean(coef.opt[2]+ coef.opt[3]*t) >
mean(coef.opt[5]+coef.opt[6]*t))&
((coef.opt[2]+coef.opt[3]*T) < (coef.opt[5]+coef.opt[6]*T)))
{ll[k]<-9999999999}
if ((mean(coef.opt[2]+ coef.opt[3]*t) >
mean(coef.opt[5]+coef.opt[6]*t))&
(coef.opt[2]<coef.opt[5]))
{ll[k]<-9999999999}
trial[k,] <- result.opt$par
}
penalty<-which.min(ll)
coef.opt <- trial[penalty,]
ll[which(ll == "NaN")] <- 9999999999
if (min(ll)>5000) {coef.opt <- c(0,0,0,0,0,0,0)}
if ((fun.unr(coef.new)<fun.unr(coef.opt))
& (coef.new[2]<coef.new[5])
&
((coef.new[2]+coef.new[3]*T) < (coef.new[5]+coef.new[6]*T))
){coef.opt <- coef.new}
opt1<-coef.opt[2]+coef.opt[3]*t
opt2<-coef.opt[5]+coef.opt[6]*t
plot(t,y)
lines(t,fit1,col="red")
lines(t,fit2,col="red")
lines(t,opt1,col="blue")
lines(t,opt2,col="blue")
fun(coef.new)
fun(coef.opt)
fun.unr(coef.new)
fun.unr(coef.opt)
#log-likelihood
p_final_restrict_penalty9 [i,] <- c(coef.opt[1],

```

```

coef.opt[2], coef.opt[3],
coef.opt[4], coef.opt[5], coef.opt[6], coef.opt[7])
l_final_restrict_penalty9 [i] <- -(fun.unr(coef.opt))
}
cbind(l_final_restrict_penalty8, l_final_restrict_penalty9)
#choose between same intercept and different intercept
for(i in bb){
  if (l_final_restrict_penalty8[i] == "0" |
  l_final_restrict_penalty8[i] ==
  "-Inf"){l_final_restrict_penalty8[i]<--999999}
  if (l_final_restrict_penalty9[i] == "0" |
  l_final_restrict_penalty9[i] ==
  "-Inf"){l_final_restrict_penalty9[i]<--999999}
}
for(i in bb){
  if(l_final_restrict_penalty8[i] > l_final_restrict_penalty9[i]){
    p_final_restrict_penalty[i,] <- p_final_restrict_penalty8[i,]
    l_final_restrict_penalty[i] <- l_final_restrict_penalty8[i]}
  if(l_final_restrict_penalty8[i] < l_final_restrict_penalty9[i]){
    p_final_restrict_penalty[i,] <- p_final_restrict_penalty9[i,]
    l_final_restrict_penalty[i] <- l_final_restrict_penalty9[i]}
}
##Final check on crossing trends and restrict##
#make sure the blue line is the upper trend
uu <- matrix (0, n_municipal, 1)
for (i in 1:n_municipal){
  uu[i] <- (mean(p_final_restrict_penalty[i,2]+
  p_final_restrict_penalty[i,3]*t)>mean(p_final_restrict_penalty[i,5]+
  p_final_restrict_penalty[i,6]*t))
}
ss<- which(uu==1)
#switch back the parameters
for (i in ss){
  store <- p_final_restrict_penalty[i,]
  p_final_restrict_penalty[i,1] <- 1- p_final_restrict_penalty[i,1]
  p_final_restrict_penalty[i,2] <- p_final_restrict_penalty[i,5]
  p_final_restrict_penalty[i,3] <- p_final_restrict_penalty[i,6]
  p_final_restrict_penalty[i,4] <- p_final_restrict_penalty[i,7]
  p_final_restrict_penalty[i,5] <- store[2]
  p_final_restrict_penalty[i,6] <- store[3]
  p_final_restrict_penalty[i,7] <- store[4]
}
#final check if there's error
oo <- matrix (0, n_municipal, 1)
for (i in 1:n_municipal){

```

```

if (p_final_restrict_penalty[i,1]==0){oo[i]<-1}
if (l_final_restrict_penalty[i]==0){oo[i]<-1}
}
nn<- which(oo==1)

```

Model 5 and 5R

```

p_final_penalty_cv <- matrix(0, n_municipal, 9)
l_final_penalty_cv <- matrix(0, n_municipal, 1)
starting <- c(.2,.35,.5,.65,.8)
for(i in 1:n_municipal){
y <- crop[,i+1]
#penalized log-likelihood
fun<-function(start){-(sum(log(start[1]*dnorm(y,start[2]+start[3]*t,
(start[4]+start[5]*t)^0.5)+(1-start[1])*dnorm(y,start[6]+start[7]*t,
(start[8]+start[9]*t)^0.5))))-penalty*(start[1]-0)^2)}
#log-likelihood
fun.unr<-function(start){-(sum(log(start[1]*dnorm(y,start[2]
+start[3]*t,
(start[4]+start[5]*t)^0.5)+(1-start[1])*dnorm(y,start[6]
+start[7]*t, (start[8]+start[9]*t)^0.5)))))}
tol<-0.001
conv<-0
it.conv<-0
T<-length(y)
t<-seq(1:T)
p_loop <- matrix (0, length(starting), 9)
l_loop <- matrix (0, length(starting), 1)
for(s in 1:length(starting)){
result<-rq(y~t,starting[s])
w<-rep(0,T)
w[result$residuals <0]<-1
yy<-c(y,y)
x1t<-c(t,rep(0,T))
x2t<-c(rep(0,T),t)
x1<-c(rep(1,T),rep(0,T))
x2<-c(rep(0,T),rep(1,T))
yy<-c(y,y)
ww<-c(w,1-w)
XXR<-cbind(x1,x1t,x2,x2t)
coef.last<-rep(0,9)
temp.a<-10000
j<-1
TT<-T+1
TTT<-2*T
while (conv < 1){
ww<-c(w,1-w)
result<-lm(yy~XXR-1,weights=ww)
ss1 <- lm((result$residuals[1:T]^2) ~ t, weights = w)
if (ss1$coefficients[1]<0) {sss1<-

```

```

lm((result$residuals[1:T]^2) ~ t-1, weights = w)
ss1$coefficients[1]<-0
ss1$coefficients[2]<-ss1$coefficients[1]
ss1$fitted.values<-ss1$fitted.values}
if (min(ss1$fitted.values)<0)
{ss1$coefficients[1]<-mean(result$residuals[1:T]^2)
ss1$coefficients[2]<-0
ss1$fitted.values<-ss1$coefficients[1]}
ss2 <- lm((result$residuals[TT:TTT]^2) ~ t, weights = 1-w)
if (ss2$coefficients[1]<0) {ss2<-
lm((result$residuals[TT:TTT]^2) ~ t-1, weights = 1-w)
ss2$coefficients[1]<-0
ss2$coefficients[2]<-ss2$coefficients[1]
ss2$fitted.values<-ss2$fitted.values}
if (min(ss2$fitted.values)<0)
{ss2$coefficients[1]<-mean(result$residuals[TT:TTT]^2)
ss2$coefficients[2]<-0
ss2$fitted.values<-ss2$coefficients[1]}
p1<-dnorm(y,result$fitted.values[1:T],(ss1$fitted.values)^0.5)
p2<-dnorm(y,result$fitted.values[TT:TTT],(ss2$fitted.values)^0.5)
w <- p1/(p1+p2)
mw<-mean(w)
coef.new<-c(mw,result$coefficients[1:2],ss1$coefficients,
result$coefficients[3:4],ss2$coefficients)
temp.a <- sum(abs(coef.new-coef.last))
if (j>500) temp.a<-0
if (temp.a<tol) {conv<-1
it.conv<-j}
coef.last<-coef.new
j<-j+1
}
p_loop[s,] <- coef.new
l_loop[s] <- fun.unr(coef.new)
}
coef.new <- p_loop[which.max(l_loop),]
fit1<-coef.new[2]+coef.new[3]*t
fit2<-coef.new[6]+coef.new[7]*t
start<-as.numeric(coef.new)
start2 <- c(pmin(1, start[1]), start[2:9])
ll<-rep(99999999,100)
trial <- matrix(0, 100, 9)
for (k in 1:100){
penalty<-k/2
result.opt<-optim(start2,fun)
coef.opt<-result.opt$par

```

```

ll[k]<-fun.unr(coef.opt)
if (coef.opt[1]>1) ll[k]<-9999999999 #avoid lambda > 1
if (coef.opt[1]<0) ll[k]<-9999999999 #avoid lambda < 0
if (coef.opt[2]<0) ll[k]<-9999999999 #avoid a1 < 0
if (coef.opt[3]<0) ll[k]<-9999999999 #avoid a1 < 0
if (coef.opt[4]<0) ll[k]<-9999999999 #avoid g1 < 0
if (coef.opt[6]<0) ll[k]<-9999999999 #avoid a2 < 0
if (coef.opt[8]<0) ll[k]<-9999999999 #avoid g2 < 0
trial[k,] <- result.opt$par
}
penalty<-which.min(ll)
coef.opt <- trial[penalty,]
ll[which(ll == "NaN")] <- 9999999999
if(min(ll)>fun.unr(coef.new)){coef.opt <- coef.new}
if(min(ll)>5000){coef.opt <- coef.new}
opt1<-coef.opt[2]+coef.opt[3]*t
opt2<-coef.opt[6]+coef.opt[7]*t
plot(t,y)
lines(t,fit1,col="red")
lines(t,fit2,col="red")
lines(t,opt1,col="blue")
lines(t,opt2,col="blue")
fun(coef.new)
fun(coef.opt)
fun.unr(coef.new)
fun.unr(coef.opt)
#log-likelihood
p_final_penalty_cv [i,] <- c(coef.opt[1], coef.opt[2], coef.opt[3],
coef.opt[4], coef.opt[5], coef.opt[6],
coef.opt[7], coef.opt[8], coef.opt[9])
l_final_penalty_cv [i] <- -(fun.unr(coef.opt))
}
#make sure the blue line is the upper trend
uu <- matrix (0, n_municipal, 1)
for (i in 1:n_municipal){
if(mean(p_final_penalty_cv[i,2]+ p_final_penalty_cv[i,3]*t) >
mean(p_final_penalty_cv[i,6]+p_final_penalty_cv[i,7]*t)){uu[i]<-1}
}
mm<- which(uu==1)
#switch back the parameters
for (i in mm){
store <- p_final_penalty_cv[i,]
p_final_penalty_cv[i,1] <- 1- p_final_penalty_cv[i,1]
p_final_penalty_cv[i,2] <- p_final_penalty_cv[i,6]
p_final_penalty_cv[i,3] <- p_final_penalty_cv[i,7]

```

```

p_final_penalty_cv[i,4] <- p_final_penalty_cv[i,8]
p_final_penalty_cv[i,5] <- p_final_penalty_cv[i,9]
p_final_penalty_cv[i,6] <- store[2]
p_final_penalty_cv[i,7] <- store[3]
p_final_penalty_cv[i,8] <- store[4]
p_final_penalty_cv[i,9] <- store[5]
}
#Find out crossing trends and negative variance in 2028 and restrict
oo <- matrix(0, n_municipal, 2)
#column 1 for crossing , column 2 for negative var
#year2028-year1949
YY <- 79
#crossing
for (i in 1:n_municipal){
  if((p_final_penalty_cv[i,2]>p_final_penalty_cv[i,6])
  |
  ((p_final_penalty_cv[i,2]+p_final_penalty_cv[i,3]*T)>
  (p_final_penalty_cv[i,6]+p_final_penalty_cv[i,7]*T)))
  {oo[i,1]<-1}
}
#negative var
for (i in 1:n_municipal){
  if(((p_final_penalty_cv[i,4]+ p_final_penalty_cv[i,5]*YY)<0)
  |
  ((p_final_penalty_cv[i,8]+ p_final_penalty_cv[i,9]*YY)<0))
  {oo[i,2]<-1}
}
aa <- which (oo[,1] == 0 & oo[,2] == 0) #no need restriction
bb <- which (oo[,1] == 1 & oo[,2] == 0) #crossing , good var
cc <- which (oo[,1] == 0 & oo[,2] == 1) #not crossing , bad var
dd <- which (oo[,1] == 1 & oo[,2] == 1) #crossing and bad var
ee <- which (oo[,1] == 1) #crossing
ff <- which (oo[,2] == 1) #bad var
#EM and Penalized Restricted , 2-component & Changing variances
p_final_restrict_penalty_cv <- matrix(0, n_municipal, 9)
l_final_restrict_penalty_cv <- matrix(0, n_municipal, 1)
for(i in aa){
  p_final_restrict_penalty_cv[i,] <- p_final_penalty_cv[i,]
  l_final_restrict_penalty_cv[i] <- l_final_penalty_cv[i]
}
p_final_restrict_penalty8_cv <- matrix(0, n_municipal, 9)
l_final_restrict_penalty8_cv <- matrix(0, n_municipal, 1)
#Set counties that cross to start from the same intercept
for(i in sort(c(ee, ff))){
  #penalized log-likelihood

```

```

fun<-function(start) {-(sum(log(start[1]*dnorm(y,start[2]
+start[3]*t,(start[4]+start[5]*t)^0.5)+(1-start[1])*
dnorm(y,start[2]+start[6]*t,
(start[7]+start[8]*t)^0.5)))-penalty*(start[1]-0)^2)}
#log-likelihood
fun.unr<-function(start) {-(sum(log(start[1]*dnorm(y,start[2]
+start[3]*t,(start[4]+start[5]*t)^0.5)+(1-start[1])
*dnorm(y,start[2]+start[6]*t,(start[7]+start[8]*t)^0.5))))}
y<-crop[,i+1]
tol<-0.001
conv<-0
it.conv<-0
T<-length(y)
t<-seq(1:T)
p_loop <- matrix(0, length(starting), 8)
l_loop <- matrix(0, length(starting), 1)
for(s in 1:length(starting)){
result<-rq(y~t,starting[s])
w<-rep(0,T)
w[result$residuals < 0]<-1
yy<-c(y,y)
x1t<-c(t,rep(0,T))
x2t<-c(rep(0,T),t)
x1<-c(rep(1,T),rep(1,T))
yy<-c(y,y)
ww<-c(w,1-w)
XXR<-cbind(x1,x1t,x2t)
conv<-0
coef.last<-rep(0,8)
temp.a<-10000
j<-1
TT<-T+1
TTT<-2*T
while (conv < 1){
ww<-c(w,1-w)
result<-lm(yy~XXR-1,weights=ww)
ss1 <- lm((result$residuals[1:T]^2) ~ t, weights = w)
if (ss1$coefficients[1]<0) {sss1<-
lm((result$residuals[1:T]^2) ~ t-1, weights = w)
ss1$coefficients[1]<-0
ss1$coefficients[2]<-sss1$coefficients[1]
ss1$fitted.values<-sss1$fitted.values}
if ((min(ss1$fitted.values)<0)|((ss1$coefficients[1]+
ss1$coefficients[2]*YY<0))
{ss1$coefficients[1]<-mean(result$residuals[1:T]^2)}

```

```

ss1$coefficients[2]<-0
ss1$fitted.values<-ss1$coefficients[1]}
ss2 <- lm((result$residuals[TT:TTT]^2) ~ t, weights = 1-w)
if (ss2$coefficients[1]<0) {sss2<-
lm((result$residuals[TT:TTT]^2) ~ t-1, weights = 1-w)
ss2$coefficients[1]<-0
ss2$coefficients[2]<-sss2$coefficients[1]
ss2$fitted.values<-sss2$fitted.values}
if ((min(ss2$fitted.values)<0)|(ss2$coefficients[1]+
ss2$coefficients[2]*YY<0))
{ss2$coefficients[1]<-mean(result$residuals[TT:TTT]^2)
ss2$coefficients[2]<-0
ss2$fitted.values<-ss2$coefficients[1]}
mw<-mean(w)
p1<-dnorm(y,result$fitted.values[1:T],(ss1$fitted.values)^0.5)
p2<-dnorm(y,result$fitted.values[TT:TTT],(ss2$fitted.values)^0.5)
w <- p1/(p1+p2)
coef.new<-c(mw,result$coefficients[1:2],ss1$coefficients,
result$coefficients[3],ss2$coefficients)
temp.a <- sum(abs(coef.new-coef.last))
if (j>500) temp.a<-0
if (temp.a<tol) {conv<-1
it.conv<-j}
coef.last<-coef.new
j<-j+1
}
p_loop[s,] <- coef.new
l_loop[s] <- fun.unr(coef.new)
}
coef.new <- p_loop [which.max(l_loop),]
fit1<-coef.new[2]+coef.new[3]*t
fit2<-coef.new[2]+coef.new[6]*t
start<-as.numeric(coef.new)
start2 <- c(pmin(1, start[1]), start[2:8])
ll<-rep(99999999,100)
trial <- matrix(0, 100, 8)
for (k in 1:100){
penalty<-k
result.opt<-optim(start2,fun)
coef.opt<-result.opt$par
ll[k]<-fun.unr(coef.opt)
if (coef.opt[1]>1) ll[k]<-9999999999 #avoid lambda is > 1
if (coef.opt[1]<0) ll[k]<-9999999999 #avoid lambda is < 0
if (coef.opt[2]<0) ll[k]<-9999999999 #avoid a1 is < 0
if (coef.opt[3]<0) ll[k]<-9999999999 #avoid b1 is < 0

```

```

if (coef.opt[4]<0) ll[k]<-9999999999 #avoid g1 is < 0
if (coef.opt[7]<0) ll[k]<-9999999999 #avoid g2 is < 0
if (((coef.opt[4]+coef.opt[5]*YY)<0)|
((coef.opt[7]+coef.opt[8]*YY)<0)){ll[k]<-9999999999}
trial[k,] <- result.opt$par
}
penalty<-which.min(ll)
coef.opt <- trial[penalty,]
if(min(ll)>fun.unr(coef.new)){coef.opt <- coef.new}
opt1<-coef.opt[2]+coef.opt[3]*t
opt2<-coef.opt[2]+coef.opt[6]*t
plot(t,y)
lines(t,fit1,col="red")
lines(t,fit2,col="red")
lines(t,opt1,col="blue")
lines(t,opt2,col="blue")
fun(coef.new)
fun(coef.opt)
fun.unr(coef.new)
fun.unr(coef.opt)
p_final_restrict_penalty8_cv [i,] <- c(coef.opt[1], coef.opt[2],
coef.opt[3], coef.opt[4], coef.opt[5], coef.opt[2], coef.opt[6],
coef.opt[7], coef.opt[8])
l_final_restrict_penalty8_cv [i] <- -(fun.unr(coef.opt))
}
#Set counties that cross to not cross
p_final_restrict_penalty9_cv <- matrix (0, n_municipal, 9)
l_final_restrict_penalty9_cv <- matrix (0, n_municipal, 1)
starting <- c(.2,.35,.5,.65,.8)
for(i in sort(c(ee, ff))){
y <- crop[,i+1]
#penalized log-likelihood
fun<-function(start){-(sum(log(start[1]*dnorm(y,start[2]+
start[3]*t,(start[4]+start[5]*t)^0.5)+(1-start[1])*
dnorm(y,start[6]+start[7]*t,(start[8]+start[9]*t)^0.5)))
-penalty*(start[1]-0)^2)}
#log-likelihood
fun.unr<-function(start){-(sum(log(start[1]*
dnorm(y,start[2]+start[3]*t,(start[4]+start[5]*t)^0.5)+
(1-start[1])*dnorm(y,start[6]+start[7]*t,
(start[8]+start[9]*t)^0.5))))}
tol<-0.001
conv<-0
it.conv<-0
T<-length(y)

```

```

t<-seq(1:T)
p_loop <- matrix (0, length(starting), 9)
l_loop <- matrix (0, length(starting), 1)
for(s in 1:length(starting)){
result<-rq(y~t,starting[s])
w<-rep(0,T)
w[result$residuals <0]<-1
yy<-c(y,y)
x1t<-c(t,rep(0,T))
x2t<-c(rep(0,T),t)
x1<-c(rep(1,T),rep(0,T))
x2<-c(rep(0,T),rep(1,T))
yy<-c(y,y)
ww<-c(w,1-w)
XXR<-cbind(x1,x1t,x2,x2t)
coef.last<-rep(0,9)
temp.a<-10000
j<-1
TT<-T+1
TTT<-2*T
while (conv < 1){
ww<-c(w,1-w)
result<-lm(yy~XXR-1,weights=ww)
ss1 <- lm((result$residuals[1:T]^2) ~ t, weights = w)
if (ss1$coefficients[1]<0) {sss1<-
lm((result$residuals[1:T]^2) ~ t-1, weights = w)
ss1$coefficients[1]<-0
ss1$coefficients[2]<-sss1$coefficients[1]
ss1$fitted.values<-sss1$fitted.values}
if (min(ss1$fitted.values)<0)
{ss1$coefficients[1]<-mean(result$residuals[1:T]^2)
ss1$coefficients[2]<-0
ss1$fitted.values<-ss1$coefficients[1]}
ss2 <- lm((result$residuals[TT:TTT]^2) ~ t, weights = 1-w)
if (ss2$coefficients[1]<0) {sss2<-
lm((result$residuals[TT:TTT]^2) ~ t-1, weights = 1-w)
ss2$coefficients[1]<-0
ss2$coefficients[2]<-sss2$coefficients[1]
ss2$fitted.values<-sss2$fitted.values}
if (min(ss2$fitted.values)<0)
{ss2$coefficients[1]<-mean(result$residuals[TT:TTT]^2)
ss2$coefficients[2]<-0
ss2$fitted.values<-ss2$coefficients[1]}
p1<-dnorm(y,result$fitted.values[1:T],(ss1$fitted.values)^0.5)
p2<-dnorm(y,result$fitted.values[TT:TTT],(ss2$fitted.values)^0.5)

```

```

w <- p1/(p1+p2)
mw<-mean(w)
coef.new<-c(mw,result$coefficients[1:2],ss1$coefficients,
result$coefficients[3:4],ss2$coefficients)
temp.a <- sum(abs(coef.new-coef.last))
if (j>500) temp.a<-0
if (temp.a<tol) {conv<-1
it.conv<-j}
coef.last<-coef.new
j<-j+1
}
p_loop[s,] <- coef.new
l_loop[s] <- fun.unr(coef.new)
}
coef.new <- p_loop [which.max(l_loop),]
fit1<-coef.new[2]+coef.new[3]*t
fit2<-coef.new[6]+coef.new[7]*t
start<-as.numeric(coef.new)
start2 <- c(pmin(1, start[1]), start[2:9])
ll<-rep(99999999,100)
trial <- matrix(0, 100, 9)
for (k in 1:100){
penalty<-k/2
result.opt<-optim(start2,fun)
coef.opt<-result.opt$par
ll[k]<-fun.unr(coef.opt)
if (coef.opt[1]>1) ll[k]<-9999999999 #avoid lambda > 1
if (coef.opt[1]<0) ll[k]<-9999999999 #avoid lambda < 0
if (coef.opt[2]<0) ll[k]<-9999999999 #avoid a1 < 0
if (coef.opt[3]<0) ll[k]<-9999999999 #avoid a1 < 0
if (coef.opt[4]<0) ll[k]<-9999999999 #avoid g1 < 0
if (coef.opt[6]<0) ll[k]<-9999999999 #avoid a2 < 0
if (coef.opt[8]<0) ll[k]<-9999999999 #avoid g2 < 0
if (((coef.opt[4]+coef.opt[5]*YY)<0)|
((coef.opt[7]+coef.opt[8]*YY)<0)){ ll[k]<-9999999999}
if ((mean(coef.opt[2]+coef.opt[3]*t) <
mean(coef.opt[6]+coef.opt[7]*t))&
((coef.opt[2]+coef.opt[3]*T) >
(coef.opt[6]+coef.opt[7]*T))){ ll[k]<-9999999999}
if ((mean(coef.opt[2]+coef.opt[3]*t) <
mean(coef.opt[6]+coef.opt[7]*t))&
(coef.opt[2]>coef.opt[6])){ ll[k]<-9999999999}
if ((mean(coef.opt[2]+coef.opt[3]*t) > mean(coef.opt[6]
+coef.opt[7]*t))&((coef.opt[2]+coef.opt[3]*T) <
(coef.opt[6]+coef.opt[7]*T))){ ll[k]<-9999999999}

```

```

if((mean(coef.opt[2]+ coef.opt[3]*t) > mean(coef.opt[6]
+coef.opt[7]*t))&coef.opt[2]<coef.opt[6])){ ll[k]<-9999999999}
trial[k,] <- result.opt$par
}
penalty<-which.min(ll)
coef.opt <- trial[penalty,]
ll[which(ll == "NaN")] <- 9999999999
if(min(ll)>5000) {coef.opt <- c(0,0,0,0,0,0,0)}
if(
(fun.unr(coef.new)<fun.unr(coef.opt))
&
(coef.new[2]<coef.new[6])
&
((coef.new[2]+coef.new[3]*T) < (coef.new[6]+coef.new[7]*T))
){coef.opt <- coef.new}
opt1<-coef.opt[2]+coef.opt[3]*t
opt2<-coef.opt[6]+coef.opt[7]*t
plot(t,y)
lines(t,fit1,col="red")
lines(t,fit2,col="red")
lines(t,opt1,col="blue")
lines(t,opt2,col="blue")
fun(coef.new)
fun(coef.opt)
fun.unr(coef.new)
fun.unr(coef.opt)
#log-likelihood
p_final_restrict_penalty9_cv [i,] <- c(coef.opt[1], coef.opt[2],
coef.opt[3], coef.opt[4], coef.opt[5], coef.opt[6], coef.opt[7],
coef.opt[8], coef.opt[9])
l_final_restrict_penalty9_cv [i] <- -(fun.unr(coef.opt))
}
l_final_restrict_penalty8_cv
[which(is.na(l_final_restrict_penalty8_cv))] <- -999999
l_final_restrict_penalty9_cv
[which(is.na(l_final_restrict_penalty9_cv))] <- -999999
#choose between same intercept and different intercept
for(i in sort(c(ee, ff))){
if (l_final_restrict_penalty8_cv[i] == "0")
{l_final_restrict_penalty8_cv[i]<--999999}
if (l_final_restrict_penalty9_cv[i] == "0")
{l_final_restrict_penalty9_cv[i]<--999999}
}
cbind(l_final_restrict_penalty8_cv, l_final_restrict_penalty9_cv)
for(i in sort(c(ee, ff))){

```

```

if(l_final_restrict_penalty8_cv[i] >
l_final_restrict_penalty9_cv[i]){
p_final_restrict_penalty_cv[i,] <-
p_final_restrict_penalty8_cv[i,]
l_final_restrict_penalty_cv[i] <-
l_final_restrict_penalty8_cv[i]}
if(l_final_restrict_penalty8_cv[i] <
l_final_restrict_penalty9_cv[i]){
p_final_restrict_penalty_cv[i,] <-
p_final_restrict_penalty9_cv[i,]
l_final_restrict_penalty_cv[i] <-
l_final_restrict_penalty9_cv[i]}
}
##Final check on crossing trends and restrict##
#make sure the blue line is the upper trend
uu <- matrix (0, n_municipal, 1)
for (i in 1:n_municipal){
uu[i] <- (mean(p_final_restrict_penalty_cv[i,2]+
p_final_restrict_penalty_cv[i,3]*t) >
mean(p_final_restrict_penalty_cv[i,6]+
p_final_restrict_penalty_cv[i,7]*t))
}
ss<- which(uu==1)
#switch back the parameters
for (i in ss){
store <- p_final_restrict_penalty_cv[i,]
p_final_restrict_penalty_cv[i,1] <-
1-p_final_restrict_penalty_cv[i,1]
p_final_restrict_penalty_cv[i,2] <- p_final_restrict_penalty_cv[i,6]
p_final_restrict_penalty_cv[i,3] <- p_final_restrict_penalty_cv[i,7]
p_final_restrict_penalty_cv[i,4] <- p_final_restrict_penalty_cv[i,8]
p_final_restrict_penalty_cv[i,5] <- p_final_restrict_penalty_cv[i,9]
p_final_restrict_penalty_cv[i,6] <- store[2]
p_final_restrict_penalty_cv[i,7] <- store[3]
p_final_restrict_penalty_cv[i,8] <- store[4]
p_final_restrict_penalty_cv[i,9] <- store[5]
}
#final check if there's error
oo <- matrix (0, n_municipal, 1)
for (i in 1:n_municipal){
if(p_final_restrict_penalty_cv[i,1]==0){oo[i]<-1}
if(l_final_restrict_penalty_cv[i]==0){oo[i]<-1}
}
nn<- which(oo==1)

```

Expected Yield Loss

```
#Based on Model4R
parm_r <- read.csv("Parameters.csv")

#from year 1949 to 2028
n <- 80
t1 <- seq(1,n)
prem90_penalty_cv <- matrix(0, length(t1), nrow(parm_r))
for(k in 1:nrow(parm_r)){
  for(i in 1:length(t1)){
    lambda_penalty_cv <- parm_r[k,1]
    a1_penalty_cv <- parm_r[k,2]
    b1_penalty_cv <- parm_r[k,3]
    s1_penalty_cv <- parm_r[k,4]
    a2_penalty_cv <- parm_r[k,5]
    b2_penalty_cv <- parm_r[k,6]
    s2_penalty_cv <- parm_r[k,7]
    mu1_penalty_cv <- a1_penalty_cv + b1_penalty_cv * t1[i]
    mu2_penalty_cv <- a2_penalty_cv + b2_penalty_cv * t1[i]
    gaur90_penalty_cv <- (lambda_penalty_cv * mu1_penalty_cv +
      (1 - lambda_penalty_cv) * mu2_penalty_cv) * .9
    grid <- seq(0,300,.1)
    pr_penalty_cv <- lambda_penalty_cv * dnorm(grid, mu1_penalty_cv,
      s1_penalty_cv ^ 0.5) + (1 - lambda_penalty_cv) *
      dnorm(grid, mu2_penalty_cv, s2_penalty_cv ^ 0.5)
    prem90_penalty_cv[i,k] <- sum(pmax(0, gaur90_penalty_cv - grid)
      * pr_penalty_cv * .1)
  }
}
prem90_penalty_cv_2018 <- matrix(0, nrow(parm_r), 1)
prem90_penalty_cv_2023 <- matrix(0, nrow(parm_r), 1)
prem90_penalty_cv_2028 <- matrix(0, nrow(parm_r), 1)
for(k in 1:nrow(parm_r)){
  prem90_penalty_cv_2018[k] <- prem90_penalty_cv[n-10,k]
  prem90_penalty_cv_2023[k] <- prem90_penalty_cv[n-5,k]
  prem90_penalty_cv_2028[k] <- prem90_penalty_cv[n,k]
}
re <- prem90_penalty_cv_2023 / prem90_penalty_cv_2018
ms <- prem90_penalty_cv_2028 / prem90_penalty_cv_2018
```