Empirical Studies of Corn Yield Distribution Modeling

by

Yun Long

A research paper
presented to the University of Waterloo
in partial fulfillment of the
requirement for the degree of
Master of Mathematics
in
Computational Mathematics

Supervisor: Prof. Ken Seng Tan

Waterloo, Ontario, Canada, 2015

I hereby declare that I am the sole author of this report. This is a true copy of the report, including any required final revisions, as accepted by my examiners.

I understand that my report may be made electronically available to the public.

Abstract

This research paper uses partially historical corn yields of each county over five states as training set for calibrating alternative yield models including single parametric distributions, nonparametric distributions, semiparametric distributions, mixing models, and other nontraditional models. The calibrated models are applied to predict the yield of the most recent eight years in historical yield data. The performance of alternative yield models is evaluated by the Bayesian information criterion (BIC) in the in-sample analysis and mean square error (MSE) in the out-of-sample analysis. From the in-sample and out-of-sample cases, this paper suggests that a single parametric distribution Weibull is more revealing and efficient to characterize corn yield distribution when training data set is small. The results are presented for the crop corn at the county levels.

Acknowledgements

I would like to thank my supervisor, Professor Ken Seng Tan for his guidance, assistance, and financial support through the past year. I also want to thank Professor Lysa Porth and Professor Chengguo Weng who support me to make this possible with their valuable advices. last but not least, I would like to thank all of my colleagues, faculty, and staff in the Computational Mathematics Master Program to make my master study be well-spent.

Dedication

This is dedicated to all my families, friends, and whoever reads this paper.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Agriculture producers may purchase insurance products to protect their potentially substantial loss in production and income due to weather variation, while insurance companies will manage producers' production risk to maximize their reward through allocating producers among funds for reinsurance. Hence, private insurance companies would like to share underwriting risks on their liability for producers with the Federal government. On the other hand, the government shares risks with companies by Standard Reinsurance Agreement (SRA) which allows companies to decrease their risk exposure through ceding some liability to the Federal Crop Insurance Corporation (FCIC) and selectively allocating producers' policies among several reinsurance funds including Assigned Risk Fund and Commercial Fund. However, a decision for optimal crop reinsurance selection in the fund is made by producers-level risk measurement such as loss ratio (indemnity / liability) and coefficient of variation, because value of the risk measurement means whether the corresponding policies are high profit or low profit. Then companies will allocate producers or policies to appropriate funds. A key factor of the estimated ratio (LR or CV) is to forecast farm-level crop yield correctly.

Best forecasting farm-level crop yield has been long debated in the agriculture economics literature. However, directly modeling and calibrating farm-level crop yield distribution are very difficult since historical farm-level crop yield data are much less than 20 years generally. Thus, it results in hardly calibrating yield probability density function (PDF). Modeling and forecasting county-level crop yield is commonly used as first step for analyzing the shape of farm-level yield data, because historical county-level yield data are more than 40 years which is much better than farm-level yield and lower volatility of yield data at the county. County-level yield is normalized and taken by averaging all farm-level yields within the county. Also, many studies found a relationship between county-level crop yield and farm-level crop yield which is avoiding calibrating individual farm's yield directly. Therefore, improving modeling of county-level yield data is equivalently significant in the study of agriculture risk modeling.

In this essay, the method is proposed to extensively evaluate alternative crop yield distributions for forecasting county-level yield through the in-sample analysis and the out-of-sample analysis. The parametric distributions such as Gamma (Gallagher 1987), Weibull (Lanoue et al.2010;Sherrick et al.2004), Beta (Nelson and Preckel 1989;Sherrick et al.2004),normal(Just and Weninger 1999), and mixture of normals (Goodwin, Roberts, and Vedenov 2007) has been applied comprehensively in modeling in the crop yield-related data. In comparison with these aforementioned models, some nontraditional models in agriculture crop modeling such as time series method (Gaussian process), mixture of parametric and nonparametric models (Weibull and Kernel density estimator), and Erlang mixture distributions (Porth, Zhu, and Tan 2014) are introduced. Thus, the essay is organized into 6 chapters as follows: chapter 2 discusses some typical candidate yield models; chapter 3 presents methods for detrending yield data; chapter 4 describes data set used in this essay; chapter 5 evaluates performance of alternative models through in-sample analysis and out-of-sample analysis; chapter 6 concludes with the major results.

Chapter 2

Candidate Yield Models

Over many years, various approaches are used to calibrate crop yield distributions, which can be classified into three primary groups: parametric, nonparametric, and semi-parametric. On one hand, normal distribution has been discussed and supported by Just and Weinger (1999), but other models have also been analyzed against normality in the literature such as Weibull, Beta, Gamma, and Kernel methods. In this essay, candidate yield models for calibrating historical yield distribution and forecasting yield can be segmented into 6 groups: parametric (Normal, Beta, Weibull, and Gamma), semi-parametric distribution (mixture of Erlang distributions, mixture of two Gamma distributions, and mixture of two Gaussian distributions), nonparametric (kernel density estimator), mixture of nonparametric and parametric (mixture of Weibull and kernel density estimator), stochastic process (Gaussian process), and random draw from historical yield catalog.

2.1 Normal Distribution

Parameters: $\sigma > 0, -\infty < \mu < \infty$
The probability density function is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma}}, -\infty < x < \infty$$

In the study of yield distributions by Just and Weniger (1999), the normal distribution is a reasonable empirical distribution for studying crop insurance programs.The normal distribution is symmetric, bell-shaped and unbounded below by zero.

2.2 Beta Distribution

Parameters: $\alpha, \beta > 0$
The probability density function is

$$f(x) = \frac{(x-a)^{\alpha-1}(b-x)^{\beta-1}}{B(\alpha,\beta)(b-a)^{\alpha+\beta-1}}, a \leq x \leq b$$

where $B(\alpha, \beta)$ is beta function.
Beta has been used as a prior distribution by Nelson and Prekel (1989); and Triupattur, Hauser, and Chaherli (1996) since it has the advantage of being flexible with several different shapes depending on the values of the parameters (α, β). The upper and lower bounds can be ether specified or estimated in yield model fitting.

2.3 Weibull Distribution

Parameters: $\alpha, \beta > 0$
The probability density function is

$$f(x) = \frac{ax^{\alpha-1}}{\beta^{\alpha}} \exp[-(x/\beta)^{\alpha}], x \geq 0$$

Weibull becomes prevalent as evaluation by Lanoue et al. (2010), Sherrick et al.(2004) in the literature since it has more desirable properties such as flexibility, being bounded by zero, and allowing for a wide range of skewness and kurtosis.

2.4 Gamma Distribution

Parameters: $k, \theta > 0$
The probability density function is

$$f(x) = \frac{x^{k-1}e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)}, x > 0$$

Gamma was used by Gallagher (1987) to present the evidence of negatively skewed soybean yields with an upper limit and a high chance of occasional low yields.

2.5 Mixture of Erlang Distributions

Parameters: $\theta, \alpha > 0$, and $r_1 < r_2 < ... < r_M$ are integers
The probability density function is

$$f(x) = \sum_{i=1}^{M} \alpha_i \frac{x^{r_i - 1} e^{-\frac{x}{\theta}}}{\theta^{r_i}(r_i - 1)!}, x > 0$$

Erlang mixture distribution has been studied by Porth, Zhu, and Tan (2014) in modeling agriculture crop data. They also proved that the distribution could capture the heavy tails of the data more accurately.

2.6 Mixture of Two Normal Distributions

Parameters: $p > 0, \sigma_i > 0, -\infty < \mu_i < \infty$, i=1,2
The probability density function is

$$f(x) = pg(x, u_1, \sigma_1) + (1 - p)g(x, u_2, \sigma_2)$$

where $g(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma}}, -\infty < x < \infty$

2.7 Mixture of Two Gamma Distributions

Parameters: $p, k_i, \theta_i > 0$
The probability density function is

$$f(x) = pg(x, k_1, \theta_1) + (1 - p)g(x, k_2, \theta_2)$$

where $g(x, k, \theta) = \frac{x^{k-1} e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)}, x > 0$
The aforementioned mixtures of distributions are examined for unknown yield distribution in this essay because that evaluation could be difficult when data from a different source are modeled by one of a single distribution. In the studies of Ker and Goodwin, it has shown that crop yield could be bimodal and negatively skewed due to the effects of catastrophic events such as flood, freeze and drought. Hence, a multimodal distribution might be necessary.

2.8 Kernel Density Estimator

The form is

$$f(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - x_i}{h})$$

where K $(\cdot)$ is the Kernel with satisfying the condition: $\int_{-\infty}^{\infty} K(z)dz = 1$, $h > 0$ is a smoothing parameter called the bandwidth, x_i is the observation, and n is the number of observations. The Kernel estimator was used by Goodwin and Ker (1998), Turvey and Zhao (1999), and others to estimate the shape of the conditional yield density and price crop insurance contracts. The main reason is that in the nonparametric analysis, the characteristics of the distribution can be shown such as skewness and bimodality. Furthermore, there is no any prior assumption on the shape of the density for this method.

2.9 Mixture of Weibull and Kernel Density Estimator

Parameters: $p, \alpha, \beta > 0$
The probability density function is

$$f(x) = p\frac{ax^{\alpha-1}}{\beta^{\alpha}}\exp[-(x/\beta)^{\alpha}] + (1-p)\frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - x_i}{h}), x \geq 0$$

The mixture is the representative of a mixture of parametric and nonparametric in this essay and this method is aimed to take both advantages of parametric nature and nonparametric techniques for fitting crop yield data.

2.10 Gaussian Process for Nonstationary Time Serious

A Gaussian process is a collection of random variables, any finite set of which has a joint Gaussian distribution. In the paper of Belhouari and Bermak (2004), Gaussian distribution has been successfully applied to nonstationary time series as a prediction approach.

As shown in Appendix A, the predictive distribution is Gaussian (Williams and Rasmussen,1996):

$$p(y^{(n+1)}|x^{(n+1)}, X, Y) = N(\mu_{y^{(n+1)}}, \sigma^2_{y^{(n+1)}})$$

where the mean and variance are given by

$$\begin{aligned}\mu_{y^{(n+1)}} &= K(x^{(n+1)}, X)(K(X, X) + \sigma_n^2 I)^{-1} Y \\ \sigma^2_{y^{(n+1)}} &= K(x^{(n+1)}, X)(K(X, X) + \sigma_n^2 I)^{-1} K(X, x^{(n+1)})\end{aligned}$$

In the above formula:

- $X = [x^{(1)}, ...x^{(n)}]$ is a finite collection of inputs such as year-digit in the case of forecasting crop yield of next year, and $y = [y^{(1)}, ...y^{(n)}]$ is the corresponding fucntion values such as the historical crop yield.
- σ_n is the unknown Gaussian noise.
- K $(\cdot)$ is the Kernel function.

2.11 Random Draw from Historical Yield Catalog

Unlike all previous statistical yield models, the detrended yield observations are applied to construct an empirical distribution of county-level crop yields (Ker and Goodwin 2000; Ker and Coble 2003) in the next year by assigning equal probabilities $1/n$, where n is number of historical observations, to each realization of the historical county-level crop yield. This method is also applied in the study of Vedenov, Miranda, Dismukes, and Glauber (2004). The detrended yield is the yield after removing the trend in the yield time series to make the historical yield observations "on a level" or representative of today's current yield. The methods of detrending crop yield are discussed in the next chapter.

Chapter 3

Stationary Yield Data

In order to model yield distributions accurately, it is necessary to compute deterministic trends and stochastic trends of crop yields over time due to the effects of improvements in farm technology, behavioral changes, and weather changes. Hence, detrending is a widely used technique to get stationary yield data in agriculture risk analysis, or in other words, to remove trend components is to make yields comparable throughout the years.

3.1 Detrending Yield Data

Common approaches for detrending yield data include deterministic trend models and stochastic trend models. However, a stochastic trend should be tested before estimating the potential deterministic trend in the crop yield series since a stochastic trend will cause spurious parameter estimates in regressing yield. Augmented Dickey Fuller (ADF), Phillips-Perron (PP), Dickey Fuller Generlized Least Squares (DF-GLS), and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests are most frequently used tests for the existence of a stochastic trend. The ADF test is employed in this essay.

ADF test: Said and Dickey (1984) discussed the basic autoregressive unit root test to accommodate general ARMA (p,q) models with unknown orders and the test is referred to as the augmented Dickey-Fuller test:

$$y_t = c + \sigma t + \varphi y_{t-1} + \sum_{i=1}^{p} \beta_i \Delta y_{i-1} + \varepsilon_t$$

where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. The ADF tests the null hypothesis

$$H_0 : \phi = 1$$

against

$$H_1 : |\phi| < 1$$

assuming that the dynamics in the data follow an ARMA structure. If the unit root hypothesis fails to be rejected, the crop yield series should be taken difference before arguing the deterministic trend in yield series.

Our understanding is that the National Agriculture Statistics Service (NASS) county yield data are detrended by a linear trend. Moreover, due to the shortness of our data set, a simple linear trend is fitted for the crop yield in this essay. The trend model is

$$y_t^{trd} = \alpha_0 + \alpha_1 t$$

where t = 1,2,...,n are periods.
The county crop yields detrended to base year (t=n) equivalents are calculated as

$$y_t^{\det} = \frac{y_t}{y_t^{trd}} y_n^{trd}, \ t = 1, ...n$$

where y_t are the observed yields and y_t^{trd} are the corresponding yield trends. In doing this, the resulting values of yields are homoskedastic by the yield in the base year and are expressed in terms of base year technology. The detrended yield observations are then used for model fitting by each candidate yield model. Additionally, assumptions of this approach are that yields are distributed independently across time and year to year, and weather does not tend to be autocorrelative in the region (Jewson and Brix 2006; Woodard and Garcia 2008).

3.2 Trend-Adjusted Actual Production History (APH) Procedures

APH is yield insurance covering yield losses from a farm or unit, which is a 4- to 10-year yield average used to calculate each producer's production guarantee. Trend-Adjusted Yield option for APH is recently supported by United States Department of Agriculture Risk Management Agency (RMA) and discussed by Goeringer (2014), Edwards (2012), Sherrick and Schnitkey (2011). Using this option, APH yields could be adjusted accounting for better farming techniques, and the adjustments include trend controls for weather and spatial considerations as well. The eligible counties for this adjustment are shown in the actuarial documents, and Trend-Adjusted APH is available for soybeans and corn. Calculating the trend-adjusted yield is that trend adjustment rate times the number of years that have passed since the yield was recorded. For example, to calculate the trend adjustment for 2004 from 2014, farmer would take: $(2014 - 2004) *$trend rate. Adding this amount to actual yield is defined as trend adjusted yield. Thus if Trend-Adjusted Actual Production history is elected, trend adjusted yields are used for APH instead of nonstationary yields.

Chapter 4

Data

County level yields for the different crops over the historical period are provided by USDA National Agricultural Statistics Service (NASS). To examine the performance of candidate yield models in empirical studies, county-level corn yields in the states Illinois, Indiana, Lowa, Missouri, and Nebraska from NASS for 1970 - 2013 were applied in this essay. Counties in each states were collected only if data of forty-four years are available, resulting in 102 counties in Illinois, 92 counties in Indiana, 99 counties in lowa, 115 counties in Missouri, and 93 counties in Nebraska. The county-level yields are normalized at the unit level (bush/acre). A plot of the historical corn yield for the county, Ogle, with 44 observations in Illinois is shown in Appendix B. The nature of corn yields is that they are increased by time because of technological improvement. To reduce the effects of technological change, the corn yield data for each selected counties in every state are required to be detrended against time by the aforementioned method. The procedure of detrending one county (Ogle) corn yields is discussed in Appendix B. Repeatedly, it has detrended historical corn yields for the rest of counties in every states for evaluating the performance of alternative yield models.

Chapter 5

Empirical Studies

The empirical studies included in-sample analysis and out-of-sample analysis on the candidate yield models in corn yields. The performance of the set of candidates was evaluated with the Bayesian information criterion (BIC) criterion for in-sample analysis and Mean square error (MSE) criterion for out-of-sample analysis.

5.1 In-Sample Analysis

The candidate yield models are comprised of Weibull, Beta, Normal, Gamma, mixtures of Erlang distributions, a mixture of two gamma, a mixture of two Gaussians, and a mixture of Weibull and Gaussian Kernel. For each of counties in five states, the corn yields from 1970 - 2005 with 36 observations are used as training data set for each of single candidate model fitting. That means for every counties, it has differently estimated parameters in the alternative yield models.

First, all parameters in models were estimated with training data set using maximum likelihood estimation (MLE) except that mixtures of Erlang distributions and a mixture of two Gaussians were fit using the expectation-maximization (EM) algorithm. For the Beta distributions fitting, the lower limits were initialized as zero and the upper limits were initialized as 120% of the maximum observation, and final parameters were solved by MLE. Next, the BIC criterion was used to rank each of the models in terms of their respective in-sample fitting performance.

BIC was developed by Gideon E.Schwarz (1978) and is defined as

$$\mathrm{BIC} = -2 \cdot \log \hat{L} + \mathrm{K} \cdot \log \mathrm{n}$$

Table 5.1: Average BIC & Frequency of Model Ranked Best over Counties in each State

Criteria	Normal	Beta	Weibull	Gamma	Erlang	MTG	MTN	MWKDE
State 1, Illinois								
Av BIC	348.24	334.73	343.78	353.11	364.75	353.10	349.33	342.22
Frequency	0.0098	0.3725	0.2157	0.0000	0.0000	0.0000	0.0882	0.3137
State 2, Indiana								
Av BIC	339.15	316.95	333.71	343.45	349.23	331.39	336.42	328.92
Frequency	0.0217	0.4674	0.1957	0.0326	0.0000	0.0217	0.0326	0.2283
State 3, Lowa								
Av BIC	352.25	306.72	347.29	343.45	368.44	355.38	355.14	343.91
Frequency	0.0309	0.4845	0.2474	0.0000	0.0000	0.0206	0.0103	0.2062
State 4, Missouri								
Av BIC	356.21	303.02	354.86	360.22	378.08	354.09	358.78	352.71
Frequency	0.1171	0.3784	0.1982	0.0180	0.0000	0.0991	0.0270	0.1622
State 5, Nebraska								
Av BIC	332.23	315.85	330.67	334.17	342.71	341.36	358.78	327.71
Frequency	0.0538	0.4086	0.1613	0.0430	0.0215	0.0108	0.0323	0.2688

Notes: MTG: mixture of two Gammas, MTN: mixture of two Normals, MWKDE: mixture of Weibull and Kernel Density Estimator, Frequency: frequency of model ranked best over counties in each State in terms of the lowest value of BIC, [0 1].

where n is the number of data points or the sample size, K is the number of free parameters to be estimated, and $\hat{L}$ is the maximized value of the likelihood function of the model. The model with the lowest BIC is preferred and it is closely related to the Akaike information criterion (AIC). The advantage of using BIC is valid for both nested models and non-nested models (Burnham and Anderson, 2002)

5.1.1 In-Sample Results for Candidate Yield Models

Table 5.1 presents average BIC values over all counties of each state for each of candidate yield models evaluated and the frequency of each candidate models ranked best among all counties in each states. Moreover, the values that frequency of model ranked best over all counties in the states could give the quantified variation of the lowest BIC in each models over all counties in each state which is more important information for evaluating performances in each candidate models. In terms of the average BIC value, the best fitting models is often Beta distribution among these five states with the values of 334.73, 316.95, 306.72, 303.02, and 315.85, followed by Weibull, MWKDE, and Normal. The mixtures of Erlang distributions have very poor performance in terms of average BIC value. The rankings in table 5.1 present similar results according to the frequency of each models ranked best over all acounties in every state. Beta, Weibull, and MWKDE performed much better than the other candidate models in these five states. The mixtures of Erlang distributions again ranked very low in terms of frequency of best fitting. Overall, based on the in-sample analysis and drastic differences in the rankings, Beta, Weibull, and MWKDE with potential outperforming other were chosen as candidate yield models for out-of-sample analysis comparing with nonparametric (Kernel density estimator), Gaussian process, and random draw from historical yield catalog.

5.2 Out-of-Sample Analysis

Combining with the optimal yield models from the in-sample analysis, the candidate yield models includes Beta, Weibull, MWKDE, Kernel density estimator, Gaussian process, and random draw from historical yield catalog in the out-of-sample analysis. For Gaussian process, this time series method used actual yields and found correlation between the observations for predicting values in future years. A Gaussian Kernel was employed for Kernel density estimator with the default (standard) bandwidth. In the out-of-sample analysis, these candidate yield models were implemented to forecast the corn yield in 2006 - 2013 orderly for each county in the states.[1]Then, the performance of each models was assessed on the mean square error (MSE) which measures the average of the square differences between the estimators and true values. Moreover, this method is more intuitive to show the predicted error. In our cases, the form of MSE is

$$\text{MSE} = \frac{1}{8}\sum_{i=1}^{8}(\hat{Y}_i - Y_i)^2$$

where $\hat{Y}_i^{\prime}s$, i = 1,2,...8, are the predicted corn yields in 2006 - 2013, and Y_i is the corresponding actual corn yield.

[1]Notes that to predict corn yield in next year, the detrended corn yields from all previous years were used for calibrating parameters in the models. For example, to predict corn yield in 2006, the detrended corn yields from 1970 - 2005 were taken for calibrating parameters in the models. Thus, the training data set is increasing as predicting more recent year.

Table 5.2: MSE & Frequency of Model Ranked Best over Counties in each State for Out-of-Sample Analysis

Criteria	Beta	Weibull	MWKDE	Kdensity	GP	RD
State 1, Illinois						
MSE	1.2576	0.6282	0.8623	0.6403	1.8689	1.0076
Frequency	0.0000	0.6765	0.1275	0.1863	0.0098	0.0000
State 2, Indiana						
MSE	1.9501	0.7417	0.8168	0.7456	1.4162	0.8157
Frequency	0.0000	0.4683	0.1848	0.1848	0.0326	0.0543
State 3, Lowa						
MSE	1.5526	0.5108	0.6229	0.6092	1.2588	0.4959
Frequency	0.0000	0.2121	0.1111	0.1717	0.0101	0.4949
State 4, Missouri						
MSE	1.3581	0.3451	0.4030	0.4434	1.2205	0.4882
Frequency	0.0000	0.4086	0.2903	0.2043	0.0108	0.0860
State 5, Nebraska						
MSE	2.0376	0.9465	1.0353	0.9629	1.3350	0.9093
Frequency	0.0000	0.3565	0.1130	0.1478	0.1217	0.2609

Notes: all MSE's is *10^3. Kdensity: Kernel density estimator with Gaussian kernel, RD: random draw from detrended historical yield catalog, Frequency: frequency of model ranked best over counties in each State in terms of the smallest value of MSE, [0 1].

5.2.1 Out-of-Sample Results for Candidate Yield Models

Table 5.2 presents the result of average MSE for each yield model over all counties in each state, and the frequency of each model ranked best as well. Interestingly, Beta was the worst performing relatively but did quite well fitting in the in-sample analysis. This difference is very prone to overfitting. On the other hand, Weibull with (0.6282,0.6765), (0.7417,0.4683), (0.5108,0.2121), (0.3451, 0.4086), and (0.9465, 0.3565) is best model fitting distribution by far in the out-of-sample analysis while its frequency is quite high relative to other models, although there is one state where RD ranked best with (0.4959, 0.4949). This superior performance of Weibull is consistent with the findings of Sherrick (2004) and Lanoue (2010), who suggested that Weibull is the better distribution for farm-level crop yield. However, the mixture of parametric and nonparametric (MWKDE) has the potential to outperform other with the more highly parameterized form. In the study of Woodard and Sherrick (2011), they found that Weibull was the component in the optimal mixture models. However, referring to the in-sample cases and out-of-sample cases of this essay, mixing of candidate yield models did not perform well comparing with single distribution as the impact of overfitting. Based on the case study of this essay, the fundamental shape of Weibull is more accommodated to the distribution of corn yield data used in this study.

Chapter 6

Conclusion

This essay aims to extensively evaluate statistical models in characterizing corn yields in every counties of five states by in-sample and out-of-sample analysis. BIC criterion is used to access the performance of candidate yield models in the in-sample cases considering the advantages of validity in non-nested models (Burnham and Anderson, 2002). For the out-of-sample analysis, MSE of each model is more intuitively to present the differences between predictive yield and true detrended yield. The candidate yield models are comprising with parametric (Normal, Beta, Weibull, and Gamma), semi-parametric distribution (mixture of Erlang distributions, mixture of two Gamma distributions, and mixture of two Gaussian distributions), nonparametric (kernel density estimator), mixture of nonparametric and parametric (mixture of Weibull and kernel density estimator), stochastic process (Gaussian process), and random draw from historical yield catalog. To our knowledge, Gaussian process and random draw are the first time to evaluate crop yield models comparing with other parametric models in out-of-sample analysis.

The results in the case study of this essay indicated that the single distribution Weibull is still more reasonable characterizing yield distribution in crop corn forecasting, which has been found by Sherrick (2004) and Lanoue (2010) for fitting farm-level yield in both corn and soybeans yield data. Recently, some candidates are supported in agricultural risk modeling in the literature. For example. the Erlang distribution is suggested by Porth, Zhu, and Tan (2014) in capturing the tails of the data more accurately. The optimal mixture models with Weibull component performed well in the study of Woodard and Sherrick (2011). However, when the data set of crop yield is short such as less than 44 observations to calibrate the value of parameters in the each model, it will be prone to overfitting for mixing models. In other words, it is not always necessary to calibrate the distribution of yield data with mix models if the single component model in the best-fitting mixtures can sufficiently present the shape of yield data. Thus, the single model Weibull is more revealing in the case study of this paper.

In next steps, the candidate yield models are needed to be investigated with the different crop data such as soybean to see whether there are similar findings. To enhance farm risk management decisions in the crop reinsurance products such as APH, further study is also focus on applying proposed approaches of yield models in risk measurements including calculating coefficient variance (CV) and loss ratio (LR) in loss data. Because CV and LR are commonly used for reinsurance companies to allocate agricultural producers in different funds by ceding liabilities and sharing of risk with the Federal Crop Insurance Corporation (FCIC). Thus, it is necessary to investigate whether the differences in the estimated CV or LR by the proposed distributions have meaningful differences in the final fund allocation.

APPENDICES

Appendix A

Derivations for Gaussian Process

A.1 Predictive Distribution

Given a data set :

$$\mathrm{D} \ = \ \{X, Y\}$$

Starting with a linear regression model for the data set:

$$\mathrm{Y} \ = \ f(X) \ + \ \varepsilon \, , \text{ where } \varepsilon \sim \ N(0, \, \sigma_n^2), \text{ and } f(X) = X^T \mathrm{W}$$

Next, use Bayesian linear regression:
Prior has pdf:

$$p(W) \ = \ N(0, \, \Sigma_p)$$

Bayes rule:

$$\begin{aligned} p(W|X, Y) \ &= \ \frac{P(Y|X, W)}{\int P(Y|X, W)p(W)dW} \\ &= \mathrm{N}(\sigma_n^{-2} A^{-1} X^T Y, A^{-1}), \end{aligned}$$

where $A \ = \ \sigma_n^{-2} X X^T \ + \ \Sigma_p^{-1}$
Hence, the predictive distribution of $(y^{(n+1)}|x^{(n+1)})$ is

$$p(y^{(n+1)}|x^{(n+1)}, X, Y) \ = \ \int p(y^{(n+1)}|x^{(n+1)}, W)p(W|X, Y)$$

$$= N(\sigma_n^{-2} (x^{(n+1)})^T A^{-1} X^T Y, \, (x^{(n+1)})^T A^{-1} x^{(n+1)})$$

Expand in A, then mean and variance can be simplified as
$\sigma_n^{-2}(x^{(n+1)})^T A^{-1} X^T Y = (x^{(n+1)})^T \Sigma_p X(X^T \Sigma_p X + \sigma_n^2 I)^{-1} Y$
$(x^{(n+1)})^T A^{-1} x^{(n+1)} = (x^{(n+1)})^T \Sigma_P x^{(n+1)} - (x^{(n+1)})^T \Sigma_P X(X^T \Sigma_p X + \sigma_n^2 I)^{-1} X^T \Sigma_p x^{(n+1)}$
Apply Kernel trick

$$X^T \Sigma_p X^T \rightarrow k(X, X^T) = \phi(X)^T \Sigma_p \phi(X^T) = \psi(X)^T \psi(X^T)$$

Then the predictive distribution of $(y^{(n+1)} | x^{(n+1)})$ is

$$N(\mu_{y^{(n+1)}}, \sigma^2_{y^{(n+1)}})$$

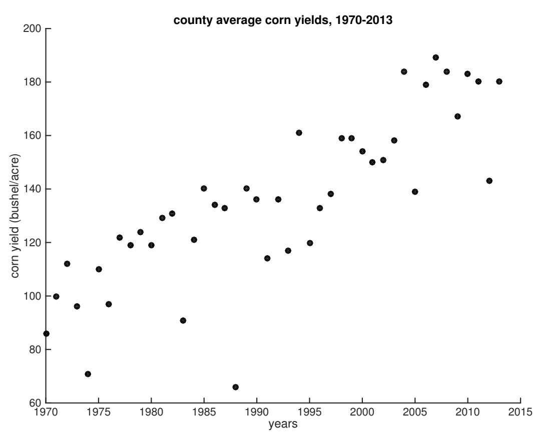
where the mean and variance are given by

$$\mu_{y^{(n+1)}} = K(x^{(n+1)}, X)(K(X,X) + \sigma_n^2 I)^{-1} Y$$
$$\sigma^2_{y^{(n+1)}} = K(x^{(n+1)}, X)(K(X,X) + \sigma_n^2 I)^{-1} K(X, x^{(n+1)})$$

Appendix B

A Case Study of Model Fitting for Corn Yield in Branch County, Ogle

B.1 Actual Corn Yields from 1970 - 2013



B.2 Trend Testing

Before accessing the performance of alternative yield models, apply detrended methods on actual county-level corn yield data to get stationary data. First, use unit root test (ADF)

Table B.1: OLS results of the two regressions

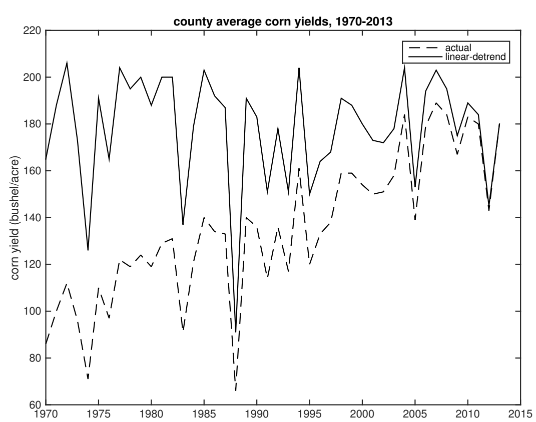
	p=2	p=0
c	126	
σ	2.7041	2.2346
φ	-0.33293	-0.13929
β_1	0.16946	
β_2	0.048868	
Test statistics	-3.9954	-7.2892
R-squared	0.66203	0.66768
Adjusted R^2	0.62447	0.65106

to see whether there is a stochastic trend in historical county yields. If there is a stochastic trend, we need to difference first (referring to section 3.1).

Results: from table B.1, rejects the null hypothesis unit root at the 5 % level for both models since $|\varphi| < 1$. That means there is a sufficient evidence to suggest that the data is trend stationary (no stochastic trend)

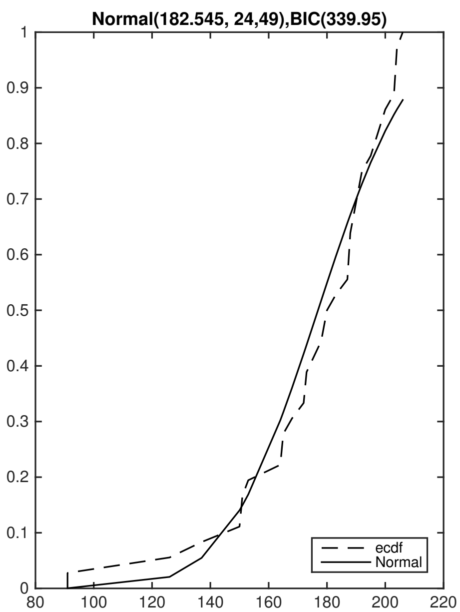
Then, detrended the corn yields by a simple linear regression (referring to section 3.1). Figure B.1 shows actual corn yields and detrended corn yields in the county from 1970 to 2013.

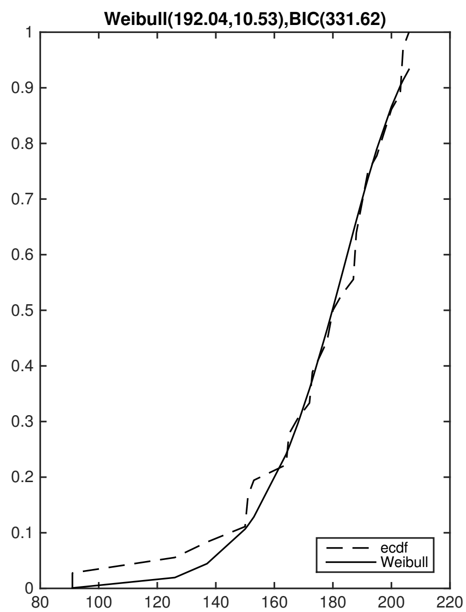
Figure B.1: Actual corn yields vs. detrended corn yields,1970-2013

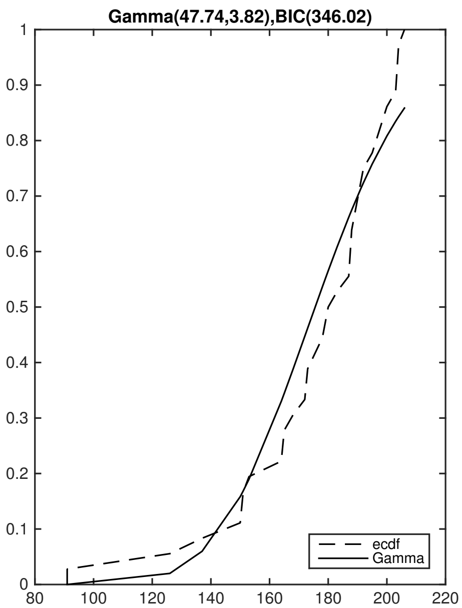


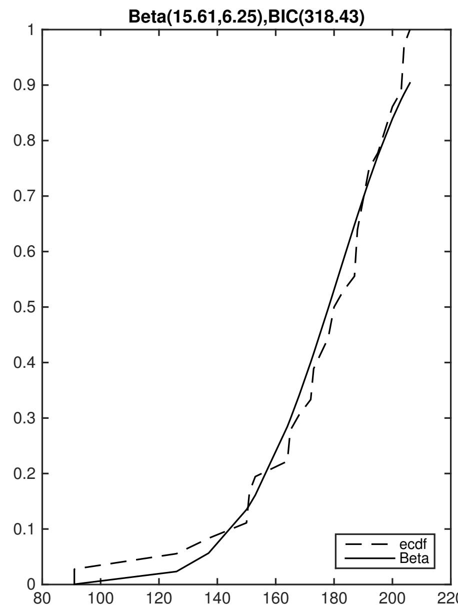
B.3 In-Sample Analysis: Empirical CDF vs.CDF Calculated by Candidate Models

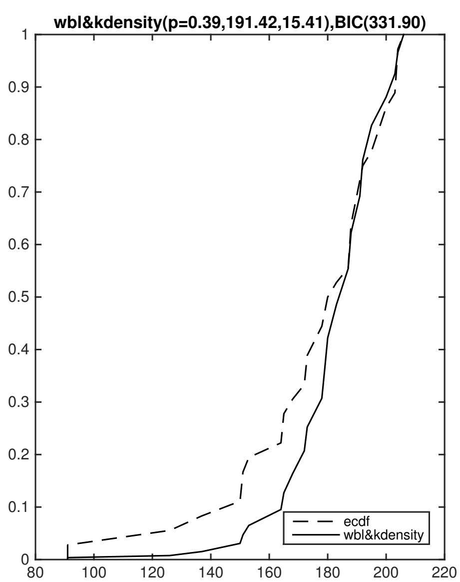
These eight figures intuitively presented the deviation of estimated cumulative distribution function (CDF) from empirical CDF. As a result, in this case, Weibull and Beta were overlapping empirical CDF much more than others. Moreover, a mixture of Weibull and kernel density estimator did well among the mixture of models.

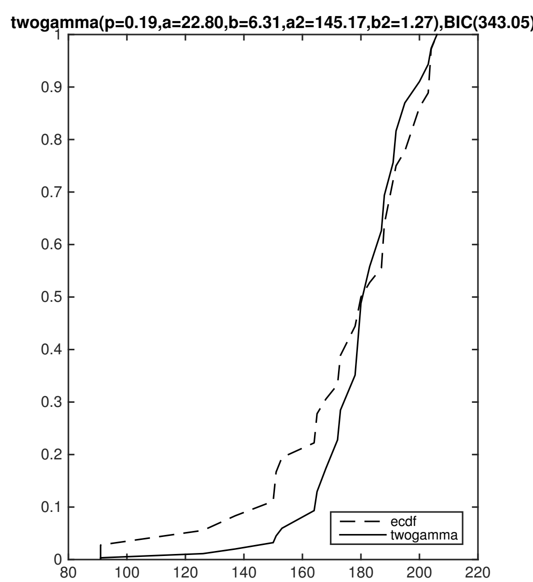


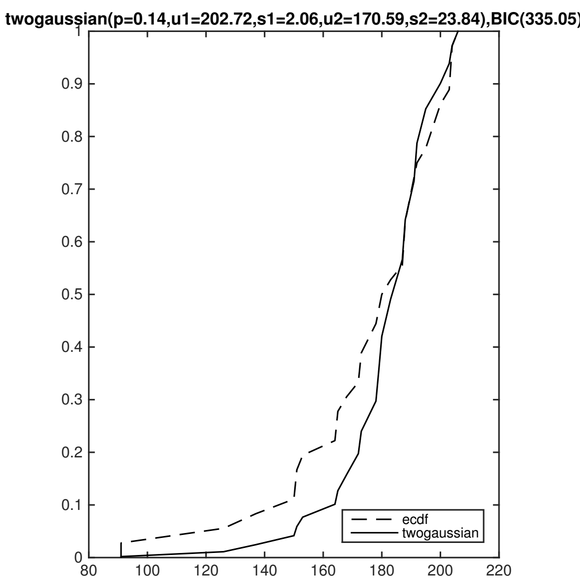


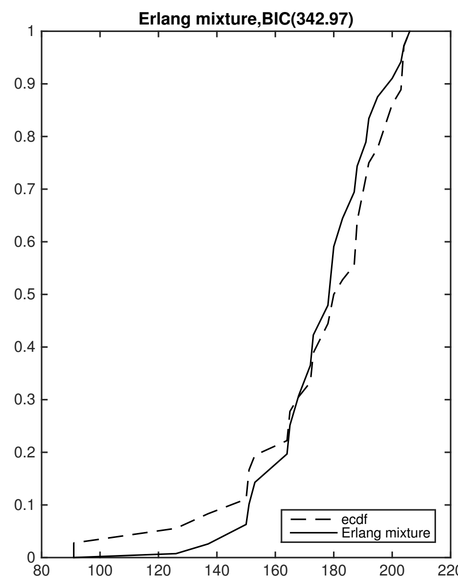

Gamma(47.74,3.82),BIC(346.02)
ecdf
Gamma


Beta(15.61,6.25),BIC(318.43)
ecdf
Beta


wbl&kdensity(p=0.39,191.42,15.41),BIC(331.90)
ecdf
wbl&kdensity


twogamma(p=0.19,a=22.80,b=6.31,a2=145.17,b2=1.27),BIC(343.05)
ecdf
twogamma


twogaussian(p=0.14,u1=202.72,s1=2.06,u2=170.59,s2=23.84),BIC(335.05)
ecdf
twogaussian


Erlang mixture,BIC(342.97)
ecdf
Erlang mixture

References

[1] Sofiane Brahim-Belhouari and Amine Bermak. Gaussian process for nonstationary time series prediction. *Computational Statistics & Data Analysis*, 47(4):705–712, 2004.

[2] Nabil M. Chaherli, Robert J. Hauser, and and Viswanath Tirupattur. Crop yield and price distributional effects on revenue hedging (undated). 1996.

[3] Robert Dismukes Joseph W. Glauber Dmitry V. Vedenov, Mario J. Miranda. Economic analysis of the standard reinsurance agreement. *Agricultural Finance Review*, 64(2):119–134, 2004.

[4] Joshua D.Woodard and Bruce J.Sherrick. Estimation of mixture models using cross-validation optimization: Implications fro crop yield distribution modeling. *Amer.J.Agr.Econ.*, 93(4):968–982, 2011.

[5] E. Edwards. Trend-adjusted actual production history. Ag Decision Maker, Iowa State University Extension and Outreach. 2012.

[6] Paul Gallagher. U.s. soybean yields: estimation and forecasting with nonsymmetric disturbances. *American Journal of Agriculture Economics*, 69(4):796–803, 1987.

[7] Barry K. Goodwin and Alan P. Ker. Nonparametric estimation of crop yield distributions: Implications for rating group-risk crop insurance contracts. *American Journal of Agricultural Economics*, 80(1):139–153, 1998.

[8] Barry K Goodwin, Matthew C Roberts, and Keith H Coble. Measurement of price risk in revenue insurance: implications of distributional assumptions. *Journal of Agricultural and Resource Economics*, pages 195–214, 2000.

[9] Richard E. Just and Quinn Weninger. Are crop yields normally distributed? *American Journal of Agricultural Economics*, 81(2):287–304, 1999.

[10] Alan P Ker and Keith Coble. Modeling conditional yield densities. *American Journal of Agricultural Economics*, 85(2):291–304, 2003.

[11] Alan P Ker and Barry K Goodwin. Nonparametric estimation of crop insurance rates revisited. *American Journal of Agricultural Economics*, 82(2):463–478, 2000.

[12] C. Lanoue, B. J. Sherrick, J. D. Woodard, and N. D Pasulson. Evaluating yield models fro crop insurance rating. Pater presented at the joint annual meeting of the Agricultural and Applied Economics Association, the Canadian Agricultural Economics Society, and the Western Agricultural Economics Associating, Denver, Colorado. 2010.

[13] Simon CK Lee and X Sheldon Lin. Modeling and evaluating insurance losses via mixtures of erlang distributions. *North American Actuarial Journal*, 14(1):107–130, 2010.

[14] Carl H. Nelson and Paul V. Preckel. Are crop yields normally distributed? *American Journal of Agricultural Economics*, 71(2):370–378, 1989.

[15] Vitor A. Ozaki, Barry K. Goodwin, and Ricardo Shirota. Parametric and nonparametric statistical modelling of crop yield: Implications for pricing crop insurance contracts. *Applied Economics*, 40(9):1151–1164, 2011.

[16] Lysa Porth, Wenjun Zhu, and Ken Seng Tan. A credibility-based erlang mixture model for pricing crop reinsurance. *Agriculture Finance Review*, 74(2):162–187, 2014.

[17] Gideon Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

[18] B. Sherrick, and G. Schnitkey F. Zanini, and S. Irwin. Crop insurance valuation under alternative yield distributions. *American Journal of Agriculture Economics*, 86(2):406–419, 2004.

[19] B. Sherrick and G. Schnitkey. Trend-adjusted aph yield endorsement. Dept. of Agr. and Consumer Econ., University of Illinois Extension. 2011.

[20] Calum Turvey, Jinhua Zhao, et al. *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, 1999.

[21] Zhiying Xu. What can we inder about farm-level crop yield pdf's from county-level pdf's? Plan B M. S. Research Paper, Department of Agricultural Economics, Michigan State University East Lansing, MI. 2004.