

Parameterization of Crop Yields and Pricing of Crop Insurance Contracts Using the Actual Production History Method

Mark Wise
Master's Project
Illinois State University

Table of Contents

| | |
|---|----|
| Introduction | 3 |
| Background Research | 4 |
| Data Description and Research Methodology | 6 |
| Statistical/Data Analysis | 8 |
| Discussion/Conclusion | 9 |
| References | 10 |
| Appendix | 11 |

Abstract

The purpose of this paper is to fit four probability distributions to crop yields, three of which are parametric, and then to price crop insurance contracts using the Actual Production History method. The method of moments was used to estimate the parameters of the distributions for twenty-four districts in the Northern Region of Ghana. It is found that the Beta distribution's premium rates are greater than the Gamma and Lognormal premium rates in every district. The gamma premium rates are greater than the lognormal premium rates in almost every district, but by very little. The empirical premium rates are very inconsistent. The more expensive premium rates happen when there is a higher skewness coefficient. The lowest premium rate occurred in the East Mamprusi District, and the highest premium occurred in the Bongo district.

Introduction

Crop insurance has been a highly studied topic in recent years. It can exist in several forms: loss crop insurance, revenue crop insurance, and yield crop insurance. Loss crop insurance is the term used to describe insurance against losses due to natural disasters such as floods, hail, and hurricanes. Revenue crop insurance protects farmers from low revenues, and yield crop insurance protects farmers from low yields. This paper will focus on this last issue.

The first question that comes to mind when one thinks of revenue or yield crop insurance is why are farmers protected against revenue and yield decreases? In any other industry, insurance is not offered for these types of protections. A clothing company can't buy insurance to protect them from low sales. So why then is revenue and yield insurance offered to farmers? The answer to that is the agriculture industry is essential for human survival. If farmers go out of business, there will be no food production and the human race will cease to exist. The alternative has almost the same effect. If a single farmer or two dominate the market, food prices will skyrocket and be unaffordable for most. These are the reasons why crop insurance exists.

One form of low yield crop insurance is the Actual Production History method. Using this method, the crop insurance premium depends on the farmer's historical yield. This paper discusses how to calculate the Actual Production History premium rate after fitting the crop yields from twenty-four different districts in three regions in Northern Ghana to three parametric distributions and the discrete empirical distribution. (Klugman, Panjer, and Willmot, 2008) define a parametric distribution as a set of distribution functions each determined by specifying one or more values, called parameters, each of which is fixed and finite.

Several methods have been developed for modeling crop insurance premiums. (Goodwin & Ker, 1998) use nonparametric methods to model the crop yield distribution, (Ker & Goodwin, 2000) discuss both Nonparametric Empirical Bayes kernel density estimation and spatial-temporal models (space-time models), and (Sherrick, Zanini, Schnitkey, & Irwin, 2004) use semiparametric methods using the Normal, Logistic, Lognormal, Weibull, and Beta distributions. (Just, R.E. & Q. Weninger, 1999) also discuss the normality of crop yields.

With the empirical distribution, each observation has the same probability of occurring. One advantage of using the empirical distribution is that conclusions can be made directly from the data, thus increasing the accuracy of the calculations. However, a disadvantage is that it is discrete, and when new data is observed, the entire distribution changes.

A continuous parametric distribution, on the other hand, has the advantage of being flexible. Instead of the entire distribution changing when new data is added, only the parameters

change, leaving the original distribution or “frame” in place. This flexibility gives a parametric distribution a big advantage in that it can have different values of skewness and kurtosis.

Skewness is defined as $\frac{\mu_3}{\sigma^3}$, and kurtosis is defined as $\frac{\mu_4}{\sigma^4}$, where μ_3 is the third central moment, μ_4 is the fourth central moment, and σ is the standard deviation (Klugman, Panjer, and Willmot, 2008). Being able to take on different values for these two quantities is important because the first four central moments of distribution are the main descriptors of its shape (Ramirez, McDonald, and Carpio, 2010). A disadvantage however, is that usually the distribution is not known, and there has to be an initial assumption about which distribution the data comes from.

The rest of this paper discusses how to fit three distributions to this yield data: the lognormal distribution, the beta distribution, and the gamma distribution. After the distributions are fit, the calculation of the Actual Production History premium rates will be discussed. Once the premium rates from the parametric distributions are calculated, they will be compared to the empirical rates. The following section examines the problems that arise when a private crop insurance industry attempts to be created.

Background Research

Developing a private crop insurance industry is not an easy task, especially in a developing country. In a developing country, such as Ghana, there are several factors that need to be considered before selling insurance: credit history, amount of risk, etc., but the data is not available. So great care must be taken before entering this market. A few of the biggest problems are moral hazard, adverse selection, and systemic risk.

Moral hazard is the term used to describe the change in the insured’s behavior after being insured, that is, once someone becomes insured, they will take less care in preventing the event from which they are insured. In some situations, they may even purposefully cause the event to happen or increase its severity. In the case of crop insurance, there can be situations such as planting the crops in low quality soil or not watering crops. Moral hazard can be a very expensive problem, and therefore could prevent the emergence of a private sector crop insurance market (Chambers, 1989).

Another big problem is adverse selection. When several different insured’s get the same coverage in the same area, there is bound to be some price discrepancies. While all insured’s paid the same premium, some insured’s have more risk and some have less. Those with less risk might realize that the insurance is not worth the price, and stop the coverage. However, those with more risk might think it is a great bargain and decide to increase the coverage. So there are less risky insured’s leaving the pool and more risky insured’s entering the pool, causing a large, risky group of insured’s. This is another problem that can be costly for insurers (Skies & Reed, 1986).

A third issue with developing a private crop insurance industry is systemic risk. Systemic risk is the concept that the entire system may fail due to large loss events, moral hazard, and adverse selection (Miranda & Glauber, 1997).

In Northern Ghana, there are eight main crops: Maize, Rice, Sorghum, Millet, Cassava, Yam, G’Nuts, and Cowpea. This analysis focuses only on maize. The reason for this is that maize is the only crop that is consistently grown. That is, it is the only crop for which the data is available for all years.

Before starting the analysis, it is important to understand the distributions used in the yield modeling. Below are three of the distributions and their various quantities. There are a few things to notice. First of all, although the beta distribution appears to be a three-parameter distribution, the parameter theta must be set in advance. Theta is determined by rounding up the maximum yield to the nearest 0.1.

Beta Distribution

The probability density function is given by:

$$(1) \quad f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left(\frac{x}{\theta}\right)^a \left(1 - \frac{x}{\theta}\right)^{b-1} \frac{1}{x}, \quad a, b > 0, \quad 0 < x < \theta$$

where $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$. The moments are given by:

$$(2) \quad E(X^k) = \frac{\theta^k \Gamma(a+b) \Gamma(a+k)}{\Gamma(a) \Gamma(a+b+k)}, \quad k > -a$$

and the method of moments estimators of a and b are:

$$(3) \quad \hat{a} = \frac{\theta m^2 - mt}{\theta t - \theta m^2}, \quad \hat{b} = \frac{(\theta m - t)(\theta - m)}{\theta t - \theta m^2}$$

Gamma Distribution

Probability Density Function:

$$(4) \quad f(x) = \frac{x^{\alpha-1} e^{-\frac{x}{\theta}}}{\theta^{\alpha} \Gamma(\alpha)}, \quad \alpha, \theta, x > 0$$

Moments:

$$(5) \quad E(X^k) = \frac{\theta^k \Gamma(\alpha + k)}{\Gamma(\alpha)}, \quad k > -\alpha$$

Method of Moments Estimates:

$$(6) \quad \hat{\alpha} = \frac{m^2}{t - m^2}, \quad \hat{\theta} = \frac{t - m^2}{m}$$

Lognormal Distribution

Probability Density Function:

$$(7) \quad f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{\left(\frac{\ln x - \mu}{\sigma}\right)^2}{2}}, \quad \mu, \sigma, x > 0$$

Moments:

$$(8) \quad E(X^k) = e^{k\mu + \frac{1}{2}k^2\sigma^2}, \quad \forall k$$

Methods of Moments Estimates:

$$(9) \quad \hat{\mu} = \ln(m) - \frac{1}{2}\hat{\sigma}^2, \quad \hat{\sigma} = \sqrt{\ln(t) - 2\ln(m)}$$

where

$$(10) \quad m = \frac{\sum_{i=1}^n x_i}{n}, \quad t = \frac{\sum_{i=1}^n x_i^2}{n}$$

Equations 1 through 10 were taken from (Klugman, Panjer, and Willmot, 2008).

These distributions were chosen for several reasons. The beta distribution was chosen for its support. Observing the data, it appears as though there is an upper bound on crop yields. Another advantage of the beta distribution is that it is flexible in the sense that it can be either positively or negatively skewed. This is not true for the gamma or lognormal distributions.

The gamma distribution, however has its advantages as well. If we let γ_3 be the skewness, then we have:

$$(11) \quad \begin{aligned} \gamma_3 &= \frac{E[(X - \mu)^3]}{\sigma^3} \\ &= \frac{E(X^3) - 3E(X^2)\mu + 2\mu^3}{\sigma^3} \\ &= \frac{\theta^3\alpha(\alpha+1)(\alpha+2) - 3\theta^2\alpha(\alpha+1)\alpha\theta + 2\alpha^3\theta^3}{\alpha^{\frac{3}{2}}\theta^3} \\ &= \frac{2\alpha\theta^3}{\alpha^{\frac{3}{2}}\theta^3} = \frac{2}{\sqrt{\alpha}} \end{aligned}$$

This shows that the skewness depends only the value of one parameter, α , and furthermore, since $\alpha > 0$, the gamma distribution only has positive skewness. The lognormal distribution also only allows for positive skewness (Ramirez & McDonald, 2006).

Data Description and Research Methodology

The data sets used in the analysis of the actuarially fair crop insurance premium rates were provided by the Katie School of Insurance, and originated from the Statistics, Research, and Information Directorate of the Ministry of Food and Agriculture using the data from years

1992 through 2008. The data used in this analysis consists of each year's production, cropped area, and yield all broken down by region and district. The data used here only reflects three regions in Northern Ghana: the Northern Region, the Upper West Region, and the Upper East Region.

In the Northern Region, in years 1992 through 2004, there are thirteen districts, and in years 2005 through 2008, there are eighteen districts. Ten of these regions were combined to reduce the number of districts back to thirteen. In the Upper West Region in years 1992 through 2004, there are five districts, and in years 2005 through 2008 there are eight districts, five of which were combined to reduce the number of districts back to five. Similarly, in the Upper East region in years 1992 through 2004, there are six districts, and in years 2005 through 2008 there are eight districts, four of which were combined to reduce the number of districts back to six. All of these changes are shown in Table 1.

Crop yield is defined as the number of units of crops produced per unit of area cropped, or total number of units of crops produced divided by total area cropped. Assuming there is only a finite amount of land that can be cropped, it appears as if crop yields are finite (there can be only a finite number of units of crop per unit of land). Therefore, the choices of parametric distributions should be limited to those with either finite support or relatively light right tails, meaning the probability of observing a high yield is small.

This leads to the three parametric distributions discussed above: Lognormal, Gamma, and Beta. The Lognormal and Gamma distributions both have support on the interval $(0, \infty)$, however they both have light right tails based on the existence of moments test, which states that the more positive moments that exist for a probability distribution, the lighter the right tail (Klugman, Panjer, & Willmot, 2008). For the lognormal distribution, all positive moments exist. For the gamma distribution, all moments k exist such that $k > -\alpha$. Similarly, for the beta distribution, all k moments exist such that $k > -a$. Since $\alpha, a > 0$, it implies that all positive, and even some negative moments exist for both the gamma and beta distributions (Klugman, Panjer, & Willmot, 2008).

Each distribution is fit to each district's crop yields by the method of moments and is defined in Appendix A. The method of moments estimates the distribution parameters by setting the raw sample moments equal to the distribution moments, thus the number of sample moments needed is equal to the number of parameters being estimated. Only the first two moments are used in this analysis.

After all the data has been fitted to each distribution, all that is left is to find the premium rate. Using Actual Production History, the premium rate depends on the historical expected yield. Let Y be the random variable representing the yield, λ be the coverage level such that $0 < \lambda < 1$, and y^e be the expected yield based on the historical data. Then the Actual Production History method pays out in the following way according to

$$(12) \quad \text{Payment} = \begin{cases} \lambda y^e - Y, & Y \leq \lambda y^e \\ 0, & Y > \lambda y^e \end{cases}$$

Therefore expected losses paid out by the insurance are:

$$\begin{aligned}
(13) \quad E(\text{Payment}) &= \int_{-\infty}^{\lambda y^e} (\lambda y^e - y) f(y) dy \\
&= \lambda y^e \int_{-\infty}^{\lambda y^e} f(y) dy - \int_{-\infty}^{\lambda y^e} y f(y) dy \\
&= \lambda y^e F_Y(\lambda y^e) - E(Y, Y < \lambda y^e) \\
&= F_Y(\lambda y^e) [\lambda y^e - E(Y|Y < \lambda y^e)]
\end{aligned}$$

The premium rate is then defined as the ratio of expected losses to maximum possible liability, or

$$(14) \quad PR = \frac{F_Y(\lambda y^e) [\lambda y^e - E(Y|Y < \lambda y^e)]}{\lambda y^e}$$

The expectation in the numerator is evaluated using the trapezoidal rule with $n=1000$. The trapezoidal rule is:

$$(15) \quad \int_a^b f(x) dx \approx \frac{b-a}{2n} [f(x_0) + 2f(x_1) + \dots + 2f(x_{n-1}) + f(x_n)]$$

where $a = x_0 < x_1 < \dots < x_n = b$ (Larson, Hostetler, and Edwards, 2002).

If we look back at equation (14), it can be shown that it is equivalent to

$$(16) \quad PR = F_Y(\lambda y^e) \left[1 - \frac{E(Y|Y < \lambda y^e)}{\lambda y^e} \right]$$

which means that the actuarially fair premium rate is bounded above by the probability the yield is less than or equal to the coverage times expected yield. It is clear from equation (16) that the premium rate is between 0 and 1. But what does the premium rate actually mean? It means that for a base price, β , of a crop, the insured has to pay $PR * \beta$ for the insurance for each crop.

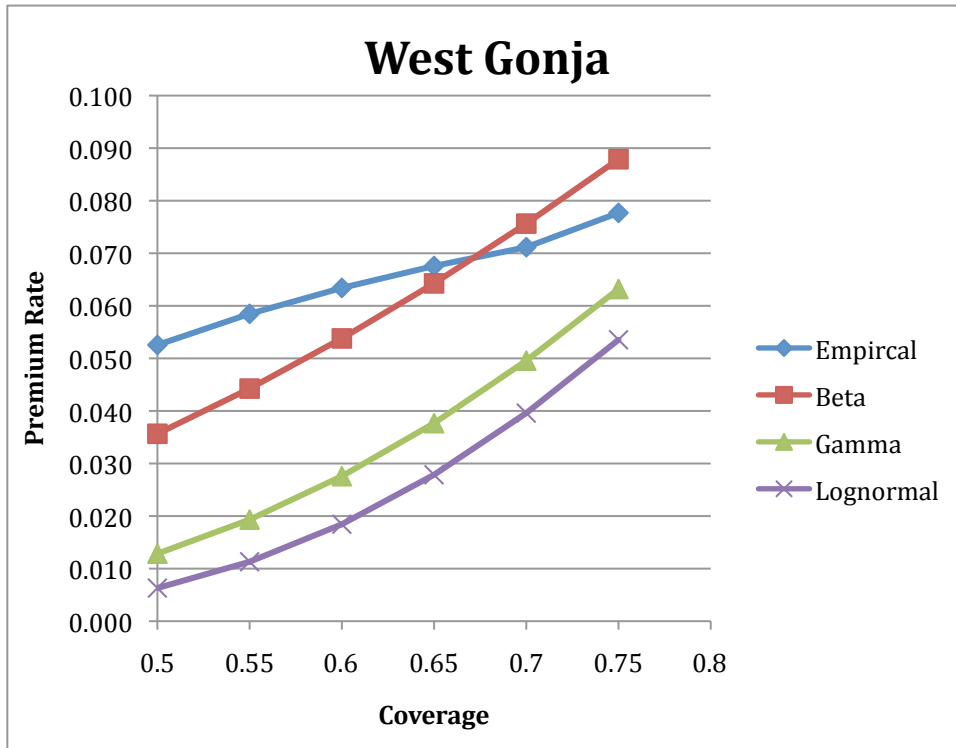
Statistical/Data Analysis

The following graph shows the change in premium rates for the four distributions in the West Gonja District of the Northern region of Ghana. The graphs of the premium rates of other districts are in the Appendix. In every district, the beta, gamma, and lognormal rates all increase together. The only line that constantly changes shape is the empirical rate line. This is due to the fact that the empirical distribution is a discrete distribution. Discrete distributions only work well when there are only a few values the distribution could take on. With regard to crop yields, a crop yield can almost be any positive number, up to a certain value. Therefore, a discrete distribution modeling crop yields is not very realistic.

The gamma and lognormal premium rates are consistency significantly lower than the beta rates. This could be due some positive skewness in the beta distribution or to some high kurtosis in the gamma and lognormal distributions. It is shown above that the skewness for the

gamma distribution is $\frac{2}{\sqrt{\alpha}}$. Thus the highest α will give the lowest skewness. The Nanumba district has the highest α of 24.15. Looking at the Nanumba graph, the gamma line almost sits directly on top of the lognormal line, indicating that a low skewness decreases the premium rate.

Figure 1
Premium Rates for the West Gonja District for the four distributions.



For each district in Northern Ghana, the Beta premium rates always overprice the gamma and lognormal premium rates. The Lognormal premium rates almost always under price the gamma premium rates. The highest premium rates come in the Bongo district of the Upper East region of Northern Ghana. This is due to several zero yield values observed in the district. The lowest premium rates observed were all in the Northern Region, in the Nanumba, East Mamprusi, and Bole districts. It appears that the low premium rates are due to lower variance, which makes sense because a higher variance would mean that there are more observations further from the mean. The higher the variance, the more likely that there are observations below the guaranteed yield. However, in these three districts, there were no observations below the guaranteed yields, resulting in a zero premium rate.

Discussion and Conclusion

In this paper it was shown that the Beta, Gamma, and Lognormal distributions could be fit to model crop yields. As it can be seen from the graphs below, the Beta premium rates are always more than the Gamma and Lognormal premium rates, and the Empirical rates vary based on the size of the yields.

The reasoning behind the expensive Beta rates is unknown, so it can only be speculated. One possible reason is the Beta distribution's lack of a right tail. Not having a right tail puts more probability on the lower values, and increases the c.d.f. According to equation (16), a higher c.d.f. results in a higher premium rate. This is because the premium rate is proportional to both the c.d.f. and $1 - E(Y|Y < \lambda y^e)$. When the c.d.f. increases, $E(Y|Y < \lambda y^e)$ also increases.

Recommendations

After this analysis, there are other areas that need to be explored. First, it would be beneficial to apply these methods to other distributions not discussed here. There are other distributions discussed in (Norwood, B, M.C. Roberts, & J.L.Lusk, 2004) that can take on more values in the skewness-kurtosis plane.

References

- Chambers, R.G. (1989). Insurability and Moral Hazard in Agricultural Insurance Markets. *American Journal of Agricultural Economics*, 71(3), 604-616.
- Goodwin, B.K. & A.P. Ker (1998). Nonparametric Estimation of Crop Yield Distributions: Implications for Rating Group-Risk Crop Insurance Contracts. *American Journal of Agricultural Economics*, 80(1), 139-153.
- Just, R.E. & Q. Weninger (1999). Are Crop Yields Normally Distributed? *American Journal of Agricultural Economics*, 81(2), 287-304.
- Ker, A.P. & B.K. Goodwin (2000). Nonparametric Estimation of Crop Insurance Rates Revisited. *American Journal of Agricultural Economics*, 83(2), 464-478.
- Klugman, S.A., Panjer, H.H., & G.E. Willmot (2008). *Loss Models: From Data to Decisions (Third Edition)*. Hoboken, NJ: John Wiley & Sons, Inc.
- Larson, R., Hostetler, R.P., Edwards, B.H., & Heyd, D.E. (2002). *Calculus of a Single Variable (Seventh Edition)*. Boston, MA: Houghton Mifflin Company.
- Miranda, M.J. & J.W. Glauber (1997). Systemic Risk, Reinsurance, and the Failure of Crop Insurance Markets. *American Journal of Agricultural Economics*, 79(1), 206-215.
- Norwood, B., M.C. Roberts, & J.L. Lusk (2004). Ranking Crop Yield Models Using Out-of-Sample Likelihood Functions. *American Journal of Agricultural Economics*, 86(4), 1032-1043.
- Norwood, B., M.C. Roberts, & J.L. Lusk (2004). Ranking Crop Yield Models Using Out-of-Sample Likelihood Functions. *American Journal of Agricultural Economics*, 86(4), 1032-1043.
- Ozaki, V.A., B.K. Goodwin, R. Shirota (2008). Parametric and nonparametric statistical modeling of crop yield: implications for pricing crop insurance contracts. *Applied Economics*, 40(7-9), 1151-1164.

Ramirez, O.A., T. McDonald (2006). Ranking Crop Yield Models: A Comment. *American Journal of Agricultural Economics*, 88(4), 1105-1110.

Sherrick, B.J., F.C. Zanini, G.D. Schnitkey, & S.H. Irwin (2004). Crop Insurance Valuation Under Alternative Yield Distributions. *American Journal of Agricultural Economics*, 86(2), 406-419.

Skees, J.R. & M.R. Reed (1986). Rate Making for Farm-Level Crop Insurance: Implications for Adverse Selection. *American Journal of Agricultural Economics*, 68(3), 653-659.

Appendix

Table 1

Combination of Districts in Northern Region.

The column on the right contains the combinations. The center column contains the names used in this study.

| 1992-2002 | 2003-2004 | 2005-2008 |
|------------------|------------------|------------------------------------|
| Damango | West Gonja | West Gonja Central Gonja |
| Yendi | East Dagomba | East Dagomba |
| Bimbilla | Nanumba | Nanumba North Nanumba South |
| Gushiegu/Karaga | Gushiegu/Karaga | Gushiegu Karaga |
| Gambaga | East Mamprusi | East Mamprusi Bunkpurugu-Yunyoo |
| Savelugu/Nanton | Savelugu/Nanton | Savelugu/Nanton |
| Salaga | East Gonja | East Gonja |
| Tamale | West Dagomba | Tamale Metropolitan |
| Bole | Bole | Bole Sawla-Tuna-kalba |
| Saboba/Chereponi | Saboba/Chereponi | Saboba/Chereponi |
| Tonlon/Kumbugu | Tonlon/Kumbugu | Tonlon/Kumbugu |
| Walewale | West Mamprusi | West Mamprusi |
| Zabzugu/Tatale | Zabzugu/Tatale | Zabzugu/Tatale |

Table 2

Combinations of Districts in Upper West Region.

The column on the right contains the combinations. The center column contains the names used in this study.

| 1992-2002 | 2003-2004 | 2005-2008 |
|-----------|------------------|------------------------------------|
| Wa | Wa | Wa West Wa East Wa Municipal |
| Lawra | Lawra | Lawra |
| Tumu | Sisala | Sisala West Sisala East |
| Jirapa | Jirapa-Lambussie | Jirapa-Lambussie |
| Nadowli | Nadowli | Nadowli |

Table 3

Combinations of Districts in Upper East Region.

The column on the right contains the combinations. The center column contains the names used in this study.

| 1992-2002 | 2003-2004 | 2005-2008 |
|-----------------|-----------------|---------------------------------|
| Builsa | Builsa | Builsa |
| Kassena/Nankana | Kassena/Nankana | Kassena/Nankana |
| Bongo | Bongo | Bongo |
| Bolgatanga | Bolgatanga | Bolga Talensi Nagdam |
| Bawku East | Bawku East | Bawku Municipal Garu Tempene |
| Bawku West | Bawku West | Bawku West |

Figure 2
 Premium Rates for the four distributions in the 24 districts in Northern Ghana.

