Agricultural Finance Review

Big assumptions for small samples in crop insurance

Ashley Elaine Hungerford Barry Goodwin

Article information:

To cite this document:
Ashley Elaine Hungerford Barry Goodwin , (2014),"Big assumptions for small samples in crop insurance", Agricultural Finance Review, Vol. 74 Iss 4 pp. 477 - 491
Permanent link to this document:
http://dx.doi.org/10.1108/AFR-09-2014-0025

Big assumptions for small samples in crop insurance

Ashley Elaine Hungerford
Economic Research Services, Washington, District of Columbia, USA, and
Barry Goodwin
Department of Agricultural and Resource Economics, North Carolina State University, Raleigh, North Carolina, USA

Abstract

Purpose – The purpose of this paper is to investigate the effects of crop insurance premiums being determined by small samples of yields that are spatially correlated. If spatial autocorrelation and small sample size are not properly accounted for in premium ratings, the premium rates may inaccurately reflect the risk of a loss.

Design/methodology/approach – The paper first examines the spatial autocorrelation among county-level yields of corn and soybeans in the Corn Belt by calculating Moran's *I* and the effective spatial degrees of freedom. After establishing the existence of spatial autocorrelation, copula models are used to estimate the joint distribution of corn yields and the joint distribution of soybean yields for a group of nine counties in Illinois. Bootstrap samples of the corn and soybean yields are generated to estimate copula models with the purpose of creating sampling distributions.

Findings – The estimated bootstrap confidence intervals demonstrate that the copula parameter estimates and the premium rates derived from the parameter estimates can vary greatly. There is also evidence of bias in the parameter estimates.

Originality/value – Although small samples will always be an issue in crop insurance ratings and assumptions must be made for the federal crop insurance program to operate at its current scale, this analysis sheds light on some of the issues caused by using small samples and will hopefully lead to the mitigation of these small sample issues.

Keywords Crop insurance, Bootstrap, Copula, Spatial autocorrelation, Systemic risk, Yield distribution

Paper type Research paper

Introduction

Recent passage of the much – delayed Farm Bill serves to further strengthen the critical role that federally – subsidized crop insurance plays in US agricultural policy. The new legislation includes options to select "Agricultural Risk Coverage", which functions as an aggregate county – level or whole – farm level revenue insurance program with payments based on historical base acreage. The new Farm Bill also introduces a "Supplemental Coverage Option," (SCO) that provides farmers with an optional county – level insurance program that covers a portion of the existing deductible. Cotton has its own version of

This paper was presented in Finance Section paper session at the 2013 Agricultural and Applied Economics Association (AAEA) annual meeting, August 4-6, in Washington, DC. The papers in these sessions are not subjected to the journal's standard refereeing process. Reviews conducted by N. Paulson and/or C.G. Turvey.

The views expressed are those of the authors and should not be attributed to the Economic Research Service or USDA. Goodwin was supported by the US Forest Service and the North Carolina Agricultural Research Service.

Agricultural Finance Review
Vol. 74 No. 4, 2014
pp. 477-491

0002-1466
DOI 10.1108/AFR-09-2014-0025

SCO in the "Stacked Income Protection Plan". The crop insurance title also mandates development of a revenue – minus – cost insurance plan. All of these new insurance plans provide revenue coverage, meaning that payments may be triggered by low prices, low yields, or a combination of both that results in a revenue shortfall.

All of these plans address coverage of multiple, dependent sources of risk. Further, the largest share of liability written in the current crop insurance program is in the form of revenue coverage, which now accounts for almost 90 percent of total liability in the federal program. In terms of individual farm – level insurance coverage, the "Revenue-Protection" plan with harvest – price replacement accounts for over 80 percent of total liability. A critical parameter in the design and rating of these revenue insurance plans is the measure of dependence or correlation among the various sources of risk. In the case of revenue insurance, one is typically concerned with the inverse correlation that exists between crop yields and prices. In the case of more complex insurance instruments, such as the newly proposed revenue minus cost plan, one must be concerned with multiple dependencies. These dependencies may be complex to model and measure. In most cases, billions of dollars of liability is rated using very small numbers of observations or even by assumed values of correlation relationships that may be only weakly related to empirical measurement of actual dependencies. These dependencies are often assumed to be constant across the individual marginal distributions for individual sources of risk. For example, fixed correlations are estimated using as few as 15 annual observations and are assumed to hold at an aggregate state level in the rating of the individual coverage revenue plans that make up the vast bulk of the insurance book of business[1].

The current collection of insurance plans that provide individual (farm-level) coverage are often termed as the "COMBO" plan. This plan consists of traditional yield insurance and revenue insurance that bases coverage either on projected or realized harvest – time prices. As we have noted, the bulk of coverage applies to revenue with losses paid at the higher of predicted or realized harvest-time prices. When setting COMBO premium rates, strong assumptions are made on the distributions of yields and prices. Assuming away complex dependencies among yields in neighboring counties and imposing practical distributions may ease computation, but these assumptions have the potential to create bias and inefficiency during estimation. In this paper we examine the effects of these assumptions when they are applied in practice to the small samples of yields and prices typically used for crop insurance rating. In particular, we explore characteristics of yield data including tail dependence and spatial autocorrelation. Tail dependence is a measure of comovement in the tails of marginals in joint distributions. Spatial autocorrelation, the dependency among regions in a defined space, and tail dependence typically lead to heteroskedasticity in estimates. White (1981) discusses misspecification in the error term (i.e. heteroskadasticity) in nonlinear models causes inefficient estimates. Yatchew and Griliches (1985) demonstrate that under certain conditions, heteroskedasticity can cause estimation bias. An example of this estimation bias can arise in maximum likelihood estimates of discrete choice models. The complication of heteroskedasticity is compounded with the issue of small sample size, which may cause large (but often understated) standard errors and decreased statistical power.

The analysis focusses on soybean and corn yields from the group of counties surrounding McLean County in Illinois. We refer to this grouping as the McLean County group[2]. The region contains several counties that are among the top producers of corn and soybeans for the USA. Although we start with a preliminary

analysis that gives an overview of spatial autocorrelation for several states in the Corn Belt, our focus is on assessing the standard errors of estimates that are important in evaluating yield distributions and COMBO insurance plan rating. Bootstrapping provides a nonparametric method for calculating robust standard errors as well as confidence intervals. Bootstrap estimates are constructed for the parameter estimates of the joint distribution of yields for the McLean County group as well as the joint distribution for log price deviates and yields for McLean County. The joint distribution of the yields for the county group is constructed using copula modeling, which allows for tail dependence, while the joint distribution for the log price deviates and yields for McLean County is constructed using the Iman-Conover procedure. One of the motivations for studying spatial autocorrelation and tail dependence among yields is to derive a better understanding of the relationship between these dependencies and systemic risk.

In the early 1990s the perils of systemic risk in crop insurance were first considered by Miranda (1991). Systemic risk pertains to the likelihood of losses occurring simultaneously and dependently. As mentioned above, one of the main causes of systemic risk in crop yields is spatial autocorrelation and tail dependence. Ignoring systemic risk can cause underestimation of the variance, which will lead to underestimating the probability of a loss. Miranda and Glauber (1997) demonstrated that, compared to other insurance markets such as automobile and homeowners' insurance, crop insurance payouts are more correlated with each other. Within crop insurance, catastrophes such as the droughts of 2011 and 2012 led to $10.8 billion and $17.4 billion in indemnity payments (RMA, 2013). With nearly $117 billion in total liability in 2013 (excluding livestock coverage), proponents of subsidized crop insurance often argue that private firms are not able to fully bear the risk of one of these major catastrophes. As a result, a complex set of subsidies and favorable reinsurance terms are provided through the Standard Reinsurance Agreement (SRA) between the FCIC and private insurers. The SRA allows private insurers to share risk with the FCIC (Goodwin, 2012). The SRA does not diminish the importance of correctly estimating systemic risk. An important dimension of the SRA involves the ceding of policies to different reinsurance pools. Further, private reinsurance plays a critical role in covering the risks that are not fully addressed by the SRA. The structure of the National Flood Insurance Program (NFIP) mirrors the structure of the FCIC. Currently, the NFIP is on the brink of collapse due to the astronomic volume of indemnity payments paid after Hurricane Katrina and Hurricane Sandy (Government Accountability Office, 2013).

Current rating methods for COMBO insurance

The current methodology for rating COMBO policies is outlined by Coble *et al.* (2010). COMBO insurance rating begins with the calculation of an unloaded target rate, which is a function of loss cost ratios (LCRs, defined by the ratio of indemnity payments to total liability) for the county of interest. This rate is the anchor rate for insurance policies within the county. The rate is referred to as "unloaded" because it is calculated without the highest 20 percent of losses for the counties[3]. These large losses are accounted for in the catastrophic loading. The unloaded target rate is a weighted average of the historical LCRs of the county and its neighbors, weights are calculated with the Bühlmann method, which is defined as:

$$R = ZX + (1 - Z)\mu, \tag{1}$$

where $Z = P/(P+K)$ and R is the county unloaded target rate; Z, the Bühlmann credibility factor; X, the sample mean of the county of interest; μ, the mean of the

adjusted LCR of the county group; P, the exposure units; and $K = v/\alpha$ where v is the sample variance of the adjusted LCR for the county of interest; and α is the sample variance of the adjusted LCR for the county group.

Once the unloaded target rate has been established, revenue policies are rated using correlated draws from the marginal distributions of yields and prices. In the case of yields, the rates are calibrated to a truncated normal density. The distribution of prices is determined using the relevant implied options volatilities and futures prices. This approach, based on the standard Black-Scholes option pricing model, imposes a log-normal distribution on prices. Revenue rates are the calculated by means of simulation of the revenue distribution. Correlated price and yield draws are taken using the Iman and Conover (1982) procedure. The Iman-Conover procedure essentially reorders simulated yields and prices in order to obtain a predetermined rank correlation. The correlation is Spearman's ρ rank correlation coefficient of the average state yields and prices. These correlated random draws of yield and price deviates are then used to establish then premium rate for 65 percent coverage.

The steps applied in estimating revenue insurance rates involve a number of obvious assumptions. These have been evaluated in some detail by Coble *et al.* (2010). In most cases, a substantial degree of uncertainty, both in terms of the specifications imposed and the parameters used to represent these specifications, is ignored. For example, calibration of yields and prices to assumed parametric distributions does not explicitly recognize the estimation error associated with shape and location parameters as well as the uncertainty associated with any specific parametric distribution. Likewise, detrended data are typically treated as though they were observed directly rather than being subject to an estimation process. To the extent that uncertainty and sampling error is relevant, considerable uncertainty may be associated with revenue premium rates. Of course, rating revenue insurance is a problem that demands a practical solution and such assumptions are certainly necessary to some degree in order to facilitate an operational insurance program. However, this does not mean that the implications of such assumptions are not relevant to the overall operation of the program.

Although a number of assumptions are embedded in the COMBO rating process, we focus here on a single aspect of rating – the measure of dependence between yields and prices. As we have noted, this critical parameter is estimated using yield and price data from 1990 to 2005. The yield data are detrended using a linear trend and deviations from the trend are put into percentage terms (as a percentage of the trend yield). The data are then used to estimate production – weighted state – level correlation coefficients, which are then adjusted downward in an ad hoc fashion to reflect the greater variability of individual yields. Revenue "loads" are calculated by simulating yield and revenue coverage rates, with the load being given by the differences. This load is then added to the underlying yield protection rate to determine the revenue rate actually used in the program. Throughout these various steps, a large degree of specification uncertainty and sampling variability is likely to be relevant to the precision of the final rates.

The correlation between prices and yields is itself an embodiment of the correlation of yields across space. The greater is the correlation of yields across space, the stronger will be the relationship between prices and disaggregated (state, county, or farm) yields, since prices are determined on a well-integrated national market. Thus, systemic risk, spatial correlation, and the correlation of prices and yields are all aspects of the same fundamental phenomenon. We examine various aspects of this dependence relationship by evaluating the robustness of the rates to alternative specifications.

Methodology

As noted, correlation/dependence is represented using a reordering of data to achieve a desired degree of rank correlation. This Iman-Conover procedure uses a score function to determine the rankings. If, as is common and as is the case with the COMBO rating, one applies a normal score function, the Iman-Conover procedure is fully analogous to modeling correlation through the application of a Gaussian copula. This choice of a specification has implications for the joint relationships being modeled. In particular, application of a Gaussian copula (or, likewise, application of the Iman-Conover methodology) necessarily imposes a condition of zero tail dependence. This could have important implications for the accuracy of rates since one would expect stronger correlation to occur during events such as droughts or major floods that would lower yields on a wide scale. By exploring other copula models, our analysis will have the flexibility of the Iman-Conover procedure, but also allow correlation to change within the distribution.

To get a broader understanding of the relationship among neighboring regions, measures such as Moran's I are appropriate (Cressie, 1993). Moran's I is very similar to Pearson's correlation coefficient and can be used to evaluate correlation for cross-sectional spatial data. This measure is defined as:

$$\text{Moran's } I = \frac{N}{\Sigma_i \Sigma_j w_{ij}} \frac{\Sigma_i \Sigma_j w_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\Sigma_i (X - \bar{X})^2}, \tag{2}$$

where N is the number of regions and w_{ij} is an element of the spatial weighting matrix. The spatial weighting matrix places higher weight on regions close to each other and less weight on regions that are far apart. For Moran's I, the value $+1$ indicates perfect positive spatial autocorrelation; 0 indicates no spatial autocorrelation; and -1 indicates perfect negative spatial autocorrelation.

The effective number of spatial degrees of freedom for time-varying data, denoted N^*_{ef}, describes the number of regions in a spatio-temporal data set that are spatially independent. This measure is commonly used in climatology to determine the optimal placement of weather stations. There are various definition for the effective number of spatial degrees of freedom, we use:

$$N^*_{ef} = \frac{(tr\,\mathbf{G})^2}{(tr\,\mathbf{G}^2)} = \frac{\left(\sum_{i=1}^{N} G_{ii}\right)^2}{\sum_{i,j=1}^{N} G_{ij}^2} \tag{3}$$

where N is the number of regions, $\boldsymbol{G}$ is the $N \times N$ covariance matrix of the N regions, and tr is the trace of the matrix (Bretherton *et al.*, 1999).

The literature exploring copula modeling has increased dramatically in the last 15 years. The backbone of copula modeling stems from Sklar (1959), which states that any joint distribution can be represented as a function (a copula) of its marginal distributions. The representation is unique if the marginal distributions are continuous. From Sklar's Theorem, one is able build a variety of joint distributions. The joint density of the d-dimensional distribution can be written as:

$$f(\mathbf{u}) = c(F_1(u_1), \ldots, F_d(u_d)) \prod_{i=1}^{d} f_i(u_i) \tag{4}$$

The density of a copula is written as:

$$c(\mathbf{u}) = \frac{f(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d))}{\prod_{i=1}^{d} f_i(F_i^{-1}(u_i))} \tag{5}$$

Nelson (2006) provides an excellent overview on copula modeling. The most popular copulas include the elliptical and Archimedean families. Elliptical copulas consist of symmetric, elliptical distributions, such as the Gaussian and Student's t distributions. The Gaussian copula of dimension d has the form:

$$C_\rho(\mathbf{u}) = \Phi_\rho(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)), \tag{6}$$

where $\Phi-1$ is the inverse cumulative distribution of the standard normal distribution and $u_i \in [0, 1]$ for $i = 1, \ldots, d$. The Gaussian copula has zero tail dependence; however, the Student's t copula exhibits symmetric tail dependence. Being from the elliptical family, the Student's t copula of dimension d has a similar form to the Gaussian copula, which is:

$$C_{v,\rho}(\mathbf{u}) = t_{v,\rho}(t_v^{-1}(u_1), \ldots, t_v^{-1}(u_d)), \tag{7}$$

where v is the degrees of freedom.

Archimedean copulas are characterized by a single parameter and are of the form:

$$C(u) = \psi(\psi^{-1}(u_1) + \ldots + \psi^{-1}(u_d)), \tag{8}$$

where $u(\cdot)$ is the copula data and $\psi(\cdot)$ is the generator function. Table I shows the functional form and characteristics of several popular Archimedean copulas.

There are several measures of comovement often used when evaluating copula models. Asides from the traditional measure of linear correlation (Pearson's correlation coefficient), Kendall's τ and Spearman's ρ are rank correlation coefficient, are commonly considered. As discussed by Genest *et al.* (2011), the inverse of Kendall's τ is sometimes used to determine the parameter estimates (or starting values) in Archimedean copulas. Tail dependence measures the comovement of two variables at the extreme regions of the distribution (Fischer and Klein, 2007). The measure of tail dependence is defined by the copula function itself and not by the marginal distributions within the copula. There are separate definition for the lower and upper tail dependence coefficients and the specific copula function determines the nature of such tail dependencies with parameter estimates determining the exact nature of the tail dependency relationships. If we define a copula as $F_{X,Y}(x, y) = C(F_X(x), F_Y(y))$, then the lower tail dependence coefficient is defined as:

$$\lambda_L \equiv \lim_{u \to 0^+} P(Y \leqslant F_y^{-1}(u) | X \leqslant F_X^{-1}(u)) = \lim_{u \to 0^+} \frac{C(u, u)}{u}, \tag{9}$$

Table I. Popular Archimedean Copulas

Name	Generator	Inverse generator	Parameter	Tail dependence
Clayton	$(1+\theta x)^{-1/\theta}$	$\frac{1}{\theta}(x^{-\theta}-1)$	$\theta > 1$	Lower
Gumbel	$e^{-x^{1/\theta}}$	$(-\log(x))^\theta$	$\theta \geqslant 1$	Upper
Frank	$\frac{1}{\theta}\log(1-(1-e^{-\theta})e^{-x})$	$-\log\left(\frac{e^{-\theta x}-1}{e^{-\theta}-1}\right)$	$\theta > 1$	None

while the upper tail dependence coefficient is defined as:

$$\lambda_U \equiv \lim_{u \to 1^-} P(Y > F_y^{-1}(u) | X > F_X^{-1}(u)) = \lim_{u \to 1^-} \frac{1 - 2u + C(u,u)}{1 - u}. \tag{10}$$

Bootstrapping is a parametric or nonparametric technique for calculating standard errors and confidence intervals. It is typically applied in a nonparametric context by drawing random samples with replacement from the observed sample and estimating the parameters of interest for each of the replicates to approximate the distribution of the parameter estimate (Efron, 1993). The approximated distribution of the parameter estimate is not guaranteed to be centered at the parameter estimate calculated from the observed sample. Therefore, we correct this bias by adjusting the center of the confidence interval with:

$$z_{adj} = \Phi^{-1}\left(\frac{\sum_{i=1}^{M} I(\tilde{\beta}_i < \hat{\beta})}{M}\right) \tag{11}$$

where Φ^{-1} is the inverse of the normal cumulative density function, M is the number of random samples drawn from the observed sample, $\tilde{\beta}_i$ is the parameter estimate from random sample i, and $\hat{\beta}$ is the parameter estimate from the observed sample. The table statistic for a $(1-\alpha)$ percent bias-corrected confidence interval is adjusted from $(\pm\Phi^{-1}(\alpha/2))$ to $(z\_adj \pm \Phi^{-1}(\alpha/2))$.

If the approximated distribution of the parameter estimate is skewed, then a basic percentile confidence interval is inappropriate. This issue can be addressed through acceleration, which adjusts the bounds of the confidence interval to account for the skewness. The adjustment begins by calculating the value:

$$\hat{\alpha} = \frac{\sum_{i=1}^{n} (\bar{\beta}_i - \beta_j)^3}{6\sum_{i=1}^{n} (\bar{\beta}_i - \beta_j^2)^{3/2}}, \tag{12}$$

where n is the number of observations, $\bar{\beta}_i$ is the parameter estimate calculated without the ith observation, and $\beta_j = \frac{1}{n}\sum_{i=1}^{n} \bar{\beta}_i$. Then the $(1-\alpha)$ percent bias-corrected and accelerate confidence interval, referred to as a $(1-\alpha)$ percent BC_a confidence interval, has percentiles that are adjusted from $\alpha/2$ and $(1-\alpha/2)$ to:

$$\tilde{\alpha} = \Phi\left(z_{adj} + \frac{z_{adj} \pm \Phi^{-1}\left(\frac{\alpha}{2}\right)}{1 - \hat{\alpha} \times (z_{adj} \pm \Phi^{-1}\left(\frac{\alpha}{2}\right)}\right). \tag{13}$$

Empirical application

Annual county yields of soybeans and corn were obtained from the National Agricultural Statistical Services. After 2007 several counties were aggregated in to "Other Counties" in the available data. Therefore, the calculations for Moran's I and the effective number of spatial degrees of freedom, which encompass all the counties of Iowa, Illinois, and Indiana, have a sample period of 1960-2007. However, the analysis for the McLean County group has yields from 1960 to 2012. Since many technological and policy changes have occurred during the sample period, a nonparametric (LOESS)

regression model is used to detrend the yields. The detrended yields are used in both copula estimation and in the Iman-Conover procedure. These detrended yields have the form:

$$\hat{y}_{y,2012} = y_{2012}\left(1 + \frac{e_t}{\hat{y}_t}\right), \tag{14}$$

where e_t is the residual from the LOESS regression for year $t = 1960, \ldots, 2012$ and $\hat{y}_t$ is the predicted yield for year t.

Prices for futures contracts are obtained from the Chicago Board of Trade. We examine the log differences in price for the future contracts from February to November for soybeans and from February to December for corn. These are referred to as log price deviates.

To derive an overview of the spatial correlation in the Corn Belt, we calculate Moran's I and the Nef for the county yields of corn and soybeans of Iowa, Illinois, and Indiana. Table II shows Moran's I for 2007 as well as Nef for Iowa, Illinois, and Indiana. The results for Moran's I are statistically significant at the 1 percent level for all three states and both crops. Moran's I for every year from 1960 to 2007 is statistically significant at the 1 percent level. Interestingly, corn consistently displays stronger spatial autocorrelation than soybeans. Although Iowa, Illinois, and Indiana have 99, 102, and 92 counties, respectively, each of the states has approximately eight spatially independent areas according to the measure Nef. Moran's I and the effective number of spatial degrees of freedom demonstrate the strong spatial autocorrelation present among the yield observations for these counties. Therefore, if the spatial autocorrelation is not accounted for in the modeling, standard errors will be inaccurate and, depending on the analytical approach, parameter estimates may also be subject to bias.

The analysis with Moran's I and Nef covers very large regions. The examination of the McLean County group focusses on a much smaller region of nine counties. As noted, one of the main advantages of copula modeling is the ability to measure the coefficient of tail dependence. Figures 1 and 2 show the McLean County's detrended yields for corn and soybeans, respectively, plotted against the detrended yields of neighboring counties. These plots show long tails for low yields. We expect if one county experiences low yields so will the neighboring counties. This leads us to believe there may be lower tail dependence. Therefore, we estimate the Gaussian, Student's t, and Clayton copulas using the detrended yields from the McLean County group for the sample period 1960-2012. The graphical analysis leads one to conclude that a Gumbel copula, which has only upper tail dependence, is not appropriate. For the marginal distributions of each county, rank-based empirical distributions are estimated.

		Corn		Soybeans	
State	n	Moran's I	N_{ef}	Moran's I	N_{ef}
Iowa	99	0.7309	7.1397	0.4100	7.5009
Illinois	102	0.8288	7.7515	0.8811	7.8375
Indiana	92	0.3541	7.8688	0.5995	7.6149

Table II. Moran's I and N_{ef} for Iowa, Illinois, and Indiana

Notes: n is the number of counties for each states. Moran' I is a measure of spatial correlation. N_{ef} represents the effective number of spatial degrees of freedom

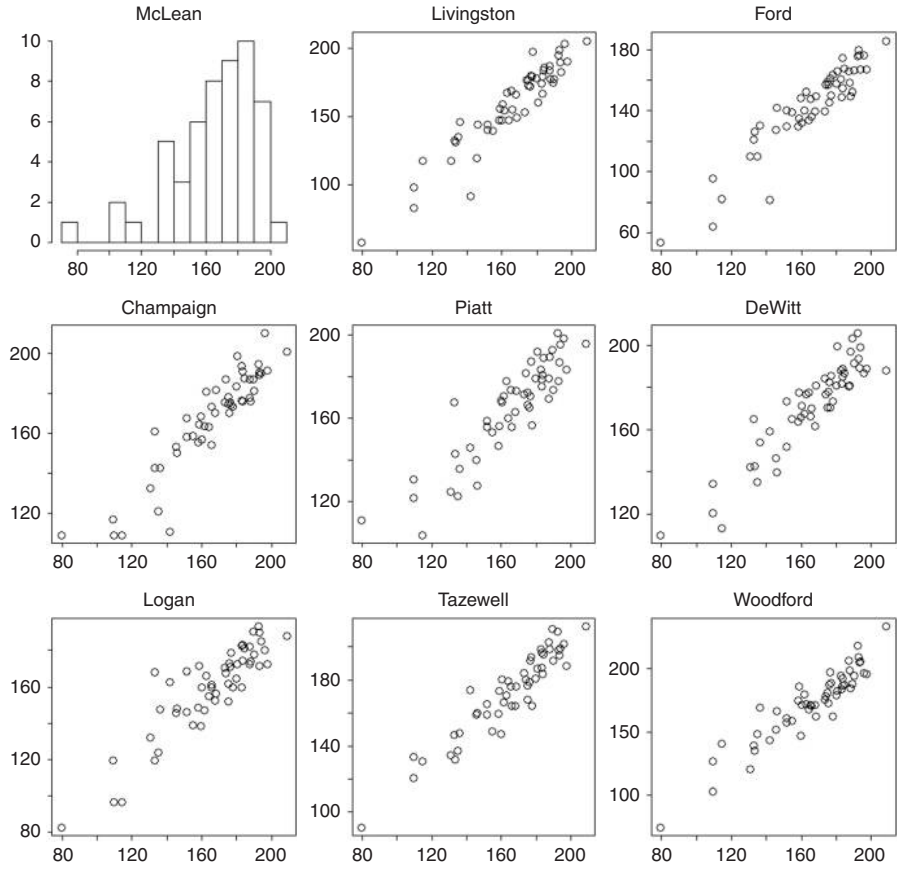

Figure 1.
The detrended yields of corn for McLean County plotted against the detrended yields of corn for its neighboring counties

Table III shows the copula parameter estimates for the McLean County group. In accordance with the results from Moran's I, corn yields have stronger dependence among the counties than soybeans. Note that the parameter estimates for the Gaussian and Student's t copulas are bounded to lie between $[-1, 1]$. The estimate for the Clayton copula is bounded from below by negative one, and higher values indicate stronger dependence. All three copulas models for both corn and soybeans have strong statistically significant parameter estimates. According to the Akaike Information Criterion (AIC), the Student's t copula provides the best fit for both corn and soybean yields. Therefore, out of the three types of copulas, the copula allowing for both upper and lower tail dependence provides the best fit. Table IV presents measures of dependence commonly associated with copulas. These measures of dependence show the relationship between two variables given the parameter estimates of the copulas in Table III. Kendall's τ and Spearman's ρ both show high correlation. Under the column "Tail Dependence", the value for the Student's t copula applies to both the lower and upper tail, while the value for the Clayton copula only applies to the lower tail[4]. The tail dependence coefficient for the Student's t copula shows that as one of the counties' yield approaches the lower (higher) extreme for corn, the probability of another counties' yield approaching the lower (higher) extreme is equal to 0.4676. For soybeans,

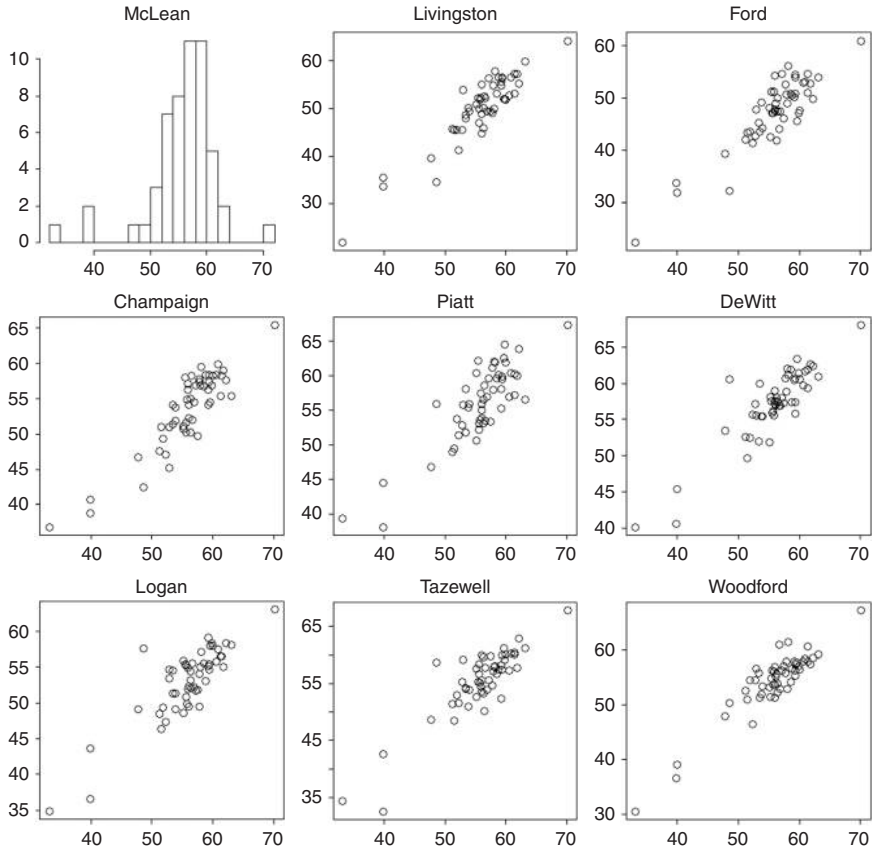

Figure 2.
The detrended yields of soybeans for McLean County plotted against the detrended yields of soybeans for its neighboring counties

Table III.
Copula estimates for the yields of the McLean county group during the sample period 1960-2012

Copula	Corn Estimate	Corn SE	Corn AIC	Soybeans Estimate	Soybeans SE	Soybeans AIC
Gauss	0.8829	0.0224	−745.6007	0.7586	0.0500	−448.2945
T	0.8921	0.0388	−781.4819	0.7814	0.0929	−484.8116
Clayton	3.1515	0.3979	−730.6254	1.8711	0.4401	−461.896

Table IV.
Measures of dependence for the copula estimates of the yields for the McLean county group during the sample period 1960-2012

Copula	Kendall's τ	Spearman's ρ	Tail dependence
Corn			
Gauss	0.6888	0.8732	0
T	0.70154	0.8830	0.4676
Clayton	0.6118	0.7975	0.8026
Soybean			
Gauss	0.5482	0.7430	0
T	0.5710	0.7666	0.2939
Clayton	0.4834	0.6441	0.6904

this probability is 0.2939. In both cases, a high degree of tail dependence relative to a Gaussian model is implied. This suggests that rates pertaining to losses deep in the tails may be significantly understated.

For the bootstrapped confidence intervals, there is no observable bias. However, there is skewness in the estimates, so the use of an accelerated confidence interval is appropriate. Results for the bootstrapped estimates are shown in Tables V and VI for corn and soybeans, respectively. The table also presents the lower and upper bounds for the 95 percent BC_a confidence intervals of the parameter estimates. The standard errors for the bootstrapped estimates are much higher than are the standard errors calculated from the observed sample.

Figure 3 illustrates the log price deviates plotted against the detrended yields for McLean County from 1960 to 2011. The correlation between the log price deviates and detrended yields for corn is −0.3503, while the correlation for soybeans is −0.3342. RMA uses the correlation of the average state yield and price deviates for the correlation in the Iman-Conover method.

The correlation of the average yields for Illinois and log price deviates is −0.3466 and −0.3270 for corn and soybeans, respectively. Therefore, the difference between the correlation using McLean County or Illinois is negligible.

Following a rating methodology similar to that used by RMA, we estimate the corn and soybean premium rates in McLean County for revenue insurance with 65 percent coverage and the Harvest Price Replacement Option. Tables VII and VIII contain the results from the Iman-Conover procedure performed on log price deviates and detrended yields of McLean County to determine the premium rates and related quantities for corn and soybeans, respectively. Skewness is present in both the bootstrapped estimates for corn and soybeans. Therefore, again the bias-corrected and

Table V. 95% BC_a confidence intervals for the parameter estimates and measures of dependence for the estimated copulas of the corn yields in McLean county group

	Bound	Gaussian	T	Clayton
Estimate	Lower	0.8447	0.8530	2.4815
	Upper	0.9215	0.9342	4.4145
Tail dependence	Lower	0	0.394	0.7563
	Upper	0	0.5726	0.8547
Kendall's τ	Lower	0.6405	0.6404	0.5537
	Upper	0.7461	0.7678	0.6882
Spearman's ρ	Lower	0.8328	0.8414	0.7404
	Upper	0.9145	0.9282	0.8644

Table VI. 95% BC_a confidence intervals for the parameter estimates and measures of dependence for the estimated copulas of the soybeans yields in McLean county group

	Bound	Gauss	T	Clayton
Estimate	Lower	0.6484	0.6504	1.1614
	Upper	0.8464	0.8724	2.7594
Tail dependence	Lower	0	0.1762	0.5505
	Upper	0	0.6749	0.5797
Kendall's τ	Lower	0.4491	0.4508	0.3674
	Upper	0.6425	0.6749	0.5798
Spearman's ρ	Lower	0.8346	0.6326	0.5227
	Upper	0.8621	0.8620	0.7779

accelerated 95 percent confidence intervals are used. The probability of a loss for corn in McLean County is 2.07 percent annually with the 95 percent BC_a confidence interval between 0.13 and 5.11 percent, whereas the probability of a loss for soybeans is 0.14 percent annually with the 95 percent BC_a confidence interval between 0.009 and 0.05 percent. The corn premium rates have a bias of 0.0002, while the soybean premiums rates have a bias of 0.00003. Overall we see a larger range for the confidence interval of corn with a lower bound premium rate of 0.00009 and upper bound of 0.000589, while the soybean premium rates have a 95 percent confidence interval from 9×10^{-7} to 0.00032. Note these results do not include any loadings or other adjustments to the rates.

The results demonstrate a simple result-revenue insurance premium rates are sensitive to assumptions regarding the correlation structure between prices and yields. Small sample sizes and specific parametric assumptions about the form of the joint

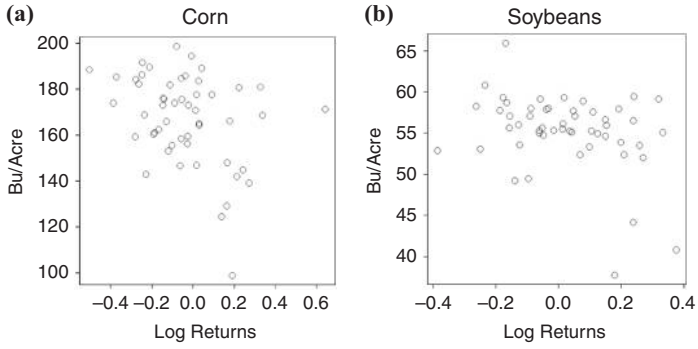

Figure 3.
The log price deviates plotted against the detrended yields for McLean County

Table VII.
Estimates related to corn COMBO rating in McLean county

	Lower bound	Sample estimate	Upper bound	SE	Bias
Probability	0.0013	0.0207	0.0511	0.0143	0.0006197
Expected loss	0.0250	0.6992	2.3388	0.7136	0.1114
Cond. Exp. loss	16.33	33.78	46.80	7.697	−0.5552
Premium rate	0.0000921	0.0014905	0.0058924	0.0015214	0.0002375

Notes: 95% BC_a confidence intervals, observed sample estimates, standard errors, and bias for the probability of a loss, expected loss, expected loss conditioned on a loss occurring, and premium rate

Table VIII.
Estimates related to soybean COMBO rating in McLean county

	Lower bound	Sample estimate	Upper bound	SE	Bias
Probability	0.0000921	0.0014905	0.0058924	0.0015214	0.0002375
Expected loss	0.000118	0.0245	0.1449	0.0389	0.0138
Cond. Exp. loss	1.1789	14.4212	29.7387	6.5948	3.6951
Premium rate	0.0000009	0.0000528	0.0003287	0.0000839	0.0000298

Notes: 95% BC_a confidence intervals, observed sample estimates, standard errors, and bias for the probability of a loss, expected loss, expected loss conditioned on a loss occurring, and premium rate

distribution have important implications for the accuracy of results. In particular, estimates of the correlation structure obtained using such small samples tend to be relatively imprecise, leading to fairly wide confidence bands for revenue rates. The nature of the relationships, in particular the degree of tail dependence, also has important implications for the accuracy of rates. In light of the billions of dollars currently insured under revenue insurance plans, even small differences in rate translate into significant cost implications for the Treasury and for farmers, who pay a portion of the rate.

Summary and concluding remarks

Current rating methods used in the federal crop insurance program are necessarily dependent upon a number of assumptions and are subject to numerous limitations. While highlighting the various aspects of rating that may be subject to these shortcomings, we focus on the critical role of dependence in the rating of revenue insurance contracts. Participation in the federal program is heavily skewed toward revenue coverage and the recent Farm Bill will provide numerous revenue – based enhancements to the current system of safety nets. We demonstrate that revenue rates may be subject to considerable variation that results from specification choices and estimation error. The small samples that one is forced to work with are also an important limitation that has implications for the accuracy of revenue rates.

Yields among neighboring counties will naturally have spatial autocorrelation, which can lead to inefficient estimates. This compounds with the issue of small sample size because spatial autocorrelation lessens the spatial degrees of freedom. The small sample size is an unavoidable issue when modeling yields. However, assuming away these issues may be hazardous because of the uncertainty in estimates. The current methodology only accounts for spatial weighting in the target rate, which is itself a point estimate. The issue of spatial dependence mainly concerns the second moments although in certain cases it can also cause biased estimates.

Our analysis demonstrates the fact that small sample sizes and spatial autocorrelation greatly affect standard errors. From Moran's I and the effective number of spatial degrees of freedom, we observe significant spatial autocorrelation among yields. This leads into the analysis using copula models, which shows us that not only is there correlation among yields for a county group but there is also nonzero tail dependence that is not recognized in current rating methods. Although there is insufficient evidence for tail dependence between the detrended yields and log price deviates, we do see large confidence intervals for the probabilities of a loss, expected losses, and premiums rates. In further analysis, we would like to incorporate the dependence seen among the county yields into the crop ratings. This could be accomplished with a nested model, which contains a copula for the county group. Then the Iman-Conover procedure would be used to determine the correlated draws of not only one county and the price deviates, but for the group of counties This method would better incorporate the uncertainty caused by the spatial dependencies.

Our analysis is not meant to imply that deficiencies exist in current rating methods. Rating insurance contracts is a problem that demands practical solutions and assumptions and limited samples will always be an issue. Our intent is rather to highlight some of the less obvious shortcomings that may be associated with existing rating practices. We have only addressed one aspect of the specification and estimation uncertainty associated with rating methods. A wide variety of rating issues presents

a number of interesting modeling and research challenges that merit the attention of applied crop insurance researchers.

Notes

1. See the comprehensive review of the COMBO rating methodology presented by Coble *et al.* (2010) for a detailed description of current rating practices.
2. The eight counties surrounding McLean County include Livingston, Ford, Champaign, Piatt, DeWitt, Logan, Tazewell, and Woodford.
3. Recent changes to the catastrophic loading procedures have reduced this to 10 percent.
4. Archimedean copulas represent dependence using a single parameter and thus must necessarily have zero tail dependence in one of the tails.

References

Bretherton, C.S., Widmann, M., Dymnikov, V.P., Wallace, J.M. and Blade, I. (1999), "The effective number of spatial degrees of freedom of a time-varying field", *Journal of Climate*, Vol. 12 No. 7, pp. 1990-2009.

Coble, K.H., Knight, T.O., Goodwin, B.K., Miller, M.F. and Rejesus, R.M. (2010), "A comprehensive review of the RMA APH and combo rating methodology", technical report, Risk Management Agency, Washington, DC, available at: www.rma.usda.gov/pubs/2009/comprehensivereview.pdf

Cressie, N.A. (1993), *Statistics for Spatial Data*, Wiley & Sons, Hoboken, NJ.

Efron, T. (1993), *An Introduction to the Bootstrap*, Chapman & Hall/CRC, Boca Raton, FL.

Fischer, M. and Klein, I. (2007), "Some results on weak and strong tail dependence coefficients for means of copulas", working paper, Discussion paper/University Erlangen-Nürnberg, Friedrich-Alexander, Chair of Statistics and Econometrics 78/2007.

Genest, C., Něslehov, J. and Ghorbal, J.B. (2011), "Estimators based on Kendall's tau in multivariate copula models", *Australian & New Zealand Journal of Statistics*, Vol. 53 No. 2, pp. 157-177.

Goodwin, B.K. (2012), "Copula-based models of systemic risk in us agriculture: implications for crop insurance and reinsurance contracts", working paper, available at: www4.ncsu.edu/~bkgoodwi/personal

Government Accountability Office (2013), "High-risk series: an update", technical report, GAO, Washington, DC.

Iman, R.L. and Conover, W.J. (1982), "A distribution-free approach to inducing rank correlation among input variables", *Communications in Statistics-Simulation and Computation*, Vol. 11 No. 3, pp. 311-333.

Miranda, M. (1991), "Area-yields crop insurance, reconsidered", *American Journal of Agricultural Economics*, Vol. 73 No. 2, pp. 233-242.

Miranda, M. and Glauber, J. (1997), "Systemic risk, reinsurance, and the failure of crop insurance markets", *American Journal of Agricultural Economics*, Vol. 79 No. 1, pp. 206-215.

Nelson (2006), *An Introduction to Copulas*, Springer, New York, NY.

Risk Management Agency (2013), *Federal Crop Insurance Corporation: Nationwide Summary – by Crop*, Washington, DC, available at: www.rma.usda.gov/data/sob.html

Sklar, A. (1959), "Fonctions de répartition àn dimensions et leurs marges", *Publ. Inst. Statisque*, No. 8, pp. 229-331.

White, H. (1981), "Consequences and detection of misspecified nonlinear regression models", *Journal of the American Statistical Association*, Vol. 76 No. 374, pp. 419-433.

Yatchew, A. and Griliches, Z. (1985), "Specification error in probit models", *Review of Economics and Statistics*, Vol. 67 No. 1, pp. 134-139.

Further reading

Goodwin, B.K. and Ker, A.P. (1998), "Nonparametric estimation of crop yield distributions: implications for rating group-risk crop insurance contracts", *American Journal of Agricultural Economics*, Vol. 80 No. 1, pp. 139-153.

About the authors

Dr Ashley Elaine Hungerford is a Research Agricultural Economist at the Economic Research Service – United States Department of Agriculture. Dr Ashley Elaine Hungerford is the corresponding author and can be contacted at: ashley.e.hungerford@ers.usda.gov

Dr Barry Goodwin is a William Neal Reynolds Distinguished Professor in the Departments of Economics and Agricultural and Resource Economics at the North Carolina State University.