How High the Hedge:
Relationships between Prices and Yields
in the Federal Crop Insurance Program

A. Ford Ramsey, Barry K. Goodwin, and Sujit K. Ghosh

The theory of the natural hedge states that agricultural yields and prices are inversely related. Actuarial rules for U.S. crop revenue insurance assume that dependence between yield and price is constant across all counties within a state and that dependence can be adequately described by the Gaussian copula. We use nonlinear measures of association and a selection of bivariate copulas to empirically characterize spatially-varying dependence between prices and yields and examine premium rate sensitivity for all corn producing counties in the United States. A simulation analysis across copula types and parameter values exposes hypothetical impacts of actuarial changes.

Key words: copulas, dependence, revenue insurance, risk management

Economic theory implies that agricultural yields and prices are inversely related; this relationship is known as the natural hedge. Strength and form of dependence are not specified by theory and must be empirically determined. The degree of inverse dependence between yields and prices is of practical importance in the federal crop insurance program. Most of the liability in the program derives from revenue insurance policies. Rating such policies requires a distribution function for crop revenue that is constructed in the federal program as a joint distribution function across yield and price. The joint distribution can be uniquely characterized by the marginal distributions of yield and price and an associated copula function.

The U.S. Department of Agriculture's Risk Management Agency (RMA) is charged with ensuring the actuarial fairness of policies offered through the federal crop insurance program. In pricing revenue insurance policies, the RMA assumes that correlation (and, implicitly, other dependence measures) between yield and price is fixed across counties within a state. The RMA also assumes that the dependence structure between yield and price can be adequately captured by a Gaussian copula. In effect, the same copula model is used for all counties within a state. Both assumptions constitute *a priori* beliefs about the natural hedge.

This work investigates the practical consequences of these assumptions on crop insurance premium rates. Through a large-scale empirical study, we examine the sensitivity of premium rates to choice of copula function and marginal distributions. Differences in premium rates are mapped to current insured values at all coverage levels providing a consistent dollar value for collected premiums under different actuarial schemes. Using simulation techniques, we simultaneously vary aspects of the copula and the marginal distributions to ascertain their combined effects on the distribution of crop revenue. While the form of the best-fitting copula varies across counties,

A. Ford Ramsey (corresponding author) is an assistant professor in the Department of Agricultural and Applied Economics at Virginia Tech. Barry K. Goodwin is the William Neal Reynolds distinguished professor in the Department of Economics and Department of Agricultural and Resource Economics at North Carolina State University. Sujit K. Ghosh is a professor in the Department of Statistics at North Carolina State University.

Review coordinated by Anton Bekkerman.

the impact on premium rates and resulting subsidy is relatively modest. Our results support the continued use of RMA assumptions in the pricing of revenue insurance.

Copulas have been applied to a number of problems in finance and economics, including the rating and pricing of crop insurance policies.[1] Bozic et al. (2014) addressed the impact of copula choice on premium rates for dairy margin insurance. We deal with crop revenue insurance, as in Goodwin and Hungerford (2015), but our empirical application is fundamentally different. Goodwin and Hungerford fit multivariate copulas to yields from four counties and the Chicago Mercantile Exchange price. Because they fit multivariate copulas, their models implicitly accounted for dependence across county yields. While dependence across yields is important for examining systemic risk and pricing reinsurance, it adds an additional complication that is not relevant to rating individual policies as currently actualized in the federal crop insurance program. In contrast, we fit bivariate copulas within each county, calculate rate differences at each coverage level, and map the rate differences into the current value insured at each coverage level. We explicitly consider two key assumptions made by RMA in the present rating environment, only one of which Goodwin and Hungerford addressed.

An additional extension beyond earlier work in revenue insurance is a simulation study that simultaneously modifies the marginal distributions and the copula. Because of the small samples that are used in practice, differences in estimated dependence and assumed dependence may result from a lack of precision in statistical estimation. Recognition of this limitation provides additional impetus for the simulation exercise.

Rate Making in Federal Crop Insurance

The U.S. federal crop insurance program functions as a public–private partnership, where private insurers service the policies offered through the program. As of the 2017 fiscal year, 16 insurers were approved to provide insurance coverage under the U.S. Department of Agriculture's Standard Reinsurance Agreement. The RMA sets the parameters of the underlying policies and determines which policies will be offered through the federal program. The public–private aspect of federal crop insurance has been a component of the program since 1981 and is often cited as a major reason for the growth of insurance uptake since that time (Glauber, 2004).

The federal program offers many types of insurance; the most popular policies are revenue insurance policies, which pay out on lost revenue. Revenue is a function of prices and yields; in federal crop insurance, pricing revenue policies is an actuarial problem of the design of a joint distribution of prices and yields. The joint distribution maps to a distribution of revenue. If price and yield have an inverse relationship, then revenue variance will be less than revenue variance under independence of these components (Bohenstedt and Goldberger, 1969). The loss on the most basic revenue insurance policy is given by

$$(1) \qquad\qquad \text{Loss} = \max\left(0, Y_P P_P \lambda - Y_H P_H\right),$$

where Y_P and P_P are expected (planting time) yields and prices and Y_H and P_H are realized (harvest time) yields and prices and λ is a coverage level between 0 and 1. At the time the policy is sold, the only stochastic variables in equation (1) are Y_H and P_H. The marginal distributions of these quantities have been extensively studied (Nelson and Preckel, 1989; Goodwin and Ker, 1998; Finger, 2010; Sherrick et al., 2014; Tolhurst and Ker, 2015).

A popular extension of basic revenue insurance is the harvest price option. The loss on such a policy is

$$(2) \qquad\qquad \text{Loss} = \max\left(0, Y_P \max\left(P_P, P_H\right)\lambda - Y_H P_H\right),$$

---

[1] For an introduction to copulas, see the works of Nelsen (1993) and Joe (2015).

with variables similarly defined. The harvest price option adds additional prices to the revenue guarantee.[2] Coble et al. (2010) offer a more extensive discussion of the actuarial methods behind these policies.

One of the primary tasks in developing rates for revenue insurance is modeling the relationship between crop yields and the normalized average price of a futures contract at the Chicago Mercantile Exchange (CME). The normalized price is constructed by dividing a harvest price by a projected price, determined before the growing season begins. This normalized price can be thought of as a return over the growing season and is modeled with a normal distribution following Black and Scholes (1973). According to the theory of the natural hedge, the normalized price should be inversely related to yields. While most studies couch the natural hedge in terms of Pearson correlation, there is little to suggest that dependence between these two quantities must be linear.

Two assumptions in the pricing of federal revenue insurance relate directly to the natural hedge or, more generally, to the structure of dependence. Dependence between prices and yields is assumed to be adequately described by a Gaussian copula, further implying that yields and prices are (asymptotically) tail independent (Nelsen, 1993). The second assumption is that the correlation matrix for the Gaussian copula is fixed within states. For instance, CME prices and corn yields in Erie County, PA, have the same dependence relationship as CME prices and corn yields in Bucks County, PA. In a majority of states, these assumptions jointly imply that the variables are independent of one another.[3] Because the underlying relationship between prices and yields does not respect state boundaries but rather follows a smooth spatial process, such assumptions may result in flawed pricing.

If a large amount of data were available for the estimation of county- or unit-level copulas and yield distributions, a strong argument could be made for the use of entirely data-driven actuarial methods. But at both levels of aggregation, historical information on crop yields may only be available for a small number of years. Furthermore, older data may not be informative for estimating loss distributions. One solution to improve rating efficiency is to pool data from surrounding counties or units. Racine and Ker (2006) were among the first to suggest using data from other counties for rating purposes. Zhu, Goodwin, and Ghosh (2014) used spatial autoregressive models to investigate spatial correlation in yield distributions, while Feng and Hayes (2016) suggested that reinsurers can effectively pool risk across international boundaries. Ker, Tolhurst, and Liu (2016) used Bayesian mixture models to pool information from estimated densities, whereas Park, Brorsen, and Harri (2018) pooled information prior to the estimation of the density. We do not consider pooling data in this application, our intent being to isolate the effect of the copula, but note that efficiencies could be gained if methods were extended to pool across copulas.

Our empirical approach to assessing the suitability of assumptions made by the RMA about the dependence structure is straightforward. We calculate premium rates at various coverage levels under three scenarios: a copula specification based on RMA practice, a Gaussian copula with correlation specified by the RMA, and the form of the copula and dependence allowed to freely vary. In the last scenario, the optimal copula is selected according to the Akaike Information Criterion, as is standard in the copula literature (Joe, 2015). Given the liability at each coverage level in the last year of the sample, we calculate how much additional premium and subsidy would be generated under alternative copula specifications. This allows for regional differences in pricing to be clarified and provides a measure of the aggregate economic impact of hypothetical changes to rating methods.

---

[2] While additional prices are added to the revenue guarantee for the harvest price option, the joint distribution is the same because the projected price is not a random variable. While Revenue Protection and Revenue Protection – Harvest Price Exclusion are available for individual units, analogous policies called Area Risk Protection (ARP) exist at the county level. At both unit levels and the county level, the vast majority of liability is in revenue policies with the harvest price option.

[3] The correlation parameter in many states is assumed to be 0. If the correlation parameter of a Gaussian copula is 0, the copula takes the form of the independence copula and the random variables are independent.

Table 1. Measures of Association and Dependence

Measure	Population Formula	Sample Formula
Pearson correlation	$\rho_P = \dfrac{Cov(X,Y)}{\sqrt{Var(X),Var(Y)}}$	$\rho_P = \dfrac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$
Spearman correlation	$\rho_S = 3(P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0])$	$\rho_S = \dfrac{\sum_i (r_i - \bar{r})(s_i - \bar{s})}{\sqrt{\sum_i (r_i - \bar{r})^2 \sum_i (s_i - \bar{s})^2}}$
Kendall's tau	$\tau = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0]$	$\tau = \dfrac{\sum_{i<j} (\text{sgn}(x_i - x_j)\,\text{sgn}(y_i - y_j))}{\sqrt{(T_0 - T_1)(T_0 - T_2)}}$
Hoeffding's D	$D = \int (F_{XY} - F_X F_Y)^2 dF_{XY}$	$D = 30 \dfrac{(n-2)(n-3)D_1 + D_2 - 2(n-2)D_3}{n(n-1)(n-2)(n-3)(n-4)}$
Distance correlation	$dCor(X,Y) = \dfrac{dCov(X,Y)}{\sqrt{dVar(X)dVar(Y)}}$	$\rho_D = \begin{cases} \dfrac{V^2(x,y)}{\sqrt{V^2(x)V^2(Y)}} & V^2(X)V^2(Y) > 0 \\ 0 & V^2(X)V^2(Y) = 0 \end{cases}$

Notes: Barred variables are sample means. r_i and s_i are the ranks of x_i and y_i, respectively. For Kendall's tau, $T_0 = n(n-1)/2$, $T_1 = \sum_k t_k(t_k - 1)/2$ and $T_2 = \sum_l u_l(u_l - 1)/2$. In this case, t_k is the count of tied x values in the kth group of tied x values, and u_l is the number of tied y values in the lth group of tied y values. The sgn($\cdot$) function is defined as 1 if x is positive, 0 when x is 0, and -1 otherwise. For Hoeffding's D, $D_1 = \sum_i (q_i - 1)(q_i - 2)$, $D_2 = \sum_i (r_i - 1)(r_i - 2)(s_i - 1)(s_i - 2)$, and $D_3 = \sum_i (r_i - 2)(s_i - 2)(q_i - 1)$. The term q_i is the bivariate rank of point i, which is 1 plus the number of points with both x and y less than the value of the ith point.

Measures of Dependence and the Copula

In agricultural contexts, copulas have primarily been used to model spatial dependence in yields or to model dependence between random variables in crop revenue insurance, margin insurance, and whole-farm insurance (Bozic et al., 2014; Bulut and Collins, 2014; Ahmed and Serra, 2015; Goodwin and Hungerford, 2015; Feng and Hayes, 2016). Although applications in these studies have varied, the gist has been that the use of a Gaussian copula implies tail independence among the variables. Most authors find evidence that the Gaussian copula is not a suitable model by either selecting a different form based on fit criteria or rejecting a null hypothesis that the Gaussian is the proper copula. However, Hungerford and Goodwin (2014) showed that the form of the copula and its parameter values can be highly variable when using small samples. In spite of variability due to sample size, if the impact of different copulas on premium and subsidy is small, this can be taken as evidence that RMA assumptions about the structure of dependence may be justified. No previous studies have mapped rates to collected premiums at all available coverage levels, providing measures of economic impact across a wide range of copula functions.

Many nonlinear measures have been developed to measure association and dependence between two or more variables. These measures vary in their ability to capture different types of dependence and in their computational simplicity. Statistics of association are useful tools for exploring the dependence structure without making assumptions on functional form. They can also be used to parameterize copula models. Table 1 gives the forms of five measures of association and dependence. Dependence between price and yield is often measured in terms of the Pearson correlation coefficient, which captures the extent of linear association between variables. It is widely known that Pearson correlation is only appropriate for measuring linear relationships. Furthermore, it is fallacious to say that a joint distribution can always be defined from its marginal distributions and Pearson correlation (Embrechts, McNeil, and Straumann, 2002). The notion is false, except in some special cases such as when the joint distribution of the variables is Gaussian.

Other nonparametric measures of association can capture nonlinear and nonmonotonic dependence relationships. Monotone dependence occurs when, if one variable increases, the other variable tends to increase (or decrease) as well. Criteria for measures of monotone association, or bivariate concordance, were discussed in Scarsini (1984). Spearman's rank correlation and Kendall's tau satisfy these criteria, while Pearson correlation does not. From Table 1, it should be clear that Spearman's rank correlation has the same formula as Pearson correlation, but with the calculation based on ranks instead of the levels of the variables. Because Spearman rank correlation is only a function of ranks of the data, it is invariant to monotone transformations of the variables and does not rely on an assumption of linearity.

Kendall's tau measures the number of concordant and discordant pairs of observations in the data. If all pairs are concordant, then the statistic in Table 1 will equal 1. This indicates perfect concordance and positive dependence. If the pairs are perfectly discordant, then the value of Kendall's tau is −1, indicating negative dependence. The population versions of Kendall's tau and Spearman rank correlation have the same formula, which is essentially Nelsen's (1993) concordance function.

Hoeffding's D dependence coefficient facilitates tests of independence even when the alternative is nonmonotonic. Hoeffding's D was developed from the definition of independence, which is that two variables are independent when $F_{XY} = F_X F_Y$. Hoeffding (1948) considered the distance between the joint distribution and the product of the marginals, which should be 0 when the variables are independent. Hoeffding's D is an unbiased estimator of this distance and thus can be used to test for independence against a wide variety of alternatives.

The last measure we suggest for gauging dependence between yield and price is distance correlation, as proposed by Szekely, Rizzo, and Bakirov (2007) and applied by Szekely and Rizzo (2009). Distance correlation is a generalization of Pearson correlation, and for a bivariate normal distribution, distance correlation is a deterministic function of Pearson correlation. Much like

Spearman rank correlation, distance correlation can be computed on ranks of the data, leading to a rank distance correlation statistic. Because distance correlation is capable of finding complicated dependence structures, there is little gain in transforming the data to ranks. As demonstrated in Szekely and Rizzo, distance correlation performs well in detecting nonlinear dependence even when the sample size is relatively small. Simon and Tibshirani (2014) found distance correlation to have more power compared to Pearson correlation and the maximal information coefficient.

Even though measures of association provide quantification of the magnitude and direction of dependence, it is still the case that a joint distribution must be formed and evaluated for actuarial purposes. One approach to forming the distribution and incorporating nonlinear dependence is through the use of copula functions. Sklar's (1959) theorem is the fundamental existence theorem for copulas, although work on standardized distributions predates Sklar's theorem. Let F be a joint distribution function with univariate marginal distribution functions $F_1, \ldots, F_d$. Then there exists a copula function $C : [0,1]^d \to [0,1]$ such that

$$(3) \qquad\qquad F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)),$$

where $x_1, \ldots, x_d \in \mathbb{R}$ are random variables. Copulas provide a way of constructing joint distributions and simultaneously describing scale-free or rank dependence.

By inversion of the joint distribution in equation (3), the copula function can be written as

$$(4) \qquad\qquad C(u_1, \ldots, u_d) = F(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)),$$

where $F_1^{-1}, \ldots, F_d^{-1}$ are 1-dimensional quantile functions and $u_1, \ldots, u_d \in [0,1]$. The copula is parameterized by a vector ρ consisting of dependence parameters. Given that the copula is itself a joint distribution function, it satisfies all of the criteria for a joint distribution function.

Copulas allow for a joint distribution function to be described in terms of the marginal behavior of the underlying variables and their dependence structure. Several features of the dependence structure may be of interest. Popular measures of dependence like the Pearson correlation coefficient are often incapable of discriminating among differences in these features. From an empirical perspective, the choice of a specific parametric copula function involves *a priori* assumptions about the nature of dependence. Some of the most important properties relating to copulas are symmetry, radial symmetry, joint symmetry, associativity and Archimedeanity, max-stability, and tail dependence (Li and Genton, 2013).

A detailed discussion of symmetry concepts can be found in Nelsen (1993). A copula is radially symmetric if

$$(5) \qquad\qquad C(u_1, u_2) - C(1 - u_1, 1 - u_2) + 1 - u_1 - u_2 = 0$$

for all $(u_1, u_2) \in [0,1]^2$. Radial symmetry is often called tail symmetry. Two points that are equidistant from the middle of the unit square and lie on rays pointing in opposite directions from the middle of the unit square will have the same copula density. Intuitively, the most interesting type of symmetry is radial symmetry because radially asymmetric copulas have different dependence relationships in the lower and upper tails of the distribution. These differences in dependence can often be motivated by appeals to economic theory. From a practical standpoint, insurers and financial analysts are usually interested in behavior in the lower tail, as this is where worst-case portfolio losses occur.

An Archimedean copula is defined as a function $C : [0,1]^d \to [0,1]$ where

$$(6) \qquad\qquad C(u_1, \ldots, u_d) = \theta(\theta^{-1}(u_1) + \cdots + \theta^{-1}(u_d))$$

and the function $\theta : [0, \infty) \to [0,1]$ possesses several properties. Namely, $\theta(0) = 1$, $\theta(\infty) = 0$, and $(-1)^i \theta^{(i)} \geq 0$, for $i = 1, \ldots, \infty$, where $\theta^{(i)}$ denotes the ith derivative of θ. These conditions imply that θ is a strictly decreasing differentiable function. This function is referred to as the generator of

the copula. Many functions satisfy the requirements on the generator, and so there are many copulas within this class. All Archimedean copulas are associative with

$$C(C(u_1,u_2),u_3) = C(u_1,C(u_2,u_3)) \tag{7}$$

for all $(u_1,u_2,u_3) \in [0,1]^3$.

Copulas can also be distinguished by their tail dependence. The upper-tail dependence coefficient is defined as

$$\lambda_U = \lim_{u \to 1} Pr(U_1 \geq u | U_2 \geq u) = \lim_{u \to 1} \frac{1 - 2u + C(u,u)}{1 - u}, \tag{8}$$

while the lower-tail dependence coefficient is defined as

$$\lambda_L = \lim_{u \to 0} Pr(U_1 \leq u | U_2 \leq u) = \lim_{u \to 0} \frac{C(u,u)}{u}. \tag{9}$$

A copula has tail dependence if and only if the respective coefficients in equations (8) and (9) are nonzero. Tail dependence is an asymptotic concept, and when the tail dependence coefficient is close to 1, large values of the variables occur together. For a given copula model, the tail dependence coefficients are functions of the underlying dependence parameters. The choice of the copula family is intrinsically a choice of theoretical tail dependence.

The five most popular parametric copulas are the Gaussian, t, Gumbel, Clayton, and Frank (Joe, 2015). With various rotations, these copulas can capture a wide range of dependence behavior. Table 2 gives the forms of the copulas, parameter restrictions, and a summary of their properties. Copula parameters can be estimated from observed data by calibration- or maximum-likelihood-based approaches. With parametric marginal density functions $f(x_i|\rho_i)$, the density function of the joint distribution F can then be given as

$$f(x_1,\ldots,x_d) = c(F_1(x_1),\ldots,F_d(x_d)) \prod_{i=1}^{d} f_i(x_i). \tag{10}$$

For an *i.i.d.* random sample of size n, the log-likelihood function is then

$$L = \sum_{j=1}^{n} \log f(x_{j1},\ldots,x_{jd}). \tag{11}$$

Substituting into equation (11) from equation (10) results in

$$L = \sum_{j=1}^{n} \log c(F_1(x_{j1}),\ldots,F_d(x_{jd})) + \sum_{i=1}^{d} \sum_{j=1}^{n} \log f_i(x_{ji}). \tag{12}$$

The first term on the right of equation (12) is the contribution of the dependence structure to the log-likelihood, and the second term is the contribution of the univariate marginal distributions. One could estimate the joint distribution by maximizing the likelihood in equation (12) over the parameters of the copula and the univariate marginal distributions. Joe and Xu (1996) suggested an alternative procedure, now known as Inference for Margins (IFM), in which estimates of the parameters of the marginal distributions are first obtained by maximum likelihood estimation. Then the copula contribution to the likelihood is maximized, holding the univariate parameters fixed at their estimates from the first stage. Both approaches result in estimators that are consistent and asymptotically normal.

Because many functions meet the criteria of a copula, there are many choices for modeling dependence. A general modeling approach is to compare several parametric copulas or to utilize a

Table 2. Properties of Five Popular Copulas

Copula	Tail Dependence	Radial Symmetry	Formula	Parameters
Normal	No	Yes	$C_N(u, \Sigma) = \Phi(\phi^{-1}(u_1), \phi^{-1}(u_2))$	$\Sigma \in [-1, 1]^{2 \times 2}$
t	Yes	Yes	$C_t(u, \Sigma, v) = \mathbf{t}(t^{-1}(u_1), t^{-1}(u_2))$	$\Sigma \in [-1, 1]^{2 \times 2}, v \in [0, \infty)$
Gumbel	Yes	No	$C_G(u, \theta) = \exp\left(-\left(\sum_{i=1}^{2}(-\log u_i)^{\theta}\right)^{1/\theta}\right)$	$\theta \geq 1$
Clayton	Yes	No	$C_C(u, \theta) = \left(\sum_{i=1}^{2} u_i^{-\theta} - 2 + 1\right)^{-1/\theta}$	$\theta > 0$
Frank	No	Yes	$C_F(u, \theta) = \frac{1}{\theta} \log\left(1 + \frac{\prod_{i=1}^{2}(\exp(-\theta u_i) - 1)}{(\exp(-\theta) - 1)}\right)$	$\theta \in \mathbb{R}/0$

fully nonparametric copula. Consideration of strictly parametric copulas is often limited to a small range of copulas that may or may not be capable of capturing important aspects of the dependence structure. While several goodness-of-fit tests for parametric copulas exist, these tests rarely provide any guidance as to the appropriate choice of copula if the proposed copula is rejected (Genest, Remillard, and Beaudoin, 2009; Huang and Prokhorov, 2014; Kojadinovic, Yan, and Holmes, 2011). Such tests may also require large samples when based on comparisons of the estimated copula with empirical (nonparametric) copulas.

As noted in Joe (2015), the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) can be used to compare parametric copula models. Because the copula has tail dependence and radial symmetry features, in addition to its general strength of dependence, AIC and BIC provide measures of fit that consider all dependence aspects. However, they only provide a ranking of proposed copulas. It could be the case that all of the models under consideration are inadequate for a given problem. A related likelihood-based test from Vuong (1989) has been applied to copulas. The Vuong test can be used to determine whether one of two copula models is a better fit, but—like AIC and BIC—it does not assess the adequacy of the models under consideration.

Empirical Application

Our data come from the USDA National Agricultural Statistics Service and include annual yields in all corn-producing counties between 1980 and 2015. In some counties, yields are either not reported or 0, often to protect the identity of producers. Counties with less than 15 years of available yields were dropped from the analysis.[4] Raw yields in each county were detrended using robust regression similar to RMA procedures.[5] The robust regression technique we apply, M-estimation, is a more general case of MM-estimation (Huber, 1973). Finger (2010) shows that errors from detrending are likely to be small or moderate in size. Results using other detrending methods (linear regression and locally weighted scatterplot smoothing) are shown in the online supplement (www.jareonline.org) and indicate little aggregate sensitivity to the detrending procedure. Prices were obtained from the CME and normalized as the logarithm of the October average price (P_O) for a December corn futures contract divided by the February average price (P_F) for the same December futures contract. The joint distribution was estimated over this normalized price $\ln(P_O/P_F)$, or return, and the detrended county yield in each year.

The RMA makes two assumptions when rating revenue insurance policies. First, the correlation between prices and yields within a state is assumed to be fixed across counties. Even though counties vary in yields, each county is assumed to have the same yield–price correlation as all other counties in that state. Second, the dependence structure in all counties is specified using a Gaussian copula. We investigate the suitability of these assumptions with respect to the actuarial fairness of the federal crop insurance program and provide estimates of insurance rates under a model that allows for county variation and more general dependence structures.

The exercise of rating crop insurance policies in the federal program is a task of both inference and prediction. Inference because the variables underlying such policies are economic in nature and can be related to structural parameters of agricultural markets. Prediction because insurance is always concerned with the assessment of future losses and their probabilities of occurrence. We view these sometimes competing tasks through the lens of model selection. The number of

---

[4] Both the length of the time series and the choice of threshold for inclusion in the analysis were motivated by practical aspects of copula estimation. The models are difficult to estimate with any amount of precision below the threshold, and we hesitate to utilize longer yield series due to possible structural change in relationships between prices and yields. Little guidance is available on the optimal length of time series for estimating yield distributions, although ongoing research is being directed toward this question (Liu and Ker, 2018). A useful future exercise would consider the optimal window size for estimating the copula functions. We considered including counties with less than 15 years of available yield data and found that results were robust to their inclusion.

[5] For counties with less than 26 years of data, the RMA detrends using robust regression (L1) with a linear trend and no knots, while those with more available yield data use a 1-knot spline model.
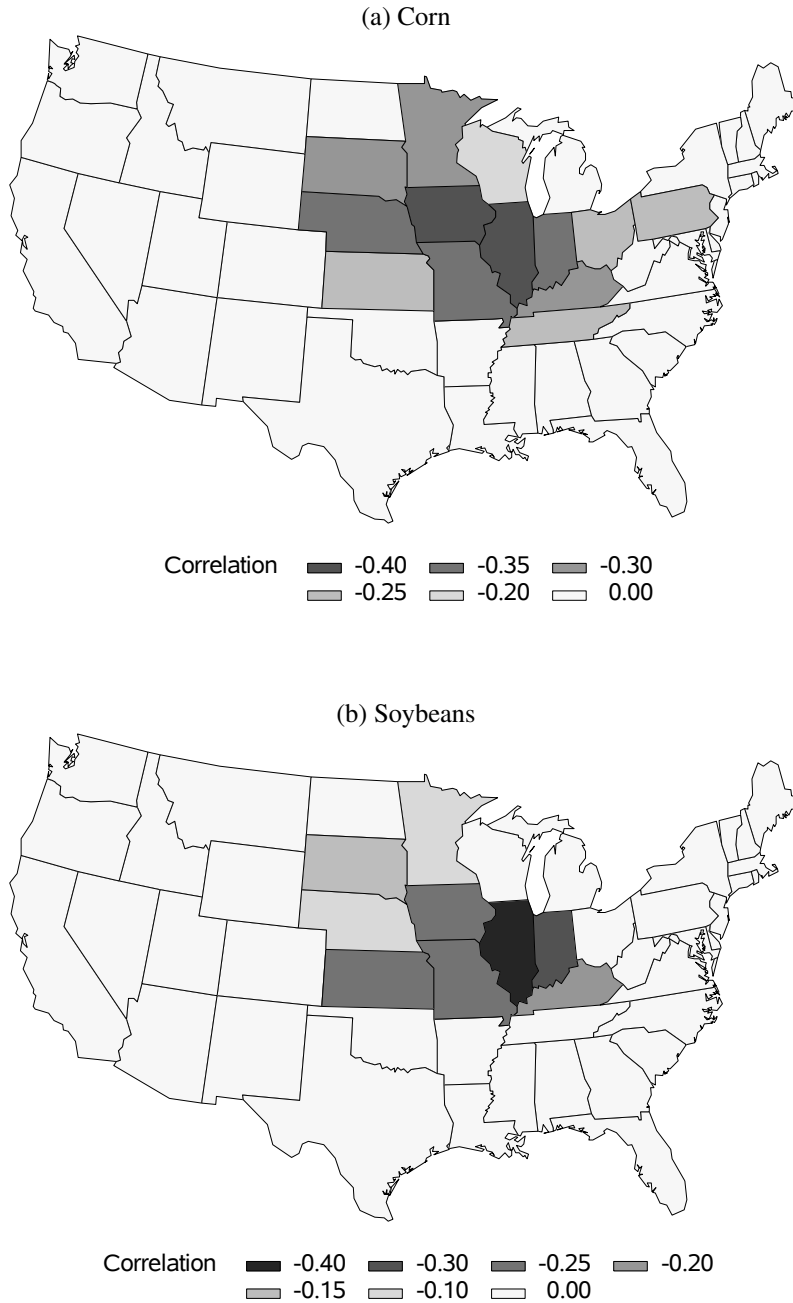
possible dependence models for yield and price relationships is infinite. As in most analyses, the set of alternative models results from *a priori* judgment. Current pricing methods make potentially strict assumptions, motivated by the theory of the natural hedge, on the types of models allowed in this set. As noted by Leamer (1983), researchers should not be overconfident in any single theory. We also make *a priori* judgments since we only consider a restricted number of copula models. However, our proposed models cover a wide range of dependence possibilities.

Figure 1 shows assumed correlations for two crops for which revenue insurance is currently available through the federal program. The Pearson correlations used in constructing premium rates for corn range from 0 in most states to −0.4 in Illinois and Iowa. The pattern of correlation follows, very roughly, the total quantities of corn produced in different states. However, there are some state boundaries with abrupt changes in assumed correlation. One instance is the boundary between Michigan and Indiana. It seems unlikely that yields in northern Indiana are strongly related to Chicago prices while those in southern Michigan are completely independent. Corn acreage, production, and yields in these counties are similar. Figure 2 shows estimated Pearson correlations between yield and price for all corn-producing counties in the sample. The estimated correlations are similar to those assumed by the RMA in states like Iowa and Illinois, but there are large differences in the South and along the Mississippi River. The RMA assumes 0 correlation in all counties in Mississippi, while counties near the Gulf of Mexico have strong negative dependence.

Strong negative correlation is realized in counties around New York City. One possible explanation, which also applies to the negative dependence observed in southern Texas and on the Mississippi, is that these areas are close to major ports and shipping lanes. While this seems plausible, it does not align with the weak dependence observed in the Carolinas. The ports of Wilmington, Charleston, and Savannah have large enough capacity to handle outbound shipments. One distinguishing feature for the Carolinas is that both states are grain-deficit and use most of their corn production to feed hogs and other livestock. Note the distance correlations shown in Figure 2. Distance correlation provides a powerful means of assessing dependence, which may be nonlinear. According to this measure, strong dependence is observed in the major growing regions, particularly Illinois, Missouri, and Iowa. There are also strong pockets of dependence in southern Texas, along the Mississippi, in New Jersey, and in northern Wisconsin. In contrast, dependence as measured by distance correlation is weak in Indiana and Ohio.

Positive Pearson correlation is estimated in many counties in eastern North Carolina, South Carolina, Minnesota, and Wisconsin. Figure S1 provides scatterplots of normalized yields and prices from selected counties with strong positive or negative correlation, and positive relationships hold when dependence is measured using Spearman's rho or Kendall's tau (see Figures S2 and S3). Findings of positive Pearson correlation may be spurious given small sample sizes. Tests based on county-level data are subject to the modifiable areal unit problem, as explained in Gehlke and Biehl (1934). Bias can be introduced because the boundaries of the counties are drawn arbitrarily (from a statistical perspective). But as the policies are written at the county level, without any pooling, it is the county results that are of interest. To investigate the variability of the correlation estimates, we implemented Efron's (1979) bootstrap. In every county, 1,000 bootstrap samples were drawn to construct a sampling distribution of Pearson correlation.

We constructed a 90% confidence interval using the sampling distribution in each county and recorded whether the correlation assumed by the RMA was within this interval. The correlation assumed by the RMA fell outside of the interval in nearly 20% of the counties in the sample. There are several noteworthy results. In general, the bootstrap results align with those from the map of distance correlation (see Figure S5). It appears that findings of positive correlation in North Carolina are indeed spurious, but there are still several counties in South Carolina where the empirical estimates hold. Areas with counties that do not support the RMA's assumptions include northern Iowa, Minnesota, Wisconsin, Indiana, and Ohio. Most of the clusters occur along or near state boundaries, highlighting the problems caused by imposing more or less arbitrary areal unit definitions on underlying spatial processes. We do not take the bootstrap results as definitive

(a) Corn

Correlation   ▉ -0.40   ▉ -0.35   ▉ -0.30
              ▉ -0.25   ▉ -0.20   ▢ 0.00

(b) Soybeans

Correlation   ▉ -0.40   ▉ -0.30   ▉ -0.25   ▉ -0.20
              ▉ -0.15   ▉ -0.10   ▢ 0.00

Figure 1. Pearson Correlation for Current Actuarial Methods

Source: Adapted from Goodwin et al. (2014).

evidence that RMA assumptions are flawed, because the sample size is small in any one county. We simply note that, if assumptions on correlation are relaxed, empirically driven estimates will diverge from what is currently imposed. Moreover, there is no theoretical or intuitive support for constant correlation within states.

Copulas were fitted to the data from each county using semiparametric maximum pseudo-likelihood techniques. Normalized prices and yields were first converted to the uniform scale using the empirical distribution function. Then the copula portion of the likelihood was maximized over the

(a) Pearson Correlation: Price vs. Yield

(b) Distance Correlation: Price vs. Yield
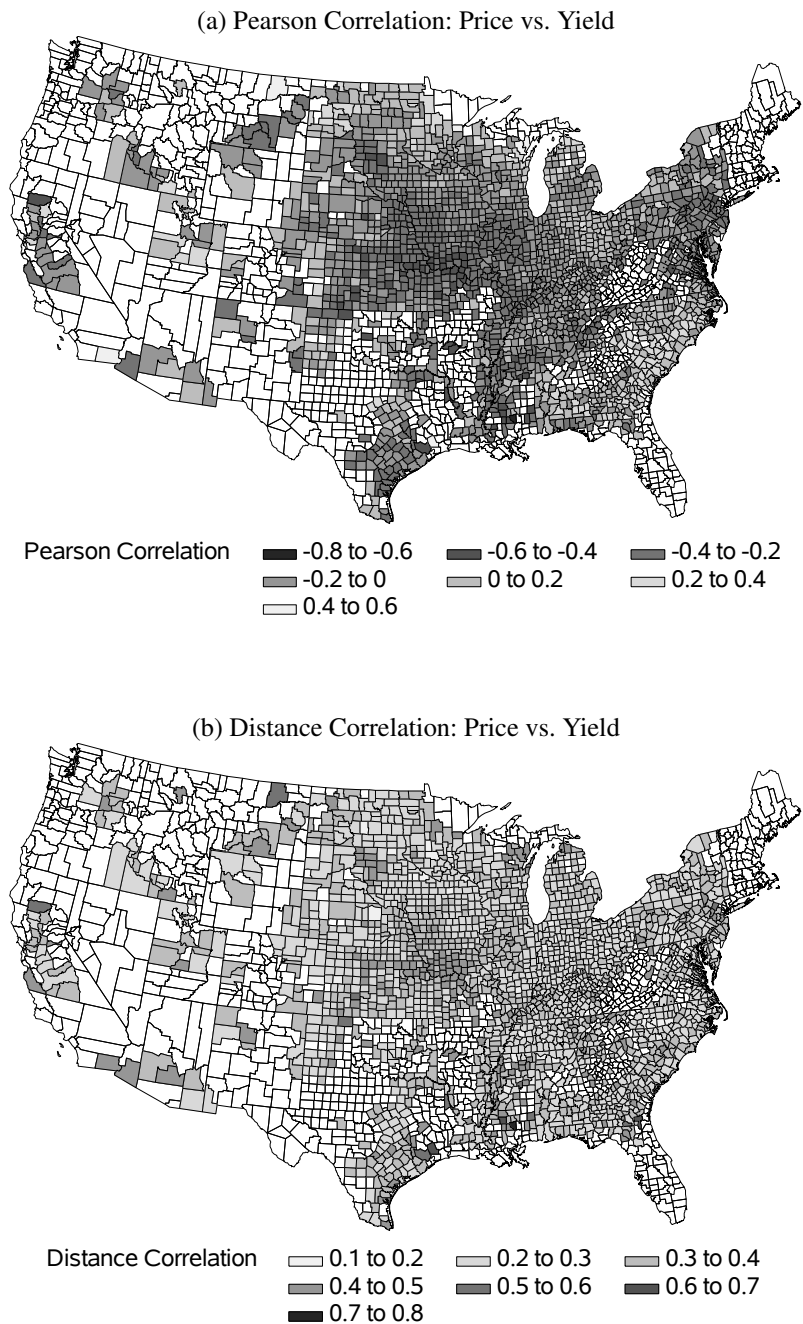
Figure 2. Measures of Association and Dependence

uniform data. We estimate normal, t, Gumbel, Clayton, Frank, 90-degree-rotated Gumbel, and 90-degree-rotated Clayton copulas. The best-fitting copula was selected according to AIC. The selected copula in each county is shown in Figure 3, with frequency of selection in Table S1. Selection based on BIC was also considered, but we found that the optimal copula differed between the two criteria in only 3%–4% of counties. The selected copula function was also largely consistent across detrending methodologies.

Copula   Clayton   Clayton Rot.   Frank
            Gumbel   Gumbel Rot.   Normal
            t
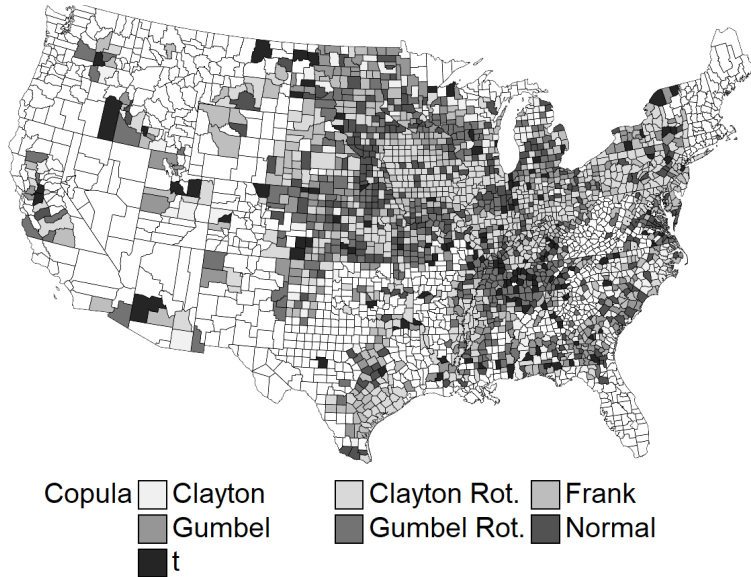
Figure 3. Best-Fitting Copula by County

Overall, there is a large amount of variation in the selected copula. In areas with negative dependence, the best-fitting copula is usually either the Frank or normal copula. Areas with positive dependence are often best described with a Clayton or *t* copula. As data determining the direction of dependence also determine the strength of tail dependence and type of joint symmetry, geographic clustering is to be expected. Frank and Gaussian copulas are similar in terms of joint symmetry and tail dependence. However, an unrotated Clayton copula implies joint occurrence of low yields and low prices. Simultaneously low yields and prices result in the largest losses under a revenue insurance policy. We might expect that counties where the Clayton copula is selected will have higher premium rates as a result of varying copula models.

We calculated premium rates in each county by assuming a February average price of $4/bushel of corn and a coverage level of 80%. 1,000 draws were taken from each of a normal copula with correlation given by Figure 1, a normal copula with correlation estimated by maximum likelihood, and the best-fitting copula estimated by maximum likelihood. Each set of uniform yield draws was then passed through the quantile function of a Weibull distribution and the uniform price draws through the quantile function of a normal distribution. The parameters of the Weibull distribution were estimated via maximum likelihood using detrended yields in each county.[6] This procedure results in 1,000 simulated yields and prices for each copula model in each county. Using these sets of yields and prices, a distribution of revenue can be formed and losses under the artificial revenue policy can be obtained.

Figure S6 maps the difference in rates between models based on a Gaussian copula with data-driven correlation and a Gaussian copula with correlation set according to Figure 1. The difference in rates resulting from misspecified correlation is small. A histogram and kernel density of the differences is shown in Figure 4. Rate differences are grouped near 0, although there are a few outliers. Even in the most extreme cases, differences in Pearson correlation only result in rate differences of 1%–1.5%. In spite of the small difference in realized rates, Figure S6 shows evidence of systematic variation across space. In counties where the estimated Pearson correlation is more

---

[6] Sherrick et al. (2014) provide support for the use of a Weibull distribution. We also considered using a nonparametric distribution for the marginals and found that rates differed substantially depending on the specified marginal distribution. When viewed against the results from the varying copulas, it appears that the marginal distribution of yields is a more important choice than the copula function when rating revenue insurance.

(a) Normal Cop. Varying Corr. - Normal Cop. RMA Corr.
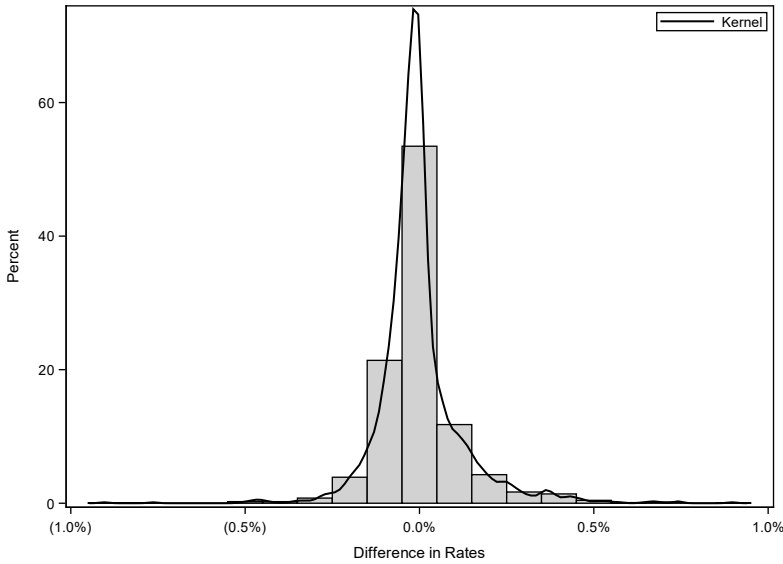
(b) Optimal Cop. - Normal Cop. RMA Corr.
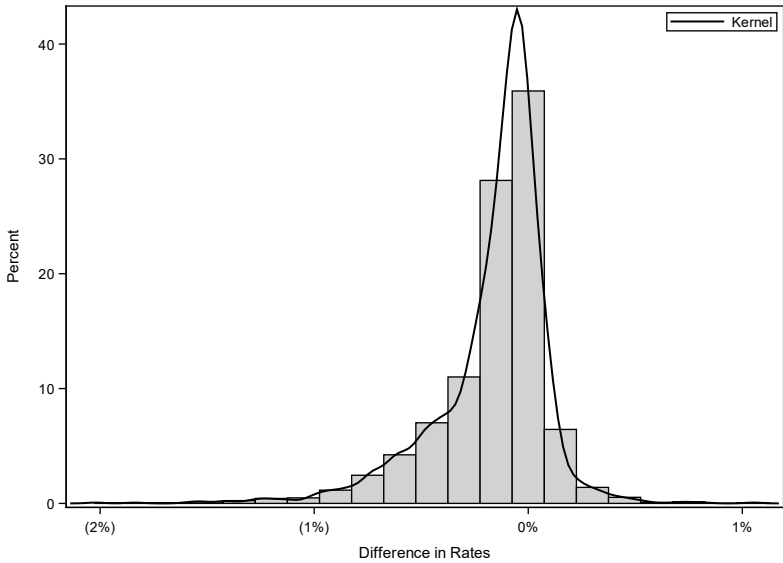
Figure 4. Histogram and Kernel Density of Rate Differences

negative than the RMA correlation, the difference is usually positive. In locations where the assumed correlation is more strongly negative, at least when compared to the estimated correlation, the difference in rates is often negative.

The second panel of Figure S6 maps the difference in rates between the model based on the best-fitting copula and the model with Gaussian copula and fixed correlation. The kernel density in Figure 4 shows that when the copula is allowed to vary with strength of dependence, premium rates in many counties are lower than the rates under fixed correlation and a normal copula. This difference is largest in counties where there is evidence of negative dependence and the rotated

Clayton copula is best fitting. Spatial patterns exhibited in the data suggest that randomness does not entirely drive these results. There is a clear distinction between the shapes of the densities in Figure 4a and Figure 4b. Figure 4b suggests a negatively skewed distribution of the difference in rates.

We took the total value insured in each county in 2015 at each coverage level and multiplied this amount by the premium rate implied by different copulas at the given coverage level. The difference in this value was summed within counties to provide a measure of economic value from rating changes and can be thought of as premium collected. Maps in Figure 5 show the difference in premium collected when using a model with a flexible copula specification versus RMA assumptions. If RMA allowed correlation to vary but maintained the use of a Gaussian copula, they would collect roughly $9 million less in premiums. If the copula were allowed to vary, premiums collected would decrease by around $64 million. While these numbers are not exact (because some counties had insured value but not enough data to calculate rates), they provide a general sense of the economic impact of rate changes. The estimates are robust to detrending method (see Table S2). Both of these aggregate changes should be viewed as economically insignificant, as total premium collected by RMA in 2015 for corn revenue protection policies exceeded $3.4 billion.

Simulation Analysis

While empirical applications illuminate many aspects of the crop insurance program, we are also interested in determining the effects that different copulas can have on rates more generally. The preceding section only considered revenue insurance policies for all corn-producing counties at current levels of insured value. Simulation provides a means of drawing conclusions that can be applied to a wider variety of crops. With the advent of the supplemental coverage option, which essentially allows for increased coverage levels above 80%, a natural question is how rates respond to changes in the copula model across the range of possible coverage levels. To this end, we simulated insurance rates across a variety of copulas, strengths of dependence, and parameter values. Our basic inquiry is to determine the conditions under which the chosen copula model has the greatest impact on pricing.

A price of $5/bushel of corn is assumed for the Revenue Protection – Harvest Price Option policies. To specify the dependence relationships, 10,000 draws are taken from each copula model across 13 Pearson correlations from −0.6 to 0.6. The marginal distribution of yields is the normalized beta with domain 0–300, and the first shape parameter is 2. The second shape parameter takes values of either 4 or 8, which corresponds to a short or long right tail. The log-price distribution is normal and parameterized by implied volatility of 20 and 40. Lastly, rates were evaluated at coverage levels from 0.50 to 0.90. With variation in marginal parameters, dependence models, and coverage levels, we price a total of 260 policies through the simulation.

Each of the surface plots is constructed by smoothing across 65 input points using a bivariate spline (Meinguet, 1979). Figure S7 shows the rates obtained from each copula model across a range of coverage levels and Pearson correlations. As expected, the minimum and maximum rates obtained from the normal copula model are exceeded in both cases by the rates obtained from the *t* copula. The maximum rate under the normal exceeds that of the 90-degree-rotated Archimedean copulas, while the minimum rate is less than the minimum rates of the same copulas. The Frank and normal copulas are similar in price, which results from the ability of these copulas to capture positive and negative dependence and their lack of tail dependence. Across all of the models, rates at the 90% coverage level are nearly double those at lower coverage levels.

Figure S8 shows differences in rates of various copula models compared to the Gaussian copula currently in use in the federal crop insurance program. The differences between the *t* and Frank copula are fairly small. The largest deviations occur between the Gumbel and Clayton copulas and the normal. When the true direction of dependence is negative, the normal copula generates rates that are higher than either of the Archimedean copulas. When dependence is positive, it underprices

(a) Normal Cop. Varying Corr. - Normal Cop. RMA Corr.
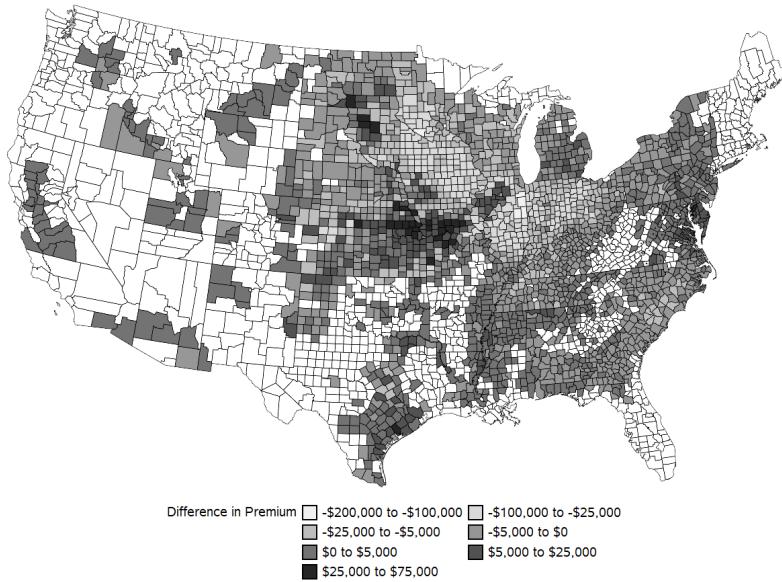
(b) Optimal Cop. - Normal Cop. RMA Corr.
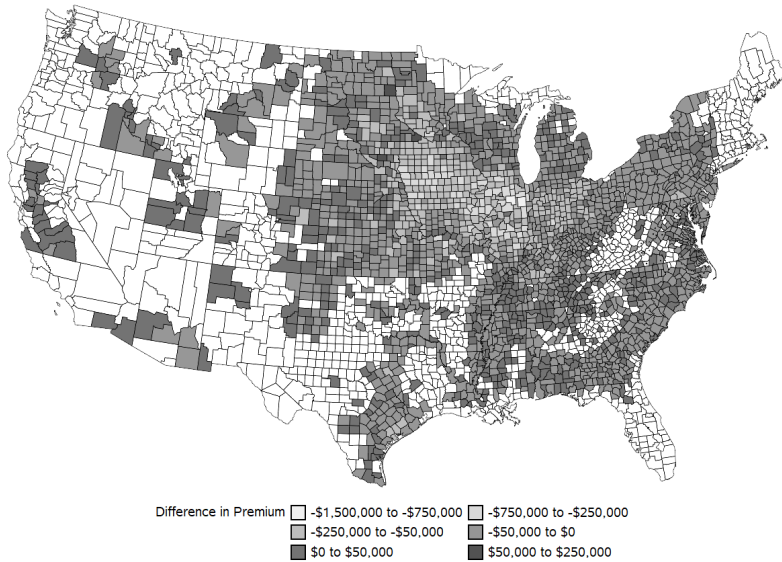
Figure 5. Difference in Premium Collected

the insurance. The rate differences are maximized when correlation is positive for the Archimedean copulas and negative for the rotated Archimedean copulas.

We also considered varying the assumed price volatility and the form of the marginal distribution of yields. Figure S9 contains plots of the same differences with a high price-volatility factor, while Figure S10 shows the same with a long right tail in the marginal distribution of yields. In general, increased price volatility retains the same pattern of rate differences, but extreme differences in rates

are accentuated. For instance, the maximum difference between a rotated Clayton copula and the normal copula at correlation -0.6 is nearly 100% larger under high price volatility. A long-right-tailed distribution shifts the pattern of rate differences from high to lower coverage levels.

Clearly, the largest rate differences occur when the copula is the Gumbel or Clayton (or a rotated variant) and the dependence parameter is exceptionally large in magnitude. The Gumbel or Clayton is frequently selected according to fit criteria in the corn counties examined above, but high or low strength of dependence is rarely observed. The vast majority of counties have Pearson correlation between -0.4 and 0.4. The scenarios that generate large differences in rates are rare, and the simulation supports the conclusion from the empirical analysis; that is, the economic impact of allowing the dependence model to vary is simply not very large.

Conclusion

Revenue insurance has assumed a prominent place in the suite of insurance policies available through the U.S. crop insurance program. Rating revenue insurance requires a joint distribution of price and yield that can be uniquely specified by marginal distributions and a copula function. Current actuarial rules assume that dependence between price and yield is adequately captured by a Gaussian copula and that correlation is constant within states. These assumptions are imposed *a priori*, motivating the question of whether, if such assumptions are not statistically justified, they have major impacts on economic outcomes. Such questions are relevant for agricultural policy because any actuarial changes directly impact insurance prices and indirectly impact subsidies paid out through the crop insurance program.

We assess the sensitivity of premium rates to assumptions about the copula function and degree of dependence both empirically and through a simulation analysis. Using common fit criteria for the copula functions, we find that the form of the copula varies across counties and the magnitude of correlation varies within states. Simulation results demonstrate that the form of the copula can have a major impact on premium rates, but only in situations unlikely to occur in reality. For the majority of counties, varying the copula results in relatively small changes in the amount of premium that would be collected. This suggests that the effort necessary to deal with the expanded set of dependence models, when balanced against costs to the Risk Management Agency of rating the policies, is not large enough to economically justify departures from current practice.

Useful future research might be directed toward a method that borrows information across space to inform the copula model. Because prices are obtained from the Chicago Mercantile Exchange, this is analogous to the problem found in yield insurance as it is the yield that varies across counties. However, it is not clear whether a best approach would smooth across the marginal distributions and copula or simply smooth across a distribution of revenue. Results presented here indicate that the effect on insurance pricing may be small in any case, but useful information about price formation and the extent of markets might be obtained.

References

Ahmed, O., and T. Serra. "Economic Analysis of the Introduction of Agricultural Revenue Insurance Contracts in Spain Using Statistical Copulas." *Agricultural Economics* 46(2015):69–79. doi: 10.1111/agec.12141.

Black, F., and M. Scholes. "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy* 81(1973):637–659. doi: 10.1086/260062.

Bohenstedt, G. W., and A. S. Goldberger. "On the Exact Covariance of Products of Random Variables." *Journal of the American Statistical Association* 64(1969):1439–1442. doi: 10.1080/01621459.1969.10501069.

Bozic, M., J. Newton, C. S. Thraen, and B. W. Gould. "Tails Curtailed: Accounting for Nonlinear Dependence in Pricing Margin Insurance for Dairy Farmers." *American Journal of Agricultural Economics* 96(2014):1117–1135. doi: 10.1093/ajae/aau033.

Bulut, H., and K. Collins. "Designing Farm Supplemental Revenue Coverage Options on Top of Crop Insurance Coverage." *Agricultural Finance Review* 74(2014):397–426. doi: 10.1108/AFR-08-2013-0032.

Coble, K. H., T. O. Knight, B. K. Goodwin, M. F. Miller, and R. M. Rejesus. "A Comprehensive Review of the RMA APH and COMBO Rating Methodology: Final Report." 2010. Prepared by Sumaria Systems for the Risk Management Agency.

Efron, B. "Bootstrap Methods: Another Look at the Jackknife." *Annals of Statistics* 7(1979):1–26.

Embrechts, P., A. J. McNeil, and D. Straumann. "Correlation and Dependence in Risk Management: Properties and Pitfalls." In M. Dempster, ed., *Risk Management: Value at Risk and Beyond,* Cambridge, UK: Cambridge University Press, 2002, 176–223.

Feng, X., and D. Hayes. "Diversifying Systemic Risk in Agriculture." *Agricultural Finance Review* 76(2016):512–531. doi: 10.1108/AFR-06-2016-0061.

Finger, R. "Revisiting the Evaluations of Robust Regression Techniques for Crop Yield Data Detrending." *American Journal of Agricultural Economics* 92(2010):205–211. doi: https://doi.org/10.1093/ajae/aap021.

Gehlke, C. E., and K. Biehl. "Certain Effects of Grouping upon the Size of the Correlation Coefficient in Census Tract Material." *Journal of the American Statistical Association* 29(1934):169–170. doi: 10.1080/01621459.1934.10506247.

Genest, C., B. Remillard, and D. Beaudoin. "Goodness-of-Fit Tests for Copulas: A Review and a Power Study." *Insurance: Mathematics and Economics* 44(2009):199–213. doi: 10.1016/j.insmatheco.2007.10.005.

Glauber, J. W. "Crop Insurance Reconsidered." *American Journal of Agricultural Economics* 85(2004):1179–1195. doi: 10.1111/j.0002-9092.2004.00663.x.

Goodwin, B., and A. Hungerford. "Copula-Based Models of Systemic Risk in U.S. Agriculture: Implications for Crop Insurance and Reinsurance Contracts." *American Journal of Agricultural Economics* 97(2015):879–896. doi: 10.1093/ajae/aau086.

Goodwin, B. K., A. Harri, R. M. Rejesus, K. H. Coble, and T. O. Knight. "Actuarial Review for Price Volatility Factor Methodology." 2014. Prepared by Sumaria Systems for the Risk Management Agency.

Goodwin, B. K., and A. P. Ker. "Nonparametric Estimation of Crop Yield Dsitributions: Implications for Rating Group-Risk Crop Insurance Contracts." *American Journal of Agricultural Economics* 80(1998):139–153. doi: 10.2307/3180276.

Hoeffding, W. "A Non–Parametric Test of Independence." *Annals of Mathematical Statistics* 19(1948):546–557.

Huang, W., and A. Prokhorov. "A Goodness-of-Fit Test for Copulas." *Econometric Reviews* 33(2014):751–771. doi: 10.1080/07474938.2012.690692.

Huber, P. J. "Robust Regression: Asymptotics, Conjectures and Monte Carlo." *Annals of Statistics* 1(1973):799–821. doi: 10.1214/aos/1176342503.

Hungerford, A. E., and B. Goodwin. "Big Assumptions for Small Samples in Crop Insurance." *Agricultural Finance Review* 74(2014):477–491. doi: 10.1108/AFR-09-2014-0025.

Joe, H. *Dependence Modeling with Copulas.* Boca Raton, FL: Chapman and Hall/CRC Press, 2015.

Joe, H., and J. J. Xu. "The Estimation Method of Inference Functions for Margins for Multivariate Models." Technical Report 166, Department of Statistics, University of British Columbia, Vancouver, Canada, 1996.

Ker, A. P., T. N. Tolhurst, and Y. Liu. "Bayesian Estimation of Possibly Similar Yield Densities: Implications for Rating Crop Insurance Contracts." *American Journal of Agricultural Economics* 98(2016):360–382. doi: 10.1093/ajae/aav065.

Kojadinovic, I., J. Yan, and M. Holmes. "Fast Large-Sample Goodness-of-Fit Tests for Copulas." *Statistica Sinica* 21(2011):841–871.

Leamer, E. E. "Model Choice and Specification Analysis." In Z. Griliches and M. D. Intriligatorr, eds., *Handbook of Econometrics,* vol. 1. New York, NY: North-Holland Publishing, 1983, 285–330.

Li, B., and M. G. Genton. "Nonparametric Identification of Copula Structures." *Journal of the American Statistical Association* 108(2013):666–675. doi: 10.1080/01621459.2013.787083.

Liu, Y., and A. Ker. "Is There Too Much History in Historical Yield Data?" 2018. Paper presented at the triennial meeting of the International Agricultural Economics Association, Vancouver, British Colubmia, July 28–August 2.

Meinguet, J. "Multivariate Interpolation at Arbitrary Points Made Simple." *Journal of Applied Mathematics and Physics* 30(1979):292–304.

Nelsen, R. B. "Some Concepts of Bivariate Symmetry." *Journal of Nonparametric Statistics* 3(1993):95–101. doi: 10.1080/10485259308832574.

Nelson, C. H., and P. V. Preckel. "The Conditional Beta Distribution as a Stochastic Production Function." *American Journal of Agricultural Economics* 71(1989):370–378. doi: 10.2307/1241595.

Park, E., B. W. Brorsen, and A. Harri. "Using Bayesian Kriging for Spatial Smoothing in Crop Insurance Rating." *American Journal of Agricultural Economics* (2018):aay045. doi: 10.1093/ajae/aay045.

Racine, J., and A. Ker. "Rating Crop Insurance Policies with Efficient Nonparametric Estimators that Admit Mixed Data Types." *Journal of Agricultural and Resource Economics* 31(2006):27–39.

Scarsini, M. "On Measures of Concordance." *Stochastica* 8(1984):201–218.

Sherrick, B. J., C. A. Lanoue, J. D. Woodard, G. D. Schnitkey, and N. D. Paulson. "Crop Yield Distributions: Fit, Efficiency, and Performance." *Agricultural Finance Review* 74(2014):348–363. doi: 10.1108/AFR-05-2013-0021.

Simon, N., and R. Tibshirani. "Comment on 'Detecting Novel Associations In Large Data Sets' by Reshef et al., *Science*, Dec. 16, 2011." *ArXiv* (2014):1–3. Available online at http://adsabs.harvard.edu/abs/2014arXiv1401.7645S.

Sklar, A. "Fonctions de Repartition a *n* Dimensions et Leurs Marges." *Publications de l'Institut de Statistique de l'Universite de Paris* 8(1959):229–231.

Szekely, G. J., and M. L. Rizzo. "Brownian Distance Covariance." *Annals of Applied Statistics* 3(2009):1236–1265.

Szekely, G. J., M. L. Rizzo, and N. K. Bakirov. "Measuring and Testing Dependence by Correlation of Distances." *Annals of Statistics* 35(2007):2769–2794. doi: 10.1214/009053607000000505.

Tolhurst, T. N., and A. P. Ker. "On Technological Change in Crop Yields." *American Journal of Agricultural Economics* 97(2015):137–158. doi: 10.1093/ajae/aau082.

Vuong, Q. H. "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses." *Econometrica* 57(1989):307–333. doi: 10.2307/1912557.

Zhu, Y., B. K. Goodwin, and S. K. Ghosh. *Modeling Dependence in the Design of Crop Insurance Contracts*. United States: Scholars' Press, 2014.

Online Supplement:
How High the Hedge:
Relationships between Prices and Yields
in the Federal Crop Insurance Program

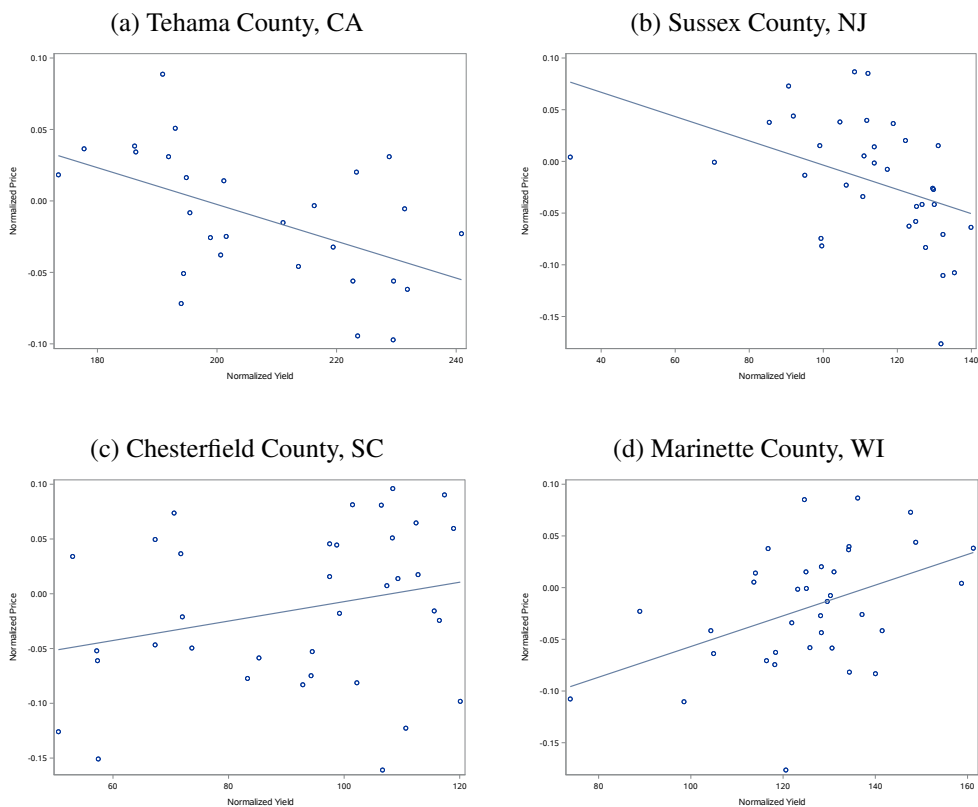**A. Ford Ramsey, Barry K. Goodwin, and Sujit K. Ghosh**

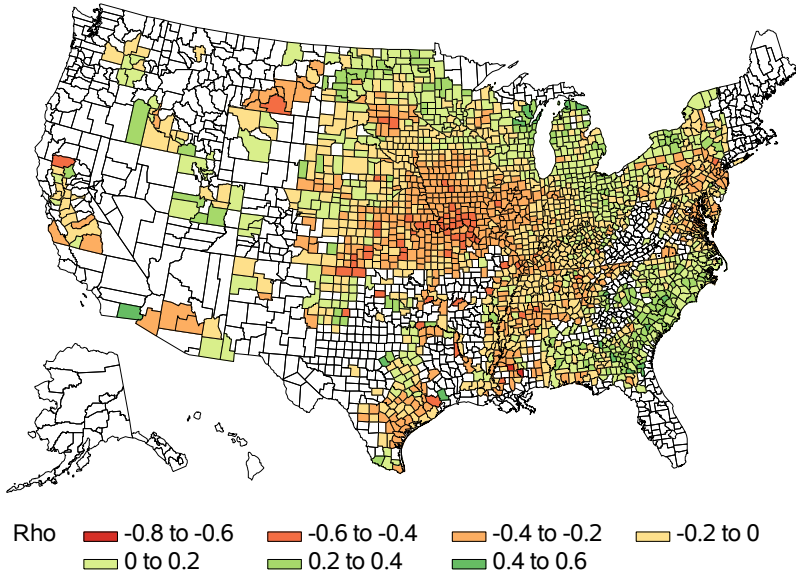**Figure S1. Scatterplots for Selected Counties**

Rho ☐ -0.8 to -0.6   ☐ -0.6 to -0.4   ☐ -0.4 to -0.2   ☐ -0.2 to 0
     ☐ 0 to 0.2   ☐ 0.2 to 0.4   ☐ 0.4 to 0.6

Figure S2. Spearman Correlation: Price vs. Yield

Tau ☐ -0.6 to -0.4   ☐ -0.4 to -0.2   ☐ -0.2 to 0
     ☐ 0 to 0.2   ☐ 0.2 to 0.4
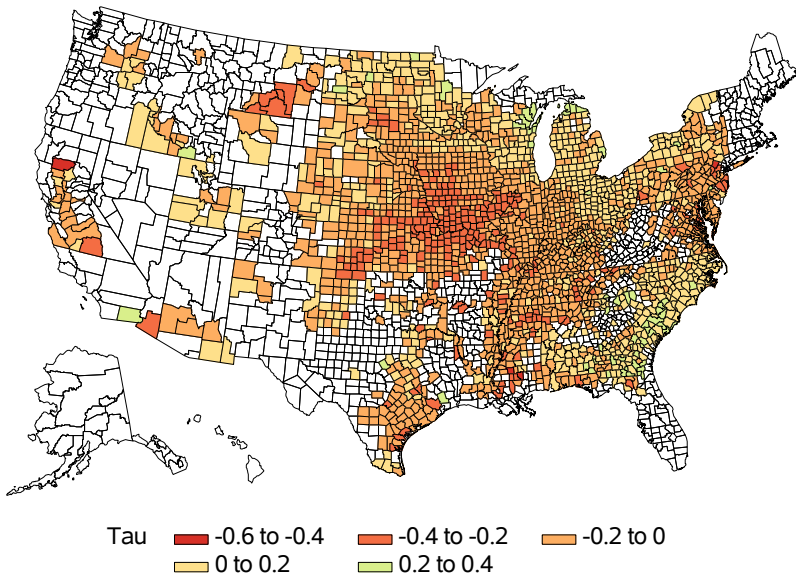
Figure S3. Kendall's Tau: Price vs. Yield
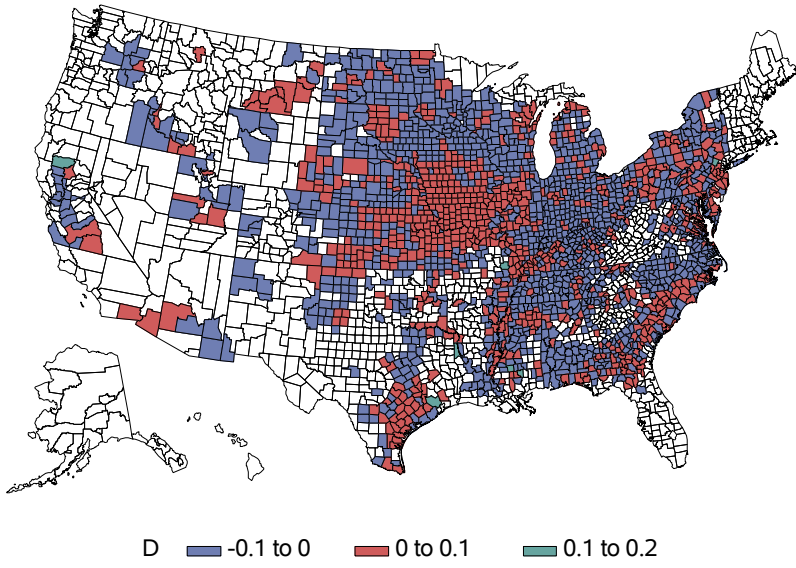
Figure S4. Hoeffding's D: Price vs. Yield
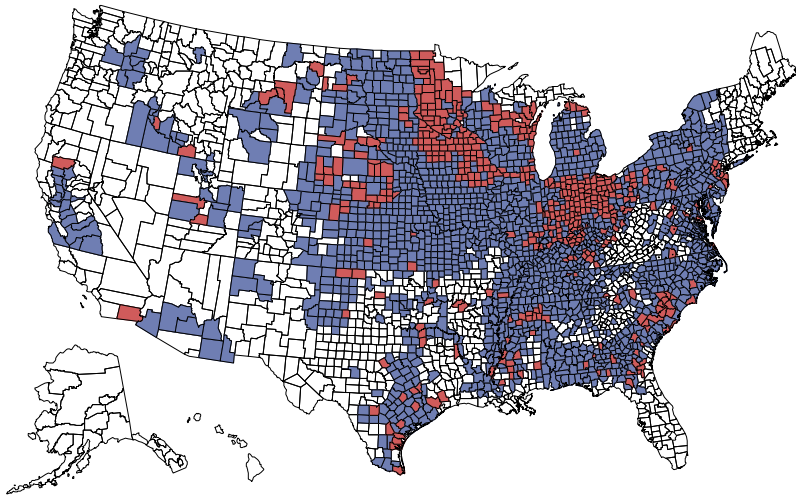
Figure S5. Bootstrap Results

(a) Normal Cop. Varying Corr. - Normal Cop. RMA Corr.

Difference in Rates

(0.90%) - (0.09%)	(0.09%) - (0.05%)
(0.05%) - (0.03%)	(0.03%) - (0.01%)
(0.01%) - 0.00%	0.00% - 0.03%
0.03% - 0.10%	0.10% - 0.90%

(b) Optimal Cop. - Normal Cop. RMA Corr.

Difference in Rates

(2.02%) - (0.45%)	(0.45%) - (0.25%)
(0.25%) - (0.15%)	(0.15%) - (0.09%)
(0.09%) - (0.05%)	(0.05%) - (0.01%)
(0.01%) - 0.05%	0.05% - 1.05%
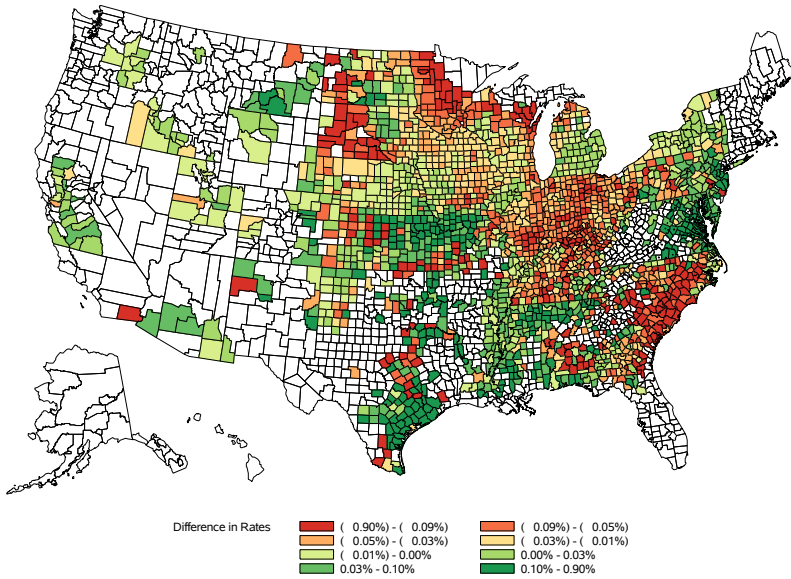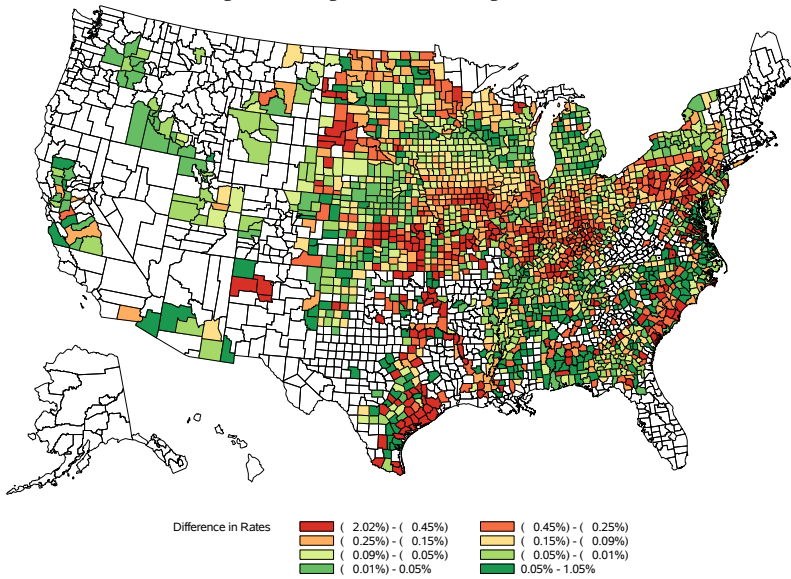
Figure S6. Difference in Rates

(a) Rates for Normal Copula

(b) Rates for *t* Copula

(c) Rates for Gumbel Copula

(d) Rates for Clayton Copula

(e) Rates for Frank Copula

(f) Rates for Gumbel 90 Copula
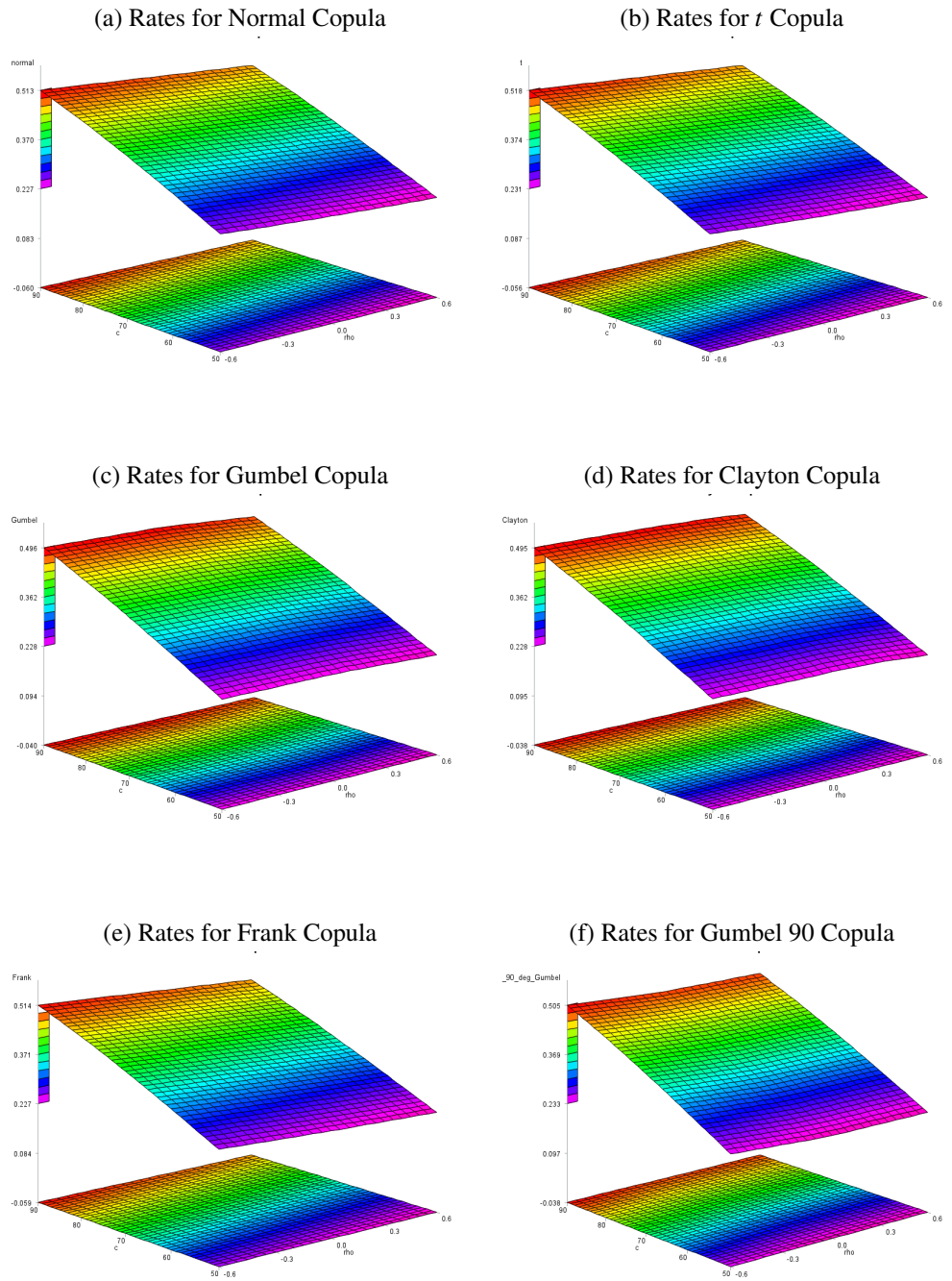
Figure S7. Simulated Rates for Selected Copula Models
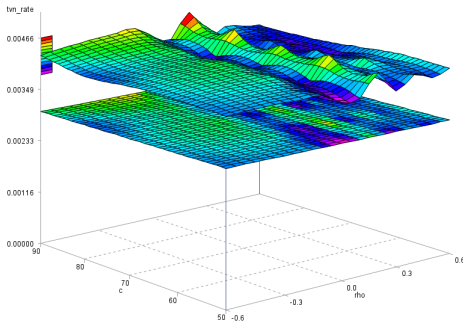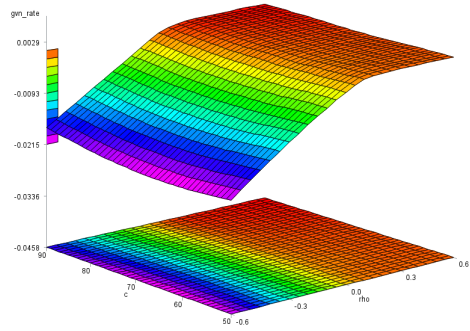
(a) *t* Rate − Normal Rate

(b) Gumbel Rate − Normal Rate

(c) Clayton Rate − Normal Rate
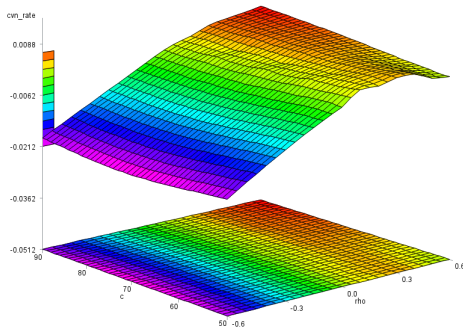
(d) Frank Rate − Normal Rate
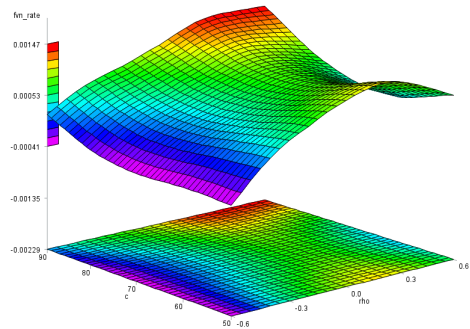
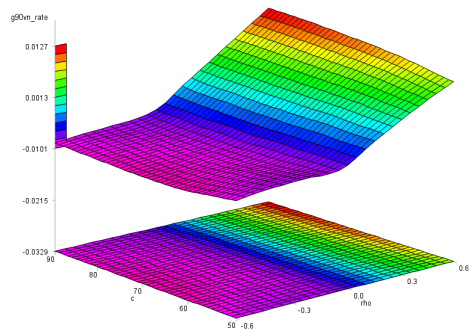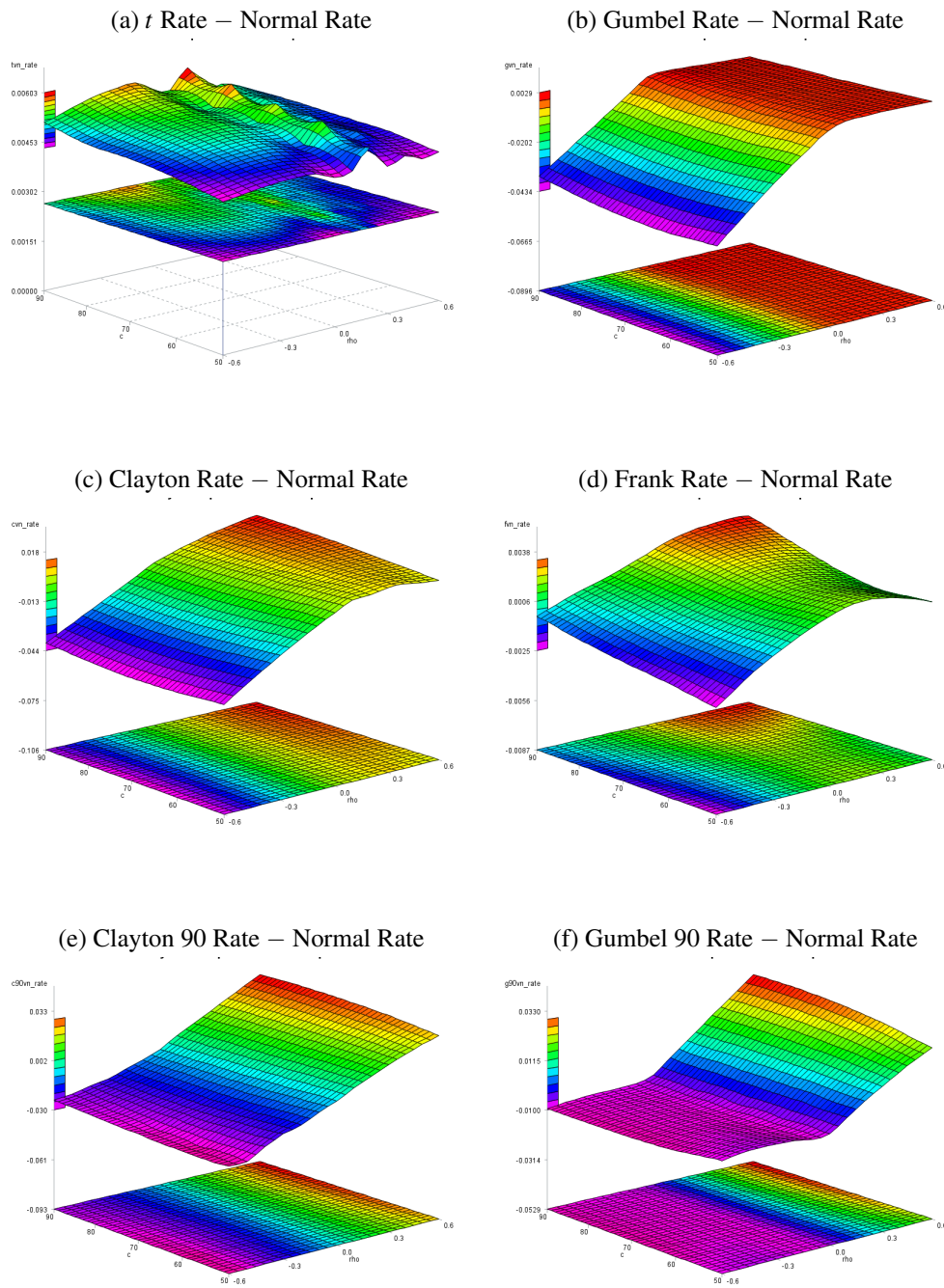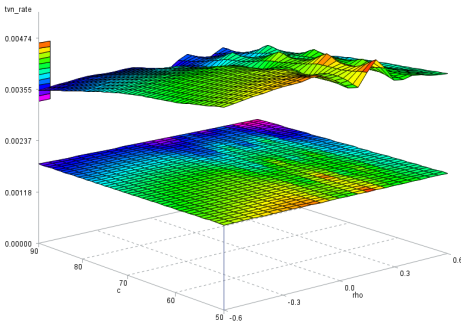(e) Clayton 90 Rate − Normal Rate
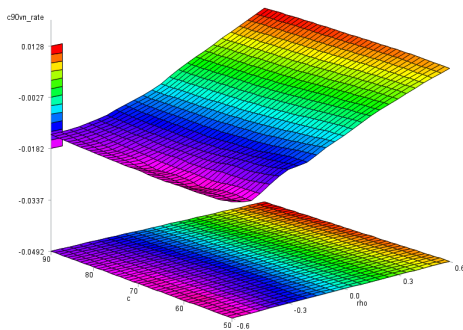
(f) Gumbel 90 Rate − Normal Rate

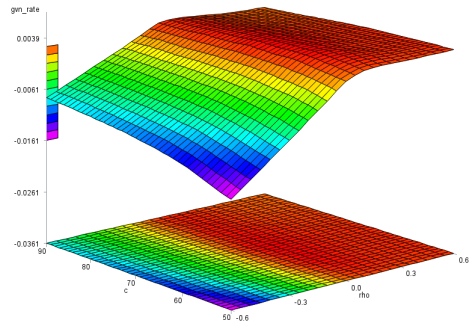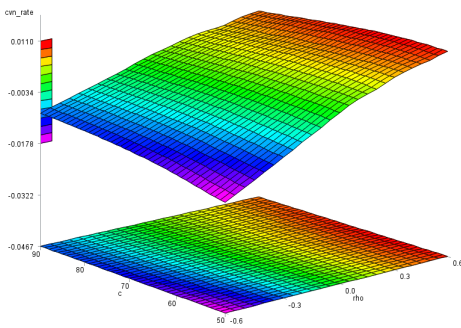**Figure S8. Simulated Differences in Rates for Selected Copula Models**
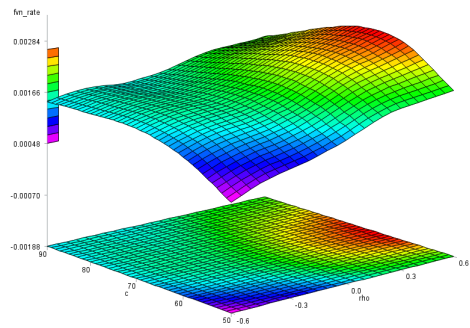
(a) *t* Rate − Normal Rate

(b) Gumbel Rate − Normal Rate

(c) Clayton Rate − Normal Rate

(d) Frank Rate − Normal Rate

(e) Clayton 90 Rate − Normal Rate

(f) Gumbel 90 Rate − Normal Rate

Figure S9. Simulated Differences in Rates for Selected Copula Models with High Price Volatility

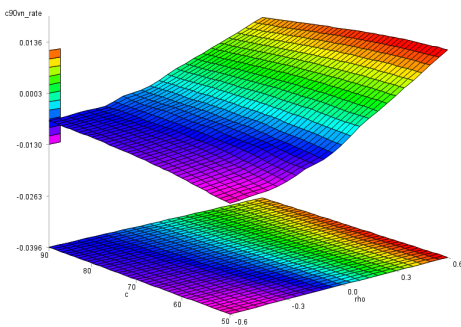(a) *t* Rate − Normal Rate

(b) Gumbel Rate − Normal Rate
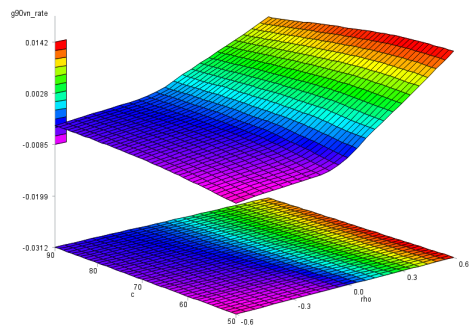
(c) Clayton Rate − Normal Rate

(d) Frank Rate − Normal Rate

(e) Clayton 90 Rate − Normal Rate

(f) Gumbel 90 Rate − Normal Rate

Figure S10. Simulated Differences in Rates for Selected Copula Models with Long Tailed Yields

Table S1. Frequencies of Copulas

	Number of Times Copula Is Selected	
Copula	**Frequency**	**Percentage**
Normal	273	13.13
t	135	6.49
Gumbel	179	8.61
Clayton	204	9.81
Frank	388	18.65
Gumbel Rot.	340	16.35
Clayton Rot.	561	26.97

Table S2. Detrending Robustness Check

	Difference in Premium under Different Detrending Methods	
Method	**Diff. Normal − RMA**	**Diff. Opt − RMA**
Linear Regression	−$9,081,656	−$62,982,398
Robust Regression	−$9,114,890	−$64,651,449
LOESS	−$7,863,085	−$64,532,124