Land Resilience and Tail Dependence Among Crop Yield Distributions

Abstract: We propose and empirically test a simple model to shed light on the nature of interactions between weather, land quality and yield. The conceptual model posits *i*) substitution relations between water stress metrics and soil quality as well as *ii*) a soil-conditioned threshold water stress level beyond which soil cannot buffer crop yields. The model implies that yield-yield dependence should vary as growing conditions vary. In comparison with intermediate growing conditions, yield-yield dependence should strengthen when growing conditions are either very good or very poor. County yield data strongly support substitution between soil and benign water availability levels but complementarity between soil and beneficial heat variables. Our estimated model provides qualified support for the hypothesis that better land is more resilient to water stress. We estimate a pseudo-copula that nests the Gaussian copula, finding strong evidence of left tail dependence among yields. Our formal model and empirical findings corroborate others' concerns about the appropriateness of current USDA rate-setting methodologies, which posit constant state-conditional rank correlations, implicitly assumed by use of the Gaussian copula. An application to aggregate crop yield rate setting suggests that current methods underprice area yield and whole farm premiums. Applying our empirical model to medium range weather projections under a climate change scenario for the Northern Great Plains, we infer that systemic yield correlations will increase in future years.

Keywords: climate change; crop insurance; Gaussian copula; pseudo-copula; reinsurance; systemic risk.

JEL Codes: G2, Q14, Q18, C58.

Suggested Running Head: Land Resilience and Crop Yield Tail Dependence

Xiaodong Du
Associate Professor, Dept. of Agricultural & Applied Economics, University of Wisconsin, Madison, Wisconsin; e-mail xdu23@wisc.edu
David A. Hennessy
Professor and Elton R. Smith Chair in Food & Agricultural Policy, Dept. of Agricultural, Food & Resource Economics, Michigan State University, East Lansing; e-mail hennes64@msu.edu
Hongli Feng
Associate Professor, Dept. of Agricultural, Food & Resource Economics, Michigan State University, East Lansing; e-mail hennes65@msu.edu
Gaurav Arora
Assistant Professor, Social Science Division, Indraprastha Institute of Information Technology, Delhi; e-mail gaurav@iiitd.ac.in

Acknowledgements: We thank two anonymous journal reviewers, editor Jim Vercammen and several conference participants for suggestions that refined and sharpened the paper's contributions. Chris Anderson helped guide us through some issues in understanding climate change projection data. Correspondence may be sent to: hennes64@msu.edu.

Arguably the three most primitive natural resources are a region's soils and climate, and the life forms available to use these resources. These assets are not to be taken for granted as they can be depleted and developed through soil erosion and humus build-up, climate change as well as habitat destruction. Our general interest is in a matter that has received sparse attention in the crop production technology literature; how do two of these endowments, soil and climate, interact in determining our capacity to cultivate the third, crops? That soil and climate interact as inputs is clear, for it is soil moisture and not rainfall that hydrates a crop. Our specific goal is to explore the interaction and its implications for yield covariation across space. In particular, we argue that soil-weather interactions should lead to strong yield correlations across space where soils are poor and in drought conditions. Beyond being fundamental attributes of yield determination, clarification on the nature of soil-weather interactions and on determination of yield covariation structures can inform on policy matters in the realms of crop insurance, climate change and commodity price volatility.

Diversification can be an important risk management strategy but its effectiveness relies on statistical dependence structures. When multiple risks are less than perfectly dependent then risk sharing can effectively reduce overall exposure. The nature of dependence structures matters greatly. Standard financial theory, as in Capital Asset Pricing models, has presumed state-invariant correlation structures. However financial market turbulence has shown state invariant correlation to be a problematic assumption. Among the approaches brought into question is the popular Gaussian copula methodology (Salmon 2012).

By considering county-level yield correlation as intercounty distance increases, Goodwin (2001) has noted that spatial dependence is state-dependent, being largest in drought years. His observation indicates a need to revisit how dependence is modelled among random variables entering crop insurance products (Goodwin 2015a, 2015b). Others too have suggested that the modelling of tail events may lead to problems with pricing crop insurance (Staudt 2010). Crop insurance rate setting typically involves the modeling of multiple risks. One approach to doing so, an algorithm due to Iman and Conover (1982), has become a central plank in United States crop insurance rate setting methodology (Coble et al. 2010) and is embedded in popular risk management software such as *@RISK* by Palisade Corporation. While quite general in the dependence structures it can model, as typically applied the algorithm essentially generates the Gaussian copula (Mildenhall 2005).

A copula maps marginal distributions into a multivariate distribution. Although the choice of marginals is discretionary, their interactions will be constrained by the copula choice. While the Gaussian copula does not fix cardinal measures of correlation, severe constraints are placed on rank correlations. It cannot capture what is commonly referred to as tail dependence, i.e., a strengthening of dependence when one of the variables of interest takes on a tail value; an example being U.S. housing prices circa 2008 (Zimmer 2012). In commodity markets, Zimmer (2015) provides evidence that U.S. corn, soybean and wheat price-price covariations strengthen when prices decline. Goodwin and Hungerford (G&H 2015) show that copulas which can accommodate at least some tail dependence, such as *t*- and mixture-copulas, fit unit-level Illinois corn yield data better than does the Gaussian copula. Ahmed and Serra (2015) find the Gaussian to be a comparatively good fit for orchard crop revenue in Spain, but their study does not emphasize tail dependence. Qiu and Rude (2016) discerned evidence of tail dependence between Ukraine domestic wheat and flour prices but not among international wheat prices.

There are several reasons why one should care about the existence of tail dependence. Crop insurance mispricing could arise from erroneously assessing systemic risk. When compared with financial markets, systemic risk is harder to estimate for crop yields because observations are annual. For unit-specific yield insurance contracts, unacknowledged tail dependence does not imply biased rates but does suggest underestimation of exposure by those who hold significant contract portfolios. For revenue, whole farm and area contracts, failure to account for tail dependence does suggest a bias in rates. Consider a whole farm product. If correlations among unit yields strengthen whenever crops become more stressed then diversification offsets will likely fail when most needed.

A second concern regards the costs of managing systemic risk, even if correctly calculated. The book of United States crop risks, with potential total liabilities of about $100 billion in recent years, is small when compared with risks assumed in global reinsurance markets (Goodwin and Smith 2013). However, the Approved Insurance Providers who work in public-private partnership with the U.S. Department of Agriculture's (USDA) Risk Management Agency (RMA) must retain or trade away certain risks while the federal government reinsures the rest.[1] For both providers and the RMA, higher systemic yield risk increases the transactions costs of managing their exposure. The possibility of tail dependence also raises separate concerns. Whether for political reasons or with intent to rectify a perceived market failure, governments often intervene in regions and sectors hit by adverse yield systemic shocks and these interventions may have lasting impacts. In addition, evidence suggests that climate may be changing. If weather extremes generate more systemic impacts on yield then, all else equal, a region's aggregate yields may become more variable as a result of shifted weather distributions.

The intents of this article are five-fold. Firstly we employ county-level data to ask how soil and weather variables interact in corn yield determination. An important contribution is to provide evidence that different weather attributes interact differently with soil attributes, where good soil complements beneficial temperature but substitutes for beneficial moisture conditions. We then take a conceptual approach to reason why state-dependent correlation structures are to be expected. Specifically, regarding yield-yield correlations we develop a model grounded in the view that yields on productive land should be comparatively more resilient to adverse weather shocks than yields on more limited soils. The model, which we call the land yield resilience model or LYRM, suggests that covariation should strengthen in better growing areas when conditions are good, and also in marginal growing areas when conditions are poor. Our third intent is to confront the LYRM hypothesis about covariation structures with empirical evidence. To do so we develop on copula comparisons in G&H (2015) by estimate a pseudo-copula that nests the Gaussian copula (Fang and Madsen 2013). We then compare insurance rates under the Gaussian pseudo-copula with that under the more restricted Gaussian copula. Finally we ask how covariation in the left tail might change in the light of climate projections that envision greater weather stress to cropping in some crop growing regions.

Together, the first two goals are intended to provide a conceptual foundation for mapping soil and weather inputs to tail attributes of joint yield outcomes. Although a connection between these inputs and how well yield responds to stress is intuitive, we have not found a formal literature that discusses the connection. We believe that formal inquiry into the idea is both new to economics and worthy of scrutiny, if only in light of implications for systemic risk. The third through fifth goals are intended to provide evidence that the model can further useful thought about important policy problems.

Our work is most closely related to G&H (2015) who question whether the Gaussian copula that the RMA implicitly assumes is appropriate. Noting Goodwin's (2001) earlier research on drought-dependent spatial yield correlations, G&H estimate alternative models that allow for strengthened dependence in the left tail. They find evidence in favor of flexible vine copulas over the Gaussian and other copula structures. A second paper broaching the theme is Tack and Holt (2016), who find spatial correlations among state-level corn yields to be stronger in abnormally bad and abnormally good years when compared with normal years. Building on their work, we turn attention to explaining *why* yield-yield dependence might strengthen in extreme weather conditions. Our resilience model is distinct from, but entirely consistent with recent work by Hennessy (2009a), Du, Hennessy and Yu (2012) and Du et al. (2015). They stress land fertility and also weather endowments as yield skewness determinants, but do not address yield-yield covariation. Du, Hennessy and Feng (2014) show that a region's natural resource endowments together with systemic yield and price-yield correlations should affect preferences for crop insurance forms (revenue or yield) and coverage levels, but do not ask how resource endowments can affect a region's systemic risk attributes.

The paper's first main section serves to provide conceptual background. It proposes a simple relation between underlying soil and weather inputs and production outcomes for heterogeneous land units. It also presents relevant dependence concepts and makes the case that interactions between growing condition inputs should affect yield covariation structures. The three sections that follow discuss weather and soils data, present an empirical model of their interaction, and estimate a technology that can test for the proposed relation. We then make two applications. The first argues that increasing dependence along the left tail will increase indemnity payouts for area yield insurance products, and that assuming constant rank correlation across yields will lead to insurance rates that are too low to cover indemnities. Empirical estimates of expected indemnities when tail dependence is ignored and when it is allowed are used to show that large biases in rates may result. The second application illustrates that our model of yield production relations, together with current climate change projections for the U.S. Western Cornbelt, imply that higher systemic yield correlations are to be expected in the future. The work concludes with a discussion on findings and with suggestions for future inquiry.

Conceptual Model

This section will first develop a model that relates soil attributes and conditions to systemic risk. Upon providing statistical characterizations of tail dependence, we will then address what the model implies for tail dependence.

Land resilience and yield-yield correlations

Our conceptual model posits an aggregate index of weather suitability for crop growth as variable *w* with distribution $F(w)$ on support $[w_l, w_h]$. A land unit's ability to convert realized weather into yield is assumed to depend on soil quality. When $w = w_h$ then we assume that all land units yield quantity y_h. We parameterize soil quality as θ, $\theta < w_h$, and yield response to weather conditions on $[\theta, w_h]$ is given as β_l.[2] Here a smaller value of θ indicates better land quality as the soils prove to be more resilient to adverse weather outcomes. A literal interpretation of θ is the soil's capacity to substitute for good weather through buffering. However buffering is limited and yield becomes more sensitive to weather on interval $[w_l, \theta)$ where the response is given as β_h, $\beta_h > \beta_l$. Thus we posit the production relation:

(1) $$y(w;\theta) = \bar{y} - \beta_l(w_h - w) - (\beta_h - \beta_l)\max[\theta - w, 0],$$

where θ represents the point of non-differentiability. Figure 1 depicts, where it is assumed that a certain threshold level of weather or land quality is required for positive output.

The model can be seen as a two input (weather, soil) variant of the von Liebig and Sprengel Law of the Minimum production technology (Berck and Helfand 1990), a distinction being that response to more favorable weather does not decline to zero after exceeding the point of non-differentiability. For convenience we assume that land quality and weather outcomes are never such that yield is zero. In the figure $\theta^a < \theta^b$ so that the better land, with $\theta = \theta^a$, has low sensitivity to adverse weather over the weather outcome set $[\theta^a, w_h]$, a larger set than $[\theta^b, w_h]$. The salient feature is that yield on better quality land is more resilient to bad weather.

As a matter of model validation note that mean yield weakly decreases in θ, i.e., $d\int_{w_l}^{w_h} y(w;\theta)dF(w)/d\theta \le 0$ and expected yield increases in land quality. The coefficient of yield variation is readily shown to be larger under θ^b than under θ^a so that the model is consistent with higher actuarial risk per bushel insured on poor quality land than on better quality land (Du et al. 2015). The impact that higher θ, i.e., lower land quality, has on skewness is also apparent. Figure 2 shows how the weather favorability distribution maps into the yield distribution, see also Hennessy (2009b). The weather index cumulative distribution function is on the horizontal axis while that of the yield distribution has domain on the vertical axis, i.e., rotated counterclockwise by 90 degrees. The distributions have the same shape in the lower tail, up to location and scaling constants. In the upper tail, however, yield is less sensitive to weather and so the yield distribution is comparatively more compact in its upper quantiles. Comparative right-tail compaction is one way to order distributions by skewness (Hao and Naiman 2007).

Now, fixing on some weather distribution, we consider how an increase in land quality affects yield distribution moments. There are three effects. One is a rightward shift in the overall distribution, i.e., first-order stochastic dominance, as reflected by $y(w;\theta^a) \ge y(w;\theta^b)$ in figure 1. A second is lower variability as β_l, rather than β_h, maps weather into yield over a larger *w* set. There will also be right tail compaction under θ^a when compared with θ^b. In sum, mean increases, variability decreases and skewness becomes less positive or more negative.

Tail dependence

Tail dependence describes the monotone relationship between two random variables. Left-tail decreasing (LTD) and right-tail increasing (RTI) stochastic orders are widely used to inform on any tail dependence (Joe 2015). Throughout $P(A)$ denotes the probability of an outcome in set A, $P(A_1 \cap A_2)$ indicates the probability of an outcome in both A_1 and A_2, and $P(A_1 \mid A_2) = P(A_1 \cap A_2)/P(A_2)$ denotes conditional probability. For two random variables, x_1 and x_2, x_2 is

(2) LTD in x_1 whenever $P[x_2 \le \hat{x}_2 \mid x_1 \le \hat{x}_1^a] \ge P[x_2 \le \hat{x}_2 \mid x_1 \le \hat{x}_1^b] \quad \forall \hat{x}_1^b \ge \hat{x}_1^a, \quad \forall \hat{x}_2$;
RTI in x_1 whenever $P[x_2 \ge \hat{x}_2 \mid x_1 \ge \hat{x}_1^a] \le P[x_2 \ge \hat{x}_2 \mid x_1 \ge \hat{x}_1^b] \quad \forall \hat{x}_1^b \ge \hat{x}_1^a, \quad \forall \hat{x}_2$.

The LTD and RTI relations are not equivalent as neither cover realizations in the sets $\{(x_1, x_2): x_1 > \hat{x}_1, x_2 \le \hat{x}_2\}$ and $\{(x_1, x_2): x_1 \le \hat{x}_1, x_2 > \hat{x}_2\}$.[3] Were data to conform with either property then the Gaussian copula would be problematic. When the $\hat{x}_2$ is sufficiently large or small then the Gaussian copula will always exhibit zero conditional correlation for extreme values of $\hat{x}_1$ (Aas 2004) and so cannot encompass tail dependence in the senses of (2).

Two other notions of tail dependence are (p. 26 in Trivedi and Zimmer 2005):[4]

$$(3) \qquad \lambda_L \equiv \lim_{v \downarrow 0} \frac{P[\{x_1 \le q_v\} \cap \{x_2 \le q_v\}]}{v}; \qquad \lambda_U \equiv \lim_{v \uparrow 1} \frac{P[\{x_1 \ge q_v\} \cap \{x_2 \ge q_v\}]}{1-v};$$

where q_v = quantile v and $v \in (0,1)$. Here notation $\{\cdot\}$ indicates a condition in the event that notation might otherwise be ambiguous. Both λ_L and λ_U have value 0 for any Gaussian copula.

Given that the LYRM model has only one source of randomness, weather favorability index w, one might ask whether yields are essentially perfectly correlated as a result. They are not although they are comonotonic, i.e., they are monotone increasing functions of the same random variable, and so are perfectly ordinally (i.e., rank) correlated. Returning to relation (3) above, one consequence is that $\lambda_L \equiv \lambda_U \equiv 1$ under the LYRM and so the Gaussian copula *cannot possibly* represent the joint yield distribution. Another consequence is that the standard cardinal, or Pearson, correlation is not unity when land units differ in quality. In figure 1 when $\theta^a < w < \theta^b$ then yield on lower quality land is comparatively more sensitive to weather. As a result, $\text{corr}[y(w;\theta^a), y(w;\theta^b) \mid w \le \theta^b] \le \text{corr}[y(w;\theta^a), y(w;\theta^b) \mid w \le \theta^a] \equiv 1$ and $\text{corr}[y(w;\theta^a), y(w;\theta^b) \mid \theta^a < w] \le \text{corr}[y(w;\theta^a), y(w;\theta^b) \mid \theta^b < w] \equiv 1$. More generally, the model supports the idea that correlations are weakest between the two tails, i.e., a U-shaped relation. This model inference can shed light on findings in Tack and Holt (2016), in which the correlation between average yields in Iowa and average yields in Illinois are compared as the extent of weather extremes vary. Consistent with their findings, LYRM proposes that correlation should be least positive when weather is moderate.

As further conceptual support for LYRM, in Item 1 of the supplementary appendix online we demonstrate the following connection between the model and tail dependence statistics:

Proposition 1: In the LYRM, all pairs of land units chosen have joint yield distributions that satisfy both the LTD property and the RTI property.

We contemplate next how figure 1 can be adapted to address differences in tail dependence across soils and weather distributions. The impact could be modeled in two ways. A weather favorability distribution could shift leftward to indicate generally less favorable weather or a yield function could shift rightward to indicate poorer soils. Figure 3 illustrates by shifting the weather distribution where the left-more distribution illustrates South Dakota and the right-more distribution illustrates Iowa. In fact, Iowa generally has better crop growing weather and soils so we can, for ease of presentation and without compromise to representation, assume that the distinction between the states is represented by a shift in the weather favorability distribution.

Finally we ask whether model inferences are robust to relaxing the comonotonicity feature. Suppose that the weather favorability index can differ across land plots through adding a zero mean, random independent land plot idiosyncratic weather component $\tilde{w}^i$, $i \in \{a, b\}$, to systemic component w so that the weather realizations are $w + \tilde{w}^a$ and $w + \tilde{w}^b$. So long as the idiosyncratic components have narrow supports then they will not affect our inferences above. This is because yield in both tails has a linear response to weather favorability. Once weather favorability is sufficiently large or small then idiosyncratic components with narrow supports will not allow for a $w + \tilde{w}^i$ realization that is the other side of the kink and so all of the findings above will continue to hold. The rest of this paper will seek to shed light on how data compare with the conceptual model, and also on model implications.

Data for Weather and Soils

While weather affects yields in many ways, for field crops the two most important weather inputs are typically moisture and temperature. As we will develop upon later, how these distinct weather inputs interact with soils may differ qualitatively. But first we will present our metrics for moisture, temperature and soils.

We use the Palmer's Z, Z_t, to measure drought and excess moisture (Xu et al. 2013). This variable quantifies how moisture conditions in a climate division deviate from normal conditions in the area. There are about nine climate divisions per state, where boundaries generally follow county and crop reporting district lines. The index and its variants seek to measure the water stock available so that precipitation over several prior months matter (Heddinghaus and Sabol 1991). As it accounts for evapotranspiration, the index embeds the season's accumulated past temperature, i.e., past temperature increases past evapotranspiration. The index 'stock' is presumed to decay at about 10% per month while current precipitation deviations from reference levels provide any net water inflow. The reference level incorporates *i*) evapotranspiration loss and *ii*) soil attribute implications for runoff and recharge as given by a model with two soil layers where the lower layer is representative of the climate division's soil traits. As the hydrological implications of weather are conditioned on soil depth and other attributes, Palmer's index has already sought to model a substitution relation between soil quality and water supply. The effort is, however, crude.

We project climate division July Palmer's Z data onto each county within the climate division. The range [-2, 2.5] is viewed as reflecting normal or moderate stress, below -2 indicates severe drought and above 2.5 indicates excess moisture (Karl 1986; NOAA 2014). Our measure of drought stress is $P_i^L = \min[Z_{i,t} - 2, 0]$ where L indicates 'Left'. Note that P_i^L is non-positive and becomes less negative whenever weather conditions improve. Our measure of excess moisture is $P_i^R = \max[Z_{i,t} - 2.5, 0]$. This measure is non-negative and becomes more positive whenever weather conditions deteriorate.

County *i* growing degree days in year *t*, $G_{i,t}$, are measured in Fahrenheit (^{o} F). Daily average temperature surplus to 50° F, but capped at 86° F, is summed over the May-September growing season, which is taken as the time interval. County *i* stress degree days in year *t*, $S_{i,t}$, are taken as daily average temperature surplus to 90° F. In the later empirical analyses, all weather related variables, including Palmer's Z, growing and stress degree days, are normalized by subtracting the mean and divided by the difference between max and min values. The same normalization procedure is also applied for the soil suitability measure described below.

Soil also matters in determining water available to a crop. Moisture must infiltrate and permeate the soil, which must be capable of holding the water (Bot and Benites 2005). Water retention capacity depends on soil depth, the (clay, silt, sand) composition, organic matter content, physical/chemical pan status and extent of soil biological activity. We use the Land Capability Classification (Helms 1992) to measure soil suitability. It is based on soil measurements that are available throughout North America and has been applied to all land of agricultural relevance in the United States. Although the classification was devised to support approaches that discourage soil erosion (Helms 1992), it is widely used to measure yield productivity.

Logic suggests that less constrained soils will be more robust to weather variability as they drain well while retaining water in deeper layers in the event of dry or hot conditions. According to Helms (1992), 'Soils in Class I have few limitations that restrict their use,' while 'Soils in Class II have some limitations that reduce the choice of plants or require moderate conservation practices,' but 'The limitations are few and the practices are easy to apply.' Limitations may include (p. 7 in Helms 1992):

> "1) gentle slopes, 2) moderate susceptibility to wind or water erosion or moderate effects of past erosion, 3) less than ideal soil depth, 4) somewhat unfavorable soil structure and workability, 5) slight to moderate salinity or sodium easily corrected but likely to recur, 6) occasional damaging overflow, 7) wetness correctable by drainage but existing permanently as a moderate limitation, and 8) slight climatic limitations on soil use and management."

Soils in classes I and II are typically cropped. Those in classes III-IV can be cropped but pose perils for crop production and typically require special management practices. For Class

III, among limitations are (p. 7 in Helms 1992):

> " … 3) frequent overflow accompanied by some crop damage; 4) very slow permeability of the subsoil; 5) wetness or some continuing waterlogging after drainage; 6) shallow depths to bedrock, hardpan, fragipan, or claypan that limit the rooting zone and the water storage; 7) low moisture-holding capacity … ."

Limitations are even more severe on Class IV soils. Soils in higher classes are generally impractical for cropping. We label Q_i as the fraction of all land (cropland or otherwise) in county *i* that is within classes I or II, and we use this as a county's overall land quality index.

Soil-Weather Interactions

As preliminary analysis we ran the following regression for USDA National Agricultural Statistics Service (NASS) county average corn yield data, $y_{i,t}$, over the period 1950-2014:

$$(4)\qquad y_{i,t} = \alpha_0 + \alpha_1 t + \alpha_2 P^L_{i,t} + \alpha_3 P^R_{i,t} + \alpha_4 G_{i,t} + \alpha_5 S_{i,t} + (\alpha_6 P^L_{i,t} + \alpha_7 P^R_{i,t} + \alpha_8 G_{i,t} + \alpha_9 S_{i,t})Q_i + \text{residual}_{i,t}.$$

This model was inspired by Arora et al. (2016) who considered corn, soybean, wheat and alfalfa in the Dakotas. Our dataset includes 777 counties in the states of IL, IN, IA, KS, MI, MN, MO, NE, ND, OH, OK, SD, and WI. The panel data set is unbalanced and county-level fixed effect models are estimated where table 1 provides regression results. Regressions (4.1) and (4.2) differ only through inclusion of a quadratic time term in addition to a linear term as proxy to technological and other trends. In each regression direct weather results are as predicted and consistent with the existing literature, e.g., Schlenker and Roberts (2009) and Xu et al. (2013).

Interaction terms are more involved. Both Palmer's Z derived variables substitute for county soil quality, i.e., for Left (L) an higher, more beneficial, value is less beneficial under better land quality while for Right (R) an higher, less beneficial, value is ameliorated by better land quality. For Growing and Stress degree days the relationships are complementary; growing degree days are more beneficial and stress degree days more detrimental in counties with better land. Note that heat is partially accommodated in the Palmer variables, in order to address drought's evapotranspiration component. In addition, complementarity between heat and soil quality is reasonable in light of the Land Capability Class definitions provided above. Any given stock of growing degree days is likely to have greater impact on less constrained soils, which are easier to work, typically more level, require fewer tillage steps and drain more rapidly. Thus earlier planting, quicker emergence and fewer weather-related setbacks are to be expected. Substitution relations are more likely to arise in regard to water stress where good soils can buffer to absorb surplus water that might impede plant development and store it until any need arises. Thus, the results in regressions (4.1)-(4.2) should not surprise.

Regression (4.3) removes the degree day variables from consideration in the interaction so as to demonstrate that substitution between land quality and water stress conditions is not an artifact of including degree day variables. To highlight relative importance, regression (4.4) removes soil interactions with Palmer's Z from consideration. We conclude that the degree day variables complement soil variables and it is not appropriate to model them with the resilience specification.

We will seek to model whether delineation of soil quality to the county level captures a substitution relation beyond that built into the Palmer index. In light of table 1 and our conceptual framework, we model a substitution relationship between soil quality and our water availability metrics as follows:

$$
\begin{aligned}
&y_{i,t} = g(t, P^L_{i,t}, P^R_{i,t}, G_{i,t}, S_{i,t}, Q_i) + \text{residual}_{i,t}; \\
(5) \quad &g(\cdot) = \gamma_0 + \gamma_1 t + \gamma_2 P^L_{i,t} + \gamma_3 P^R_{i,t} + \gamma_4 G_{i,t} + \gamma_5 S_{i,t} + \gamma_6 G_{i,t} Q_i + \gamma_7 S_{i,t} Q_i + \gamma_8 \max[\lambda_1 Q_i - P^L_{i,t}, 0] \\
&\qquad + \gamma_9 \max[\lambda_2 Q_i + P^R_{i,t}, 0].
\end{aligned}
$$

The specification captures a substitution relation between water and soils. But, as in LYRM model (1), it does more in that the max[.] statements activate the substitution relations only when water stress occurs, where stress might be due to deficient soil water, as captured by $\gamma_8 \max[\lambda_1 Q_i - P^L_{i,t}, 0]$, or excess soil water, as captured by the other max[.] statement.[5]

Regarding specific hypotheses, a consideration of (5) allows us to posit the following:

H1) yield is increasing and concave in the Left Palmer Index, i.e., $\gamma_2 > 0 > \gamma_8$;

H2) yield is decreasing and concave in the Right Palmer Index, i.e., $\gamma_3 < 0, \gamma_9 < 0$;

H3) better soil substitutes for better, less negative, Left Palmer Index readings, i.e., $\lambda_1 < 0$;

H4) better soil substitutes for better, less positive, Right Palmer Index readings, i.e., $\lambda_2 < 0$.

Bayesian Estimation

Our basic LYRM has the following specification:

$$
\begin{aligned}
&y_{i,t} = J_{i,t}\varphi + X_{i,t}\phi + \psi \eta_{i,t} + \varepsilon_{i,t}; \\
(5') \quad &\eta_{i,t} \overset{iid}{\sim} N_{(0,\infty)}(0,1); \quad \varepsilon_{i,t} \overset{iid}{\sim} N(0, \sigma^2_\varepsilon); \qquad i \in \{1, \ldots, N\}; \qquad t \in \{1, \ldots, T\}; \\
&J_{i,t} = (1, t, P^L_{i,t}, P^R_{i,t}, G_{i,t}, S_{i,t}, G_{i,t}Q_i, S_{i,t}Q_i); \qquad \varphi = (\gamma_0, \ldots, \gamma_7)'; \\
&X_{i,t} = (\max[\lambda_1 Q_i - P^L_{i,t}, 0], \max[\lambda_2 Q_i + P^R_{i,t}, 0]); \qquad \phi = (\gamma_8, \gamma_9)'.
\end{aligned}
$$

To capture any yield skewness we incorporate the right-skew variable $\eta_{i,t}$ (Koop, Poirier, and Tobias 2007), so that yield is left (right) skewed when $\psi < (>)\, 0$. Variable $\eta_{i,t}$ is unobservable, independently and identically distributed (iid) and held to follow a half-normal distribution, $N_{(0,\infty)}(0,1)$. Residual $\varepsilon_{i,t}$ is normally distributed and is assumed to be iid.

The data generate a county-level unbalanced panel.[6] Estimation takes into account county-level fixed effects, which precludes inclusion of time-invariant county-level soil quality variables, by demeaning the variables using the within-county transformation: $\tilde{y}_{i,t} = y_{i,t} - (1/T_i)\sum_{t=1}^{T_i} y_{i,t}$ and $\tilde{x}_{i,t} = x_{i,t} - (1/T_i)\sum_{t=1}^{T_i} x_{i,t}$. Here T_i represents the number of sample years for county i while $x_{i,t}$ represents the right hand side individual explanatory variables in eqn. (5'). The demeaned vectors of independent variables are written as $\tilde{J}_i$ and $\tilde{X}_i$. Then stack the equation over t within county i, letting $\tilde{H}_i = (\tilde{J}_i, \tilde{X}_i, \eta_i)$ and $\Phi = (\varphi, \phi, \psi)'$, to obtain:

(6) $\tilde{y}_i = \tilde{H}_i \Phi + \tilde{\varepsilon}_i.$

The specification is nonlinear and estimated in a Bayesian framework. The Bayesian estimation procedure, including the Gibbs sampler and related estimation details on the model specified in (6), are documented in Item 2 of the supplementary appendix online. Parallel to table 1, table 2 provides estimates of parameter coefficients together with probabilities that coefficients are positive. Version 1, i.e., (6.1) in the table, includes the degree day variables. The variables are in the main highly significant and the trend variable has value approximately equal to that found in table 1. So as to clarify whether heat interactions with soil quality impact our estimates of interest, model (6.2) removes the degree day interactions with soil quality, retaining their direct effects.

Figure 4, panels a) and b), illustrate the Palmer index relationships with soil quality given the column 1 estimates in table 2 where the Palmer variables have been normalized. Panel a) depicts the response to the left Palmer index, i.e., for drought where the variable density function over data used is also provided. The two graphs provided are yields for Iowa and South Dakota as the left Palmer index varies, where yields are model (6) fitted values using the estimated coefficients and average values of $G_{i,t}$, $S_{i,t}$ and the Palmer Right Index of selected counties in Iowa and South Dakota.[7] Inspection reveals that both H1) and H3) apply, i.e., for Left Palmer the relationship is increasing and concave while the kink point shifts left as mean land quality increases. Panel b) shows the response to the Right Palmer index, i.e., for excess moisture. The relationship is first increasing and then decreasing so that concavity applies but the monotonicity hypothesized in H2) does not apply. However, γ_3 is not significantly different from 0. The kink point in panel b) shifts to the right as land quality increases, i.e., H4) applies.

In the above we inquired into soil-weather interactions using county-level data. We have found evidence that better soils ameliorate poor soil water conditions. Thus in locations with poor soils then crops have little capacity to withstand water stress and systemic yield shortfalls may occur. In the next section we inquire into the presence of tail-dependence among yields, and also how such dependence can affect premium formation.

Implications of Tail Dependence for Crop Insurance

In this section, we do three things. In the first subsection for whole-farm or area-wide products we develop a brief conceptual framework to help think through how strengthening dependence between yields can affect actuarially fair premiums, which we interpret to be the expected indemnity payout. In the second subsection we test for how well the Gaussian copula fits data when compared with a form that nests the Gaussian but allows for non-constant correlation. We also use the fitted copula to illustrate the economic significance of incorporating tail dependence into crop insurance rate setting.

Microeconomics of Insurance and Tail-dependent Yields

For two distinct land areas we pose stochastic yield interactions in an hierarchical model having two states of nature. State *U* is benign for growing crops and occurs with probability 0.5 while harmful state *D* also occurs with probability 0.5. Conditional on the state, yields follow a two-point bivariate distribution in the manner of Dasgupta and Maskin (1987). The joint yield distribution for y_1 and y_2 is given as

$$\text{State } U \text{ with prob. 0.5: } (y_1, y_2) = \begin{cases} (\mu+\delta, \mu+\delta) & \text{with prob. } 0.25(1+\rho_U); \\ (\mu, \mu+\delta) & \text{with prob. } 0.25(1-\rho_U); \\ (\mu+\delta, \mu) & \text{with prob. } 0.25(1-\rho_U); \\ (\mu, \mu) & \text{with prob. } 0.25(1+\rho_U); \end{cases}$$

(7)

$$\text{State } D \text{ with prob. 0.5: } (y_1, y_2) = \begin{cases} (\mu, \mu) & \text{with prob. } 0.25(1+\rho_D); \\ (\mu, \mu-\delta) & \text{with prob. } 0.25(1-\rho_D); \\ (\mu-\delta, \mu) & \text{with prob. } 0.25(1-\rho_D); \\ (\mu-\delta, \mu-\delta) & \text{with prob. } 0.25(1+\rho_D); \end{cases}$$

for ρ_U and ρ_D bounded on $[0,1]$. Notice that the conditional standard Pearson correlation is ρ_U for state *U* and ρ_D for state *D*. The unconditional marginals are $(\mu-\delta, \mu, \mu+\delta)$ with respective probabilities $(0.25, 0.5, 2.5)$, mean yield μ and variance $\delta^2/2$. When each area is one acre and output price is fixed at one then coverage at level $\varphi \in [0,1)$ gives yield guarantee as $\varphi\mu$ and area indemnity per acre as $\max[\varphi\mu - 0.5y_1 - 0.5y_2, 0]$. A final comment is that the state *D* conditional variance of $0.5y_1 + 0.5y_2$ is $(1+\rho_D)\delta^2/8$ so that the variance of average yield in state *D* is larger than that in state *U* whenever $\rho_D > \rho_U$.

We seek to compare two quantities. One is when yields are characterized as the compound lottery described in (7) where correlation is allowed to be state conditioned. The other is the same but where the correlation is restricted to be common, set at the unconditional level. Observe that the unconditional covariance is given as $\delta^2/4 + \delta^2(\rho_U + \rho_D)/8$ so that the unconditional correlation between y_1 and y_2 is $\rho = \sqrt{0.5 + 0.25(\rho_U + \rho_D)}$. This statistic is invariant to changes in ρ_U and ρ_D so long as $\rho_U + \rho_D$ remains fixed. So we calculate the expected payout under *i*) specification (7), and also *ii*) specification (7) except that $(\rho_U, \rho_D) \to (\bar{\rho}, \bar{\rho})$ where $\bar{\rho} \equiv (\rho_U + \rho_D)/2$. We need only consider when mean yield is below μ, i.e., three of the states arising under *D*. Aggregated yield is $\mu - 0.5\delta$ with probability $0.25(1-\rho_D)$ and $\mu - \delta$ with probability $0.125(1+\rho_D)$.

When the yield guarantee satisfies $\mu\varphi \in [\mu - \delta, \mu - 0.5\delta)$ then the expected payoff is $p(\varphi)$ $= (1+\rho_D)(\varphi\mu - \mu + \delta)/8$. When correlation is assumed to be average correlation then the inferred expected payoff is $\hat{p}(\varphi) = (1+\bar{\rho})(\varphi\mu - \mu + \delta)/8$ and the ratio of these premiums is $p(\varphi)/\hat{p}(\varphi) = (1+\rho_D)/(1+\bar{\rho})$, greater than 1 whenever $\rho_D > \rho_U$. The difference becomes less negative as the coverage level increases. When the yield guarantee satisfies $\mu\varphi \in [\mu - 0.5\delta, \mu]$ then the expected payoff is $p(\varphi) = [2\delta + (1-\varphi)\mu(\rho_D - 3)]/8$ while the inferred expected payoff becomes $\hat{p}(\varphi) = p(\varphi) + (1-\varphi)\mu(\bar{\rho} - \rho_D)/8$. Summarizing, the premium ratio is

$$(8) \qquad \frac{p(\varphi)}{\hat{p}(\varphi)} = \begin{cases} \dfrac{1+\rho_D}{1+\bar{\rho}} & \text{for } 1 - \dfrac{\delta}{\mu} \le \varphi < 1 - \dfrac{\delta}{2\mu}; \\[2ex] \dfrac{2\delta + (\rho_D - 3)(1-\varphi)\mu}{2\delta + (\bar{\rho} - 3)(1-\varphi)\mu} & \text{for } 1 - \dfrac{\delta}{2\mu} \le \varphi \le 1; \end{cases}$$

where $\rho_D > \bar{\rho}$ implies that the ratio is decreasing in φ on $(2\mu - \delta)/(2\mu) \le \varphi \le 1$ with value 1 at $\varphi = 1$. From (8) we can conclude that when $\rho_D > \bar{\rho}$ then the following statements apply:

A) the inferred actuarially fair premium is smaller than it should be, i.e., yield insurance in the aggregate is underpriced;

B) the proportional magnitude of underpricing is most severe at low coverage levels.

Statement A) arises because the conditional variance of the aggregate is larger in state *D* than in state U whenever $\rho_D > \bar{\rho}$, a fact not recognized whenever average correlation is used to form premiums. Statement B) is due to averaging. Mispricing is most severe in the tails. Indemnifying higher yields allow more D state $(\mu-\delta,\mu)$ or $(\mu,\mu-\delta)$ outcomes to ameliorate the indemnity amount, reducing both absolute and proportional mispricing.

Pseudo-Gaussian Copula

In this subsection we conduct a simulation exercise for a generalized Gaussian copula. The exercise is similar to that in table 3 of G&H (2015), in which corn and soybean insurance contracts were considered for Illinois counties. We simulate and compare the premium rates of a hypothetical area-based yield insurance product under various coverage levels with and without tail dependence. We do so in three steps.

In Step 1 we estimate the Gaussian and Pseudo-Gaussian (PG) copulas using county level yield residuals from eqn. (6). The PG copula generalizes the regular Gaussian copula to admit flexibility in the characterization of higher moments. Following Fang and Madsen (2013), the PG copula density is defined as:[8]

$$
\begin{aligned}
c(u,v;\Theta) &= \frac{1}{K}\frac{1}{\sqrt{1-\rho^2(u,v;\Theta)}} \times \exp\left\{\frac{[\Phi^{-1}(u)]^2+[\Phi^{-1}(v)]^2}{2}\right\} \\
&\times \exp\left\{-\frac{[\Phi^{-1}(u)]^2-2\rho(u,v;\Theta)\Phi^{-1}(u)\Phi^{-1}(v)+[\Phi^{-1}(v)]^2}{2(1-\rho^2(u,v;\Theta))}\right\},
\end{aligned} \tag{9}
$$

where u and v are marginal distributions of random variables X_1 and X_2, i.e., $u=F_1(X_1)$ and $v=F_2(X_2)$. Parameter K is the normalizing factor and $\Phi^{-1}(\cdot)$ represents the inverse of the standard normal cumulative distribution. Expression $\rho(u,v;\Theta)$ is the generalized pairwise correlation functional for each pair (u,v) and depends on the unknown parameter vector $\Theta = (\alpha,\beta)'$. It can have any form subject to $\rho(u,v;\Theta)\in[-1,1]$. We modify Definition V in Fang and

Madsen (2013) and specify

(10) $\rho(u,v;\alpha,\beta)=\beta\exp[-\alpha(1-(1-u)(1-v))]$.

Parameter interpretation can be obtained from taking left tail limits for both variates, i.e., when $u\to 0$ and $v\to 0$. Then $\rho(u,v;\alpha,\beta)\to\beta$, giving β as the left tail correlation. Any shift away from the left tail such that $1-(1-u)(1-v)=r>0$ gives $\rho(u,v;\alpha,\beta)=\beta\exp[-\alpha r]$ so that if $\alpha>0$ then the shift leads to an estimated exponential decay from β toward zero. Thus α and β can be viewed as capturing convergence speed and tail shape, respectively.[9] Notice also that $\rho(u,v;\alpha,\beta)|_{\alpha=0}\equiv\beta$, in which case PG reduces to the Gaussian copula, allowing for a nested test of the Gaussian copula.

The parameters in Θ are estimated by the Maximum Likelihood method using the county level detrended yields, which are regression residuals from eqn. (6) for four groups of counties (two in IA and two in SD). The groups of counties are chosen to ensure a sufficient number of observations in each group as the NASS crop yield history for an individual county is limited to 1950-2014 only. For IA, the two groups of counties are {Calhoun, Pocahontas, Webster} and {Humboldt, Wright, Hamilton}. For SD, the sample counties are {Beadle, Jerauld, Aurora} and {Kingsbury, Sanborn, Davison}. Quantity X_i, $i\in\{1,2\}$, are the yield residuals of one county group in a state. The parameter estimate standard errors are obtained using the bootstrapping method.[10] The results are reported in table 3 where the last row reports the estimated Gaussian copula correlation coefficient. As α estimates are positive and many standard deviations above zero, the estimates find evidence in favor of the Pseudo-Gaussian over the Gaussian copula.

In Step 2 we draw random samples of yield residuals based on the estimated PG and Gaussian copulas and generate county level yield samples. We draw 5,000 random samples from the respective estimated Gaussian and PG copula. For the Gaussian copula, we use the generating function of *copularnd* in Matlab. For the PG copula, we following the two-step procedure suggested in Fang and Madsen (2013):

i) Generate a random uniform (0,1) sample, $\tilde{u}$.

ii) Generate a random sample $\tilde{v}$ from the conditional distribution of $c(u=\tilde{u},v;\Theta)$ for the given $\tilde{u}$ drawn in *i*). This gives us the desired $(\tilde{u},\tilde{v})$ pair. A random walk chain Metropolis–Hastings (M–H) algorithm (Koop, Poirier and Tobias 2007) is then applied to draw from the conditional density.

Figure 5 provides a 3-D surface plot of the PG copula using the simulated samples for Iowa. The surface indicates significant tail dependence. To emphasize the difference between Gaussian and PG copulas, figure 6 plots the surface and contour plots of the density difference, PG less Gaussian. The labeled elevations in figure 6-(b) further indicate the positive differences at the tails and negative differences at the center. The simulated $(\tilde{u},\tilde{v})$ pairs are transferred to yield residuals by inverting the empirical distribution functions of the marginals. The yield residuals are then converted to county level yields by adding back the sample means from 2014.

In Step 3 we calculate premium rates for an area yield insurance contract by aggregating county level yield losses. To illustrate the economic significance of tail dependence, we calculate expected indemnity payouts for Iowa and South Dakota. The yield guarantee depends on the chosen coverage level φ and yield loss is calculated as $0.5\max[\varphi(\bar{y}_1+\bar{y}_2)-y_1-y_2,0]$. Here $\bar{y}_1$ and $\bar{y}_2$ are mean yields for the two county groups in each state, while y_1 and y_2 are simulated yields based on the copula in question. Premiums are obtained by averaging over repeated independent draws from the respective distributions. The ratios of calculated premium rates at 75%, 85% and 95% coverage levels are reported in table 4. The results indicate that at lower coverage levels the relative divergence between premium rates is highest. At 85%

coverage the percent divergence is in the 20-25% range. At 75% coverage, rates differ by 50% in Iowa and more than 100% in South Dakota. In other words, and consistent with statements A) and B) above, the inability of the Gaussian copula to account for the tail dependence described in figure 6 generates more serious biases in estimated expected indemnities for lower coverage levels. Our findings corroborate those in table 3 of G&H (2015).

Climate Change Implications for Systemic Risk in Corn Yields

Research into how climate change may affect agricultural yield and quality levels supports the uncontroversial view that impact will be nonlinear in temperature and water availability (Schlenker and Roberts 2009; Xu et al. 2013; Tack, Barkley and Nalley 2015; Kawasaki and Uchida 2016). The existence of state-dependent correlations can also exacerbate nonlinear economic impacts. Crop can be stored forward but cannot be borrowed from the future. Grain stock-outs arising from widespread crop failure when carry-over stocks are low can induce much economic harm and large price spikes. If yields covary more when extreme weather events occur and if climate change brings more extreme weather events then, all else fixed, the costs of stock-outs and of guarding against them will increase.

To better understand how projected climate change might affect crop yields, especially regarding tail dependence, we simulate county-level yield of selected counties in Iowa and South Dakota under a climate change scenario and summarize the changes across counties.[11] Yield in county i year t ($\hat{y}_{i,t}$) is simulated using the LYRM specified in eqn. (6) where the coefficients, $\Omega = \{\gamma_0, \gamma_1, \ ... \ , \gamma_9, \lambda_1, \lambda_2\}$, take the estimated values in table 2, column 1. The Intergovernmental Panel on Climate Change (IPCC) A1B emissions scenario is chosen, under which the projected mean global temperature increase through 2090-'99 is bracketed as [1.7, 4.4°C] when benchmarked against 1980-'99 (Collins et al. 2013, figure 12.39). This model assumes rapid economic growth through this century but a balanced use of fossil and non-fossil energy sources. When comparing the 1981-2010 and 2031-2060 timeframes and April through August weather, the model projects an average increase in temperature of 2.2° C for Iowa and 2.5° C for South Dakota. It predicts 91.4 millimeter (18.6%) and 57.5 millimeter (24.3%) increases in precipitation, respectively, for Iowa and South Dakota.

Primary weather data (temperature and precipitation) are first constructed by a mean-shifting procedure using the daily projections generated by the Centre National de Recherches Météorologique's Coupled Global Climate Model (CNRM; Voldoire et al. 2011).[12] The global climate projections, both for past and future years over the range 1960-2099, were the eight degree (grid size) Conterminous U.S. (CONUS) Statistical Asynchronous Regional Regression products[13] downscaled to the county-level of disaggregation and available for all counties under scrutiny. The mean-shift approach imposes changes on weather variables derived from the simulated data, i.e., difference between a future and a historical time period, on the observed weather data over the same period. Specifically, the weather variable $W_{i,d,m,y}^{\text{Fut.}}$ for county i, day d, month m, year y used in the simulation is constructed as $W_{i,d,m,y}^{\text{Fut.}} = W_{i,d,m,y}^{\text{Past}} + (W_{i,d,m,y'}^{\text{Fut. Scen.}} - W_{i,d,m,y}^{\text{Past Scen.}})$, i.e., the historical observed weather $W_{i,d,m,y}^{\text{Past}}$ plus the day and month specific mean shift over year difference $y' - y$. We choose the historical period 1984-2014. The mean shift is generated by calculating the annual differences in weather variables between 2024-2054 and 1984-2014 using the CNRM projections. The mean shift corresponds to a 40 year lag so that the 2024 projection is constructed by shifting 1984 weather. Upon constructing future weather time series, Palmer's Z variables $\hat{P}_{i,t}^{L}$ and $\hat{P}_{i,t}^{R}$ as well as growing and stress degree days $\hat{G}_{i,t}$ and $\hat{S}_{i,t}$ are then calculated.

The data generation process for the projected weather variables is briefly described as:

i) Download daily projections of related weather variables including maximum and minimum temperatures as well as precipitation over 1984-2054.

ii) Calculate the shift in weather variables for the sample period across 40 years, i.e., $W_{i,d,m,y'}^{\text{Fut. Scen.}} - W_{i,d,m,y}^{\text{Past Scen.}}$.

iii) Impose these shifts on historical weather data for 1984-2014 to obtain projected weather.

iv) Construct growing and stress degree days over the growing season (May to September) for each year following the definitions discussed previously. Monthly Palmer's Z is calculated by fitting a linear regression model, see Item 3 of the supplementary appendix online. As in the estimation of eqn. (6), for each year we use the July Palmer's Z.

v) Simulate the yield under the new weather data set using the coefficient estimates in table 2.

County-level yields are simulated under the projected weather variables for six counties in Iowa and twelve counties in South Dakota.[14] Detrended yields, generated by subtracting the linear time trend from the simulated yields, are used for the analyses in this section. To control for the impact of empirical modeling on the yield structure, for comparison we simulate the county yield (detrended) under historical weather data over the same sample period.

Figure 7 present the one-to-one county yield pairs for each of observed historical yield (observed yield, panel a) for Iowa and panel d) for South Dakota), simulated yield under weather history (simulated historical yield, panels b) and e)), and simulated yield under projected weather (simulated projected yield, panels c) and f)) to illustrate the impact of future climate on tail dependence. As expected, under future climate scenarios county-level yields exhibit stronger tail dependence when compared with both observed and simulated historical yields. This strengthening is especially true for South Dakota. Comparing observed yield with simulated historical yield, we realize that for the selected counties the model fits Iowa's yield records better than those of South Dakota. Tail monotonicity statistic LTD is calculated based on paired county yields and the results are reported in table 5. Reading down each column, the table indicates left-tail decreasing behavior between county pairs in both Iowa and South Dakota. Furthermore, tail dependence is generally stronger under projected weather.

Concluding Remarks

The literature to date has provided remarkably little guidance on how weather and soils interact in determining yields for commercially grown crops. In this paper we have sought to do five things. With county-level data we have inquired into how interactions between weather and soils affect yields, finding complementarity between heat variables and soil quality but substitution between moisture availability and soils. Focusing on the latter, we have developed a model that emphasizes the role of natural resources in determining the structure of yield-yield correlations. More specifically, the model connects what we refer to as land yield resilience with heterogeneities in land and climate endowments so that yield-yield correlation structures can be related to growing conditions in a given year.

Thirdly, we have scrutinized yield-yield conditional correlation structures using county-level yield data and a generalization of the Gaussian copula. The additional parameter intended to account for tail dependence is significant and of the expected sign, indicating very strong dependence in the left tail and weaker dependence away from it. We then show that area yield insurance contracts are likely to be underpriced when tail dependence is ignored, and that relative mispricing will be most pronounced at low coverage levels. Our final contribution has been to use our empirical model together with downscaled climate projection data to posit that climate change will, all else fixed, increase tail dependence among yields so that aggregate yields will become more variable and demand will increase for storage across crop years.

We are not the first to assert the presence of tail-dependent yields and state-dependent correlations among yields. As best we can establish, we are however the first to provide a formal production technology framework for understanding why and how yield covariability should change with soil and weather conditions. Soil and weather are features over which humans have limited control. Substantial advances are being made in the availability of integrated soil, weather and crop choice data as well as in computation methods to study such large datasets. These data may, together with the resource use challenges that society faces, motivate further development on how climate, soils and other natural endowments interact in the determination of input choices and resulting outputs.

We close by stating a further concern about rate-setting. Constant price-yield correlation structures (whether cardinal or rank) are typically assumed when pricing revenue insurance (Coble et al. 2010). Deaton and Laroque (Theorem 1, 1992) demonstrate that price movements should become more sensitive to shocks when stocks decline. For this reason, price-yield correlations should become more negative when stocks are low and growing conditions ominous relative to when the converse applies. Thus the natural price-yield hedge may strengthen when most needed. This conjecture may be more reassuring to revenue contract underwriters than the finding that yield correlations strengthen in tails. In any case, modeling structures that force constant price-yield correlations warrant empirical scrutiny.

References

Aas, K. 2004. "Modelling the Dependence Structure of Financial Assets: A survey of four copulas." Unpublished manuscript, Norwegian Computing Center, NR Note no. SAMBA/22/04.

Ahmed, O., and T. Serra. 2015. "Economic Analysis of the Introduction of Agricultural Revenue Insurance Contracts in Spain using Statistical Copulas." *Agricultural Economics* 46(1):69-79.

Arora, G., H. Feng, D.A. Hennessy, and C.J. Anderson. 2016. "Crop Competitiveness and Future Climate Change in the Northern Great Plains." Paper presented at AAEA annual meeting, Boston MA, July 31 – August 2.

Berck, P., and G. Helfand. 1990. "Reconciling the von Liebig and Differentiable Crop Production Functions." *American Journal of Agricultural Economics* 72(4):985-996.

Bot, A., and J. Benites, eds. 2005. *Drought-resistant Soils: Optimization of soil moisture for sustainable plant production.* Proceedings of the Electronic Conference organized by the FAO of the United Nations. Rome: Food and Agriculture Organization.

Coble, K.H., T.O. Knight, B.K. Goodwin, M.F. Miller, and R.M. Rejesus. 2010. *A Comprehensive Review of the RMA APH and COMBO Rating Methodology Final Report 2010.* A report completed by Sumaria Systems for the United States Department of Agriculture, Risk Management Agency. Available at http://www.rma.usda.gov/pubs/2009/comprehensivereview.pdf.

Collins, M., R. Knutti, J. Arblaster, J.L. Dufresne, T. Fichefet, P. Friedlingstein, X. Gao, W.J Gutowski, T. Johns, G. Krinner, M. Shongwe, C. Tebaldi, A.J. Weaver, and M. Wehner. 2013. "Long-term Climate Change: Projections, Commitments and Irreversibility." In

Stocker, T.F., Qin, D., Plattner, G.K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (eds.), Ch. 12, pp. 1029-1136, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Dasgupta, P., and E. Maskin. 1987. "The Simple Economics of Research Portfolios." *Economic Journal* 97(387):581-595.

Deaton, A., and G. Laroque. 1992. "On the Behavior of Commodity Markets." *Review of Economic Studies* 59(1):1-23.

Du, X., D.A. Hennessy, and H. Feng. 2014. "A Natural Resource Theory of U.S. Crop Insurance Contract Choice." *American Journal of Agricultural Economics* 96(1):232-251.

Du, X., D.A. Hennessy, and C. Yu. 2012. "Testing Day's Conjecture that More Nitrogen Decreases Crop Yield Skewness." *American Journal of Agricultural Economics* 94(1):225-237.

Du, X., C.L. Yu, D.A. Hennessy, and R. Miao. 2015. "Geography of Crop Yield Skewness." *Agricultural Economics* 46(4):463-473.

Fang, Y., and L. Madsen. 2013. Modified Gaussian Pseudo-copula: Applications Insurance and Finance. *Insurance: Mathematics and Economics* 53: 292-301.

Furman, E., A. Kuznetsov, J. Su, and R. Zitikis. 2016. "Tail Dependence of the Gaussian Copula Revisited." *Insurance: Mathematics and Economics* 69(July):97-103.

Goodwin, B.K. 2001. "Problems with Market Insurance in Agriculture." *American Journal of Agricultural Economics* 83(3):643-649.

Goodwin, B.K. 2015a. "Agricultural Policy Analysis: The Good, The Bad, and The Ugly."

American Journal of Agricultural Economics 97(2):353-373.

Goodwin, B.K. 2015b. "Challenges in the Design of Crop Revenue Insurance." *Agricultural Finance Review* 75(1):19-30.

Goodwin, B.K., and V.H. Smith. 2013. "What Harm is Done by Subsidizing Crop Insurance?" *American Journal of Agricultural Economics* 95(2):489-497.

Goodwin, B.K., and A. Hungerford. 2015. "Copula-Based Models of Systemic Risk in U.S. Agriculture: Implications for Crop Insurance and Reinsurance Contracts." *American Journal of Agricultural Economics* 97(3):879-896.

Hao, L., and D. Q. Naiman. 2007. *Quantile Regression*. Thousand Oaks, CA: Sage Publications.

Heddinghaus, T.R., and P. Sabol. 1991. "A Review of the Drought Severity Index and Where Do We Go from Here?" Paper presented at the 7th Conference on Applied Climatology, American Meteorological Society, Boston MA, 10-13 September.

Helms, D. 1992. "The Development of the Land Capability Classification." In D. Helms (Ed.), pp. 60-73, *Readings in the History of the Soil Conservation Service*. Washington, DC: Soil Conservation Service http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1043484.pdf

Hennessy, D.A. 2009a. "Crop Yield Skewness Under Law of the Minimum Technology." *American Journal of Agricultural Economics* 91(1):197-208.

Hennessy, D.A. 2009b. "Crop Yield Skewness and the Normal Distribution." *Journal of Agricultural and Resource Economics* 34(1):34-52.

Iman, R.L., and W.J. Conover. 1982. "A Distribution-free Approach to Inducing Rank Correlation among Input Variables." *Communications in Statistics Part B – Simulation and Computation* 11(3):311-334.

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

Karl, T.R. 1986. "The Sensitivity of the Palmer Drought Severity Index and Palmer's Z Index to their Calibration Coefficients including Potential Evapotranspiration." *Journal of Applied Meteorology and Climatology* 25(1):77-86.

Kawasaki, K., and S. Uchida. 2016. "Quality Matters More than Quantity: Asymmetric Temperature Effects on Crop Yield and Quality Grade." *American Journal of Agricultural Economics* 98(4):1195-1209.

Koop, G., D.J. Poirier, and J.L. Tobias. 2007. *Bayesian Econometric Methods.* Cambridge, UK: Cambridge University Press.

Mildenhall, S.J. 2005. "The Report of the Research Working Party on Correlations and Dependencies Among All Risk Sources: Part 1-Correlation and Aggregate Loss Distributions with an Emphasis on the Iman-Conover Method." Research Working Party on Correlations and Dependencies Among All Risk Sources, Casualty Actuarial Society. Available at https://www.casact.org/research/index.cfm?fa=workingparty-corr.

National Oceanic and Atmospheric Administration (NOAA). 2014. Statewide Regional Drought. Available online at ftp://ftp.ncdc.noaa.gov/pub/data/cirs/climdiv/drought-readme.txt, visited October 4, 2016.

Patton, A.J. 2012. Copula Methods for Forecasting Multivariate Time Series. In G. Elliott and A. Timmermann (eds.) Handbook of Economic Forecasting, Volume 2, Springer Verlag.

Qiu, F., and J. Rude. 2016. "Extreme Dependence in Price Transmission." *Applied Economics* 48:4379-4392.

Salmon, F. 2012. "The Formula that Killed Wall Street." *Significance* 9(1):16-20.

Schlenker, W., and M.J. Roberts. 2009. "Nonlinear Temperature Effects Indicate Severe

Damages to U.S. Crop Yields under Climate Change." *Proceedings of the National Academies of Sciences* 106(37):15594-15598.

Staudt, A. 2010. "Tail Risk, Systemic Risk and Copulas." *Casualty Actuarial Society E-Forum* Fall Volume 2, https://www.semanticscholar.org/paper/Tail-Risk-Systemic-Risk-and-Copulas-Staudt-FCAS/c335a2be95a216cfeb9d207156999c931170f59b, visited 10/22/2017.

Tack, J.B., and M.T. Holt. 2016. "The Influence of Weather Extremes on the Spatial Correlation of Corn Yields." *Climatic Change* 134(1-2):299-309.

Tack, J.B., A. Barkley, and L.L. Nalley. 2015. "Effect of Warming Temperatures on US Wheat Yields." *Proceedings of the National Academies of Sciences* 112(22):6931-6936.

Trivedi, P.K., and D.M. Zimmer. 2005. "Copula Modeling: An Introduction for Practitioners." *Foundations and Trends® in Econometrics* 1(1):1-111.

Voldoire, A., E. Sanchez-Gomez, D. Salas y Mélia, B. Decharme, C. Cassou, S.Sénési, S. Valcke, I. Beau, A. Alias, M. Chevallier, M. Déqué, J. Deshayes, H. Douville, E. Fernandez, G. Madec, E. Maisonnave, M.-P. Moine, S. Planton, D.Saint-Martin, S. Szopa, S. Tyteca, R. Alkama, S. Belamari, A. Braun, L. Coquart, F. Chauvin. 2011. "The CNRM-CM5.1 Global Climate Model: Description and Basic Evaluation." *Climate Dynamics* 40(9):2091-2121.

Wilhelmi, O.V., and D.A. Wilhite. 2002. "Assessing Vulnerability to Agricultural Drought: A Nebraska Case Study." *Natural Hazards* 25(1):37-58.

Xu, Z., D.A. Hennessy, K. Sardana, and G. Moschini. 2013. "The Realized Yield Effect of Genetically Engineered Crops: US Maize and Soybean." *Crop Science* 53(3):735-745.

Zimmer, D.M. 2012. "The Role of Copulas in the Housing Crisis." *Review of Economics and Statistics* 94(2):607-620.

Zimmer, D.M. 2015. "Crop Price Comovements During Extreme Market Downturns."

Australian Journal of Agricultural and Resource Economics 60(2):265-283.

Table 1: Panel Data Regression of County Average Corn Yield on Weather and Land Quality Variables

Variable	Coeff.	*t* stat	Coeff.	*t* stat	Coeff.	*t* stat	Coeff.	*t* stat
	Eqn. (4.1)		Eqn. (4.2)		Eqn. (4.3)		Eqn. (4.4)	
Trend	1.63[a]	408.48	1.66[a]	104.95	1.63[a]	408.75	1.66[a]	104.85
Trend Square	—	—	-0.0004[b]	-1.64	—	—	-0.0004	-1.59
$P_{i,t}^{L}$	31.91[a]	75.47	31.91[a]	75.48	32.09[a]	75.76	31.84[a]	75.28
$P_{i,t}^{R}$	-41.30[a]	-40.16	-41.34[a]	-40.19	-41.19[a]	-39.97	-41.27[a]	-40.48
$G_{i,t}$	37.23[a]	26.13	37.28[a]	26.16	35.48[a]	24.92	37.37[a]	26.21
$S_{i,t}$	-101.87[a]	-68.37	-101.86[a]	-68.36	-97.07[a]	-66.52	-101.59[a]	-68.15
$P_{i,t}^{L} \times Q_i$	-15.67[a]	-8.55	-15.68[a]	-8.56	-2.31	-1.43	—	—
$P_{i,t}^{R} \times Q_i$	3.89	0.87	3.80	0.85	6.64	1.49	—	—
$G_{i,t} \times Q_i$	19.62[a]	3.14	19.53[a]	3.12	—	—	17.30[a]	2.77
$S_{i,t} \times Q_i$	-94.43[a]	-14.32	-94.47[a]	-14.32	—	—	-68.93[a]	-11.67
Adjusted R^2		0.8082		0.8082		0.8072		0.8079

Notes: Superscripts a/b indicate significant at 1%/10% level. Estimates of the "constant" (not reported) are close to zero and insignificant.

Table 2: Bayesian Model Estimates of Average Corn Yield on Weather and Land Quality Variables

Variable	Coeff.	Prob > 0	Coeff.	Prob > 0
	Eqn. (6.1)		Eqn. (6.2)	
γ_0, Constant	-0.0001	0.496	0.0001	0.503
γ_1, Trend	1.642	1	1.642	1
Trend Square	—	—	—	—
γ_2, $P_{i,t}^{L}$	6.461	1	7.052	1
γ_3, $P_{i,t}^{R}$	3.946	0.724	2.163	0.626
γ_4, $G_{i,t}$	7.801	0.999	8.380	1
γ_5, $S_{i,t}$	-134.913	0	-137.059	0
γ_6, $G_{i,t} \times Q_i$	-6.135	0.126	—	—
γ_7, $S_{i,t} \times Q_i$	-45.415	0	—	—
λ_1, in $P_{i,t}^{L}$ resilience	-0.124	0	-0.053	0.003
λ_2, in $P_{i,t}^{R}$ resilience	-0.029	0.193	-0.051	0.049
γ_8, $P_{i,t}^{L}$ resilience coefficient	-43.313	0	-42.408	0
γ_9, $P_{i,t}^{R}$ resilience coefficient	-38.291	0	-36.628	0
Skewness	0.025	0.525	-0.013	0.478
Variance of error	262.876	1	265.784	1

Table 3: Pseudo Gaussian copula Estimation results for Counties in Iowa and South Dakota

	Iowa	South Dakota
α	0.3002 (< 0.001)	1.2113 (< 0.001)
β	0.9821 (< 0.001)	0.9805 (< 0.001)
ρ_{12}	0.8076 (< 0.001)	0.4399 (< 0.001)

Note: Bootstrap standard errors are in the parentheses.

Table 4. Relative Premium Rates for Area Yield Insurance

Ratio, Pseudo Gaussian rate over Gaussian Copula rate	75%	85%	95%
Iowa	1.50	1.19	1.07
South Dakota	2.11	1.24	1.03

Table 5. Test Statistics for Left Tail Decreasing Property (31-year yield records; 6 counties in Iowa and 12 counties in South Dakota; standard errors in the parentheses)

LTD Statistic, Iowa	Observed yield			Simulated yield (weather history)			Simulated yield (projected weather)		
u / *v*	0.1	0.3	0.5	0.1	0.3	0.5	0.1	0.3	0.5
0.1	0.58	0.93	1.00	0.76	0.89	0.98	0.76	1.00	1.00
	(0.20)	(0.14)	(0.00)	(0.15)	(0.16)	(0.09)	(0.15)	(0.00)	(0.00)
0.3	0.30	0.79	0.98	0.32	0.68	0.89	0.33	0.90	0.96
	(0.05)	(0.10)	(0.05)	(0.04)	(0.16)	(0.09)	(0.00)	(0.10)	(0.06)
0.5	0.20	0.57	0.85	0.20	0.51	0.80	0.20	0.59	0.86
	(0.00)	(0.03)	(0.05)	(0.02)	(0.07)	(0.09)	(0.00)	(0.02)	(0.05)
LTD Statistic, South Dakota									
0.1	0.35	0.67	0.80	0.79	1.00	1.00	0.76	0.95	0.99
	(0.37)	(0.37)	(0.24)	(0.16)	(0.00)	(0.00)	(0.15)	(0.12)	(0.06)
0.3	0.25	0.61	0.82	0.33	0.75	0.91	0.33	0.81	0.99
	(0.11)	(0.16)	(0.14)	(0.00)	(0.10)	(0.08)	(0.03)	(0.08)	(0.03)
0.5	0.19	0.53	0.79	0.20	0.57	0.82	0.20	0.60	0.92
	(0.02)	(0.07)	(0.09)	(0.00)	(0.05)	(0.08)	(0.00)	(0.01)	(0.05)

Figure 1. Yield, Weather Favorability, Soil Quality Relation

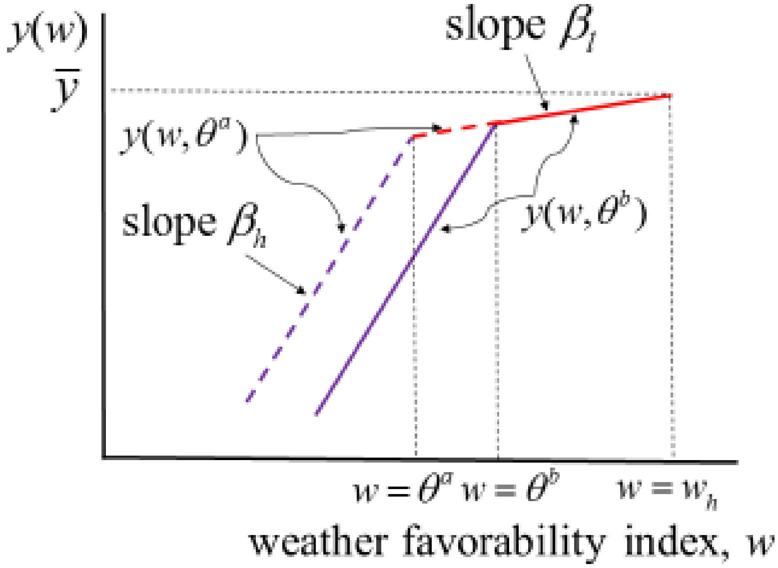
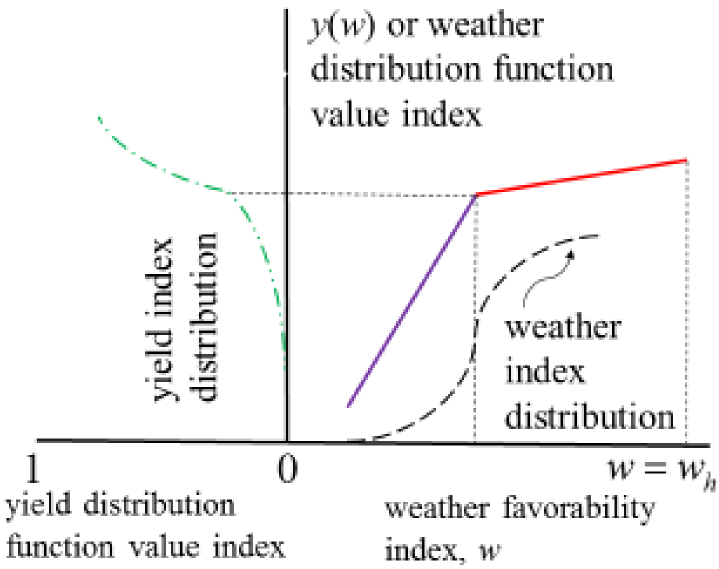
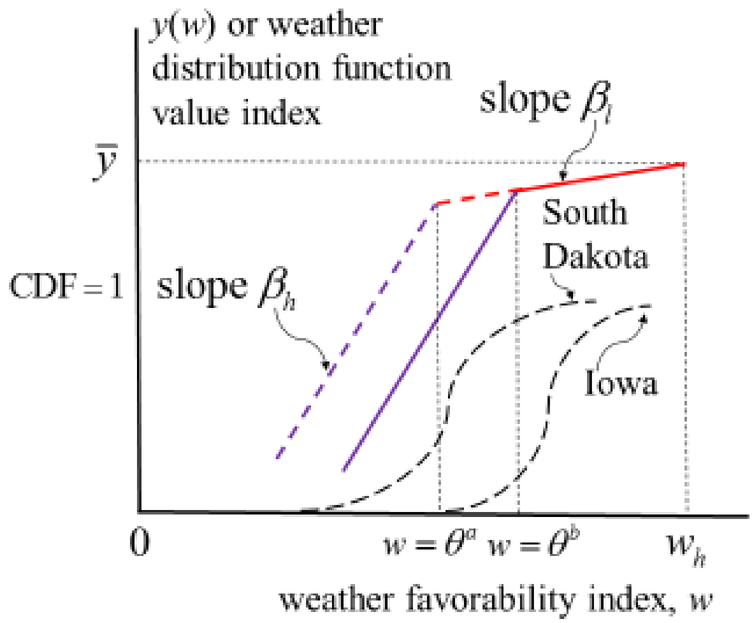


Figure 2. Map between Weather Index Distribution and Yield Distribution



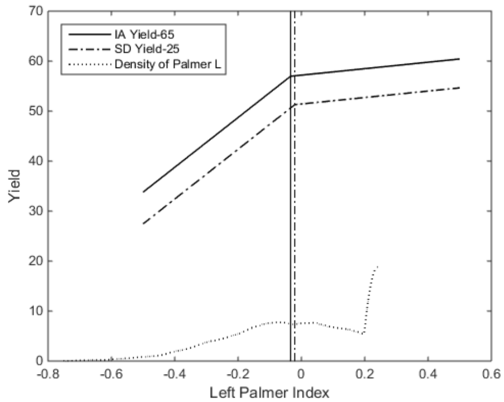
Notes: For the right quarter plane, the vertical axis measures two quantities. For the production relation it measures yield evaluation. For the weather index cumulative distribution it measures, up to a scaling factor for convenience of display, the distribution's cumulative value. For the left quarter plane, the vertical axis measures yield evaluation while the horizontal axis measures yield distribution cumulative value.

Figure 3. Characterizing Yield and Weather Favorability Differences across States

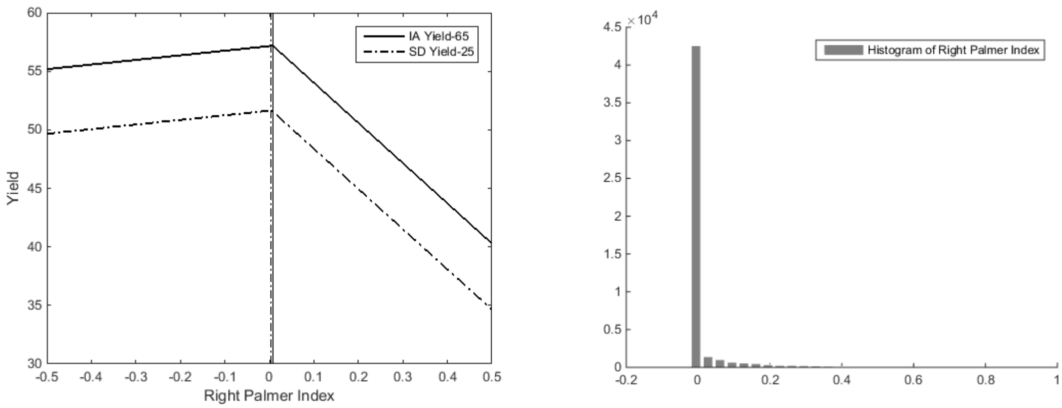


Notes: Vertical axis measures two quantities. For the production relation it measures yield. For the weather index cumulative distributions it measures, up to a scaling factor for convenience of display, a distribution's cumulative value or CDF.

Figure 4. Predicted Yields as Left and Right Palmer Indices Change, Iowa and South Dakota



(a) Left Palmer Index



(b) Right Palmer Index

Notes: *i*) In (a) the Left Palmer Index density is estimated and generated by Matlab's *epanech2* nonparametric method. Density value is multiplied by 5 to scale for presentation with the yield curve. In (b), the Right Palmer Index histogram is included because nonparametric methods don't work given the variable's sparse distribution. *ii*) Vertical lines indicate kink points for corresponding yield curves. *iii*) So that both functions can fit on the graph, Iowa and South Dakota yields were displaced downward by 65 bushels and 25 bushels respectively.

Figure 5. Surface and Contour plots for PG Copula, Iowa

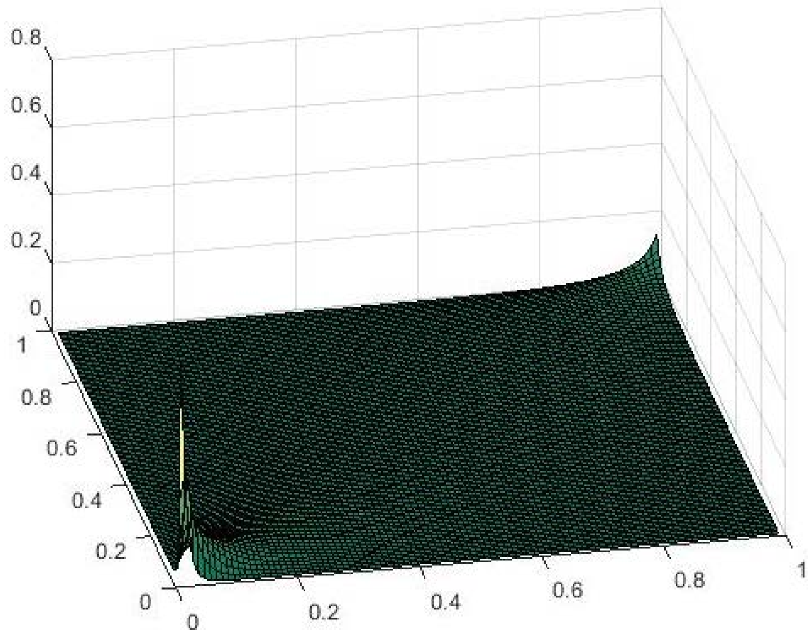
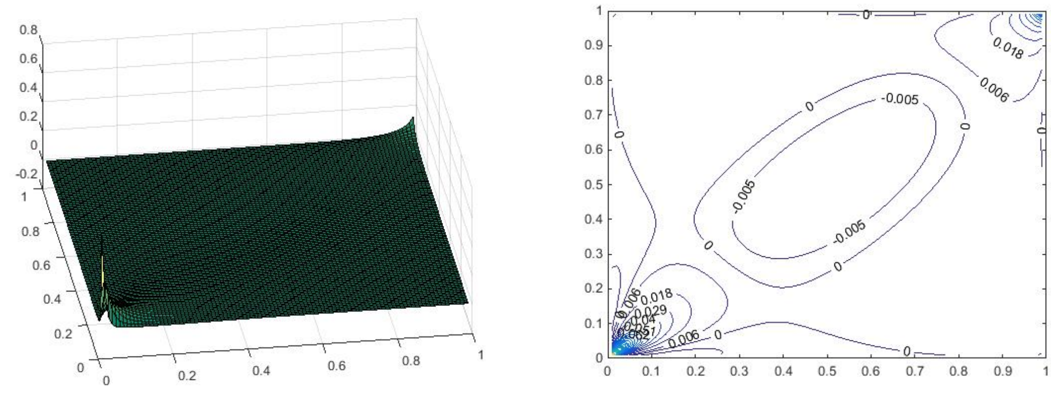
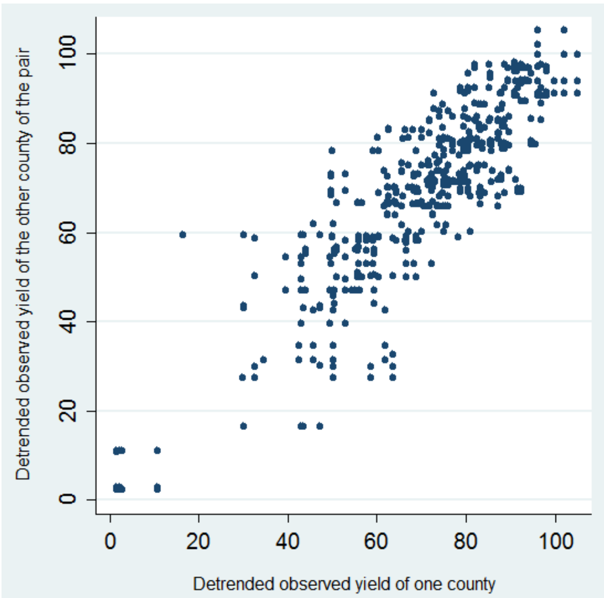


Figure 6. Surface and Contour Plots of the difference between PG and Gaussian Copula Densities, PG less Gaussian

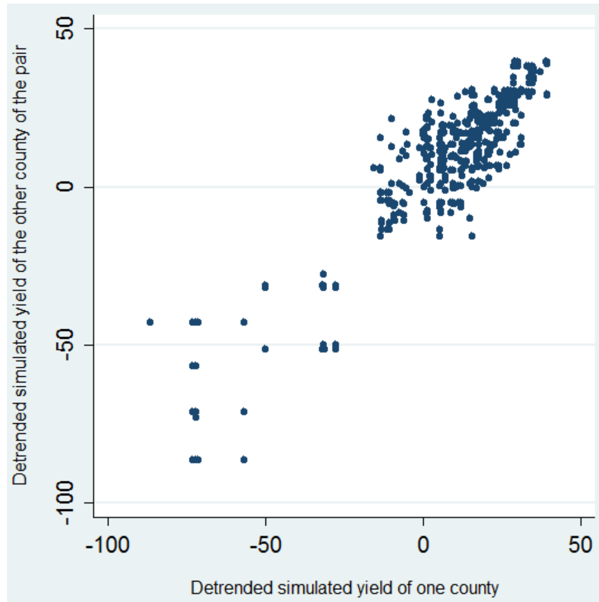


(a) (b)

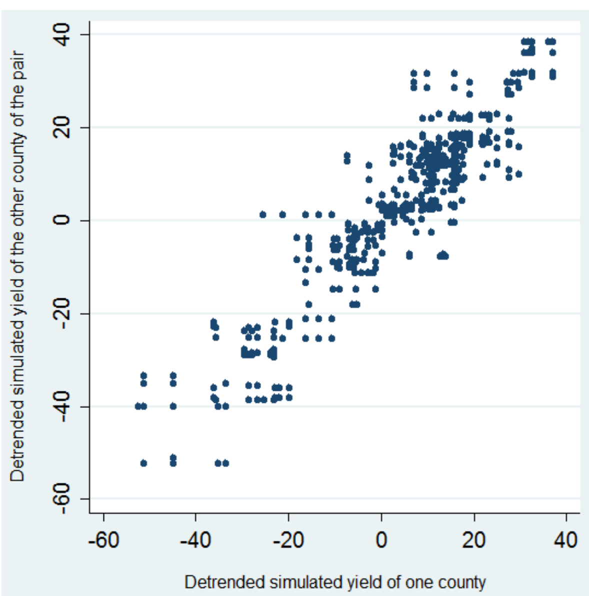
Figure 7. Scatter Plots of Observed and Simulated Yield, both Detrended, under Historical and Projected Climate



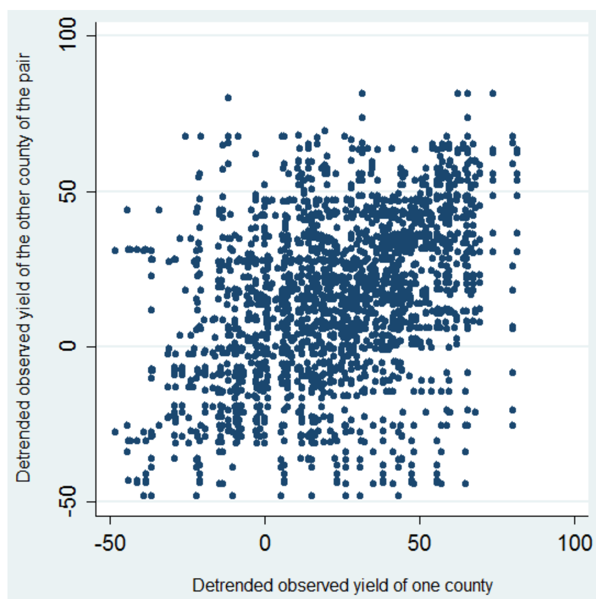
(a)



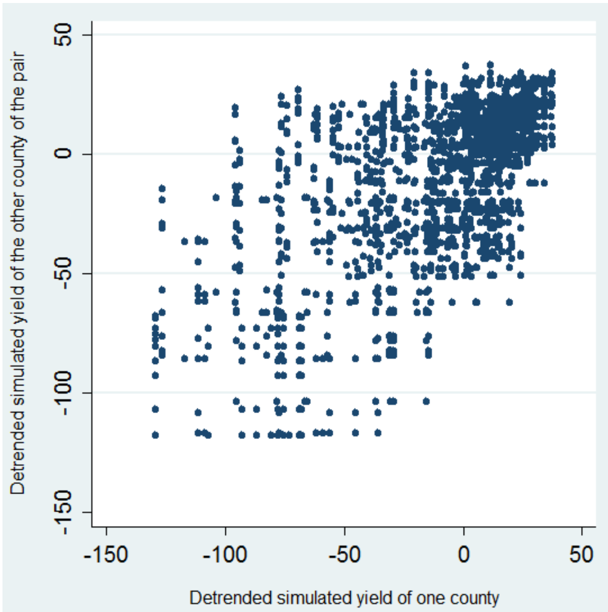
(b)

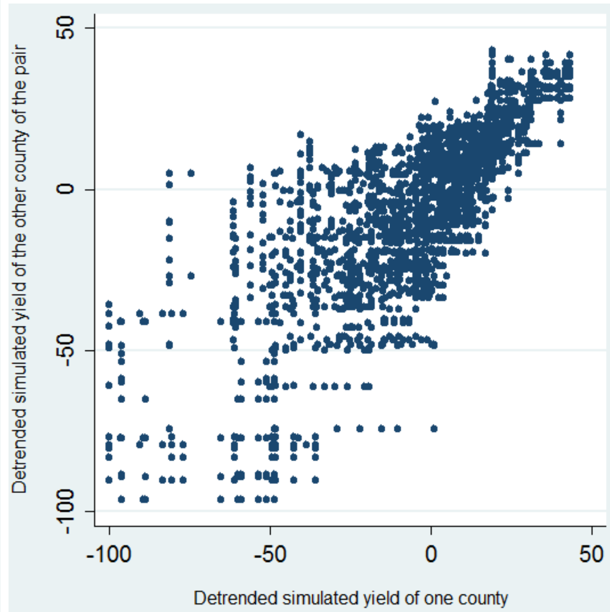


(c)



(d)





(e) (f)

Detrended

(a) observed county yield pairs (Iowa; 1984-2014)

(b) simulated county yield pairs under historical weather data (Iowa; 1984-2014)

(c) simulated county yield pairs under projected weather data (Iowa)

(d) observed county yield pairs (S. Dakota; 1984-2014)

(e) simulated county yield pairs under historical weather data (S. Dakota; 1984-2014)

(f) simulated county yield pairs under projected weather data (S. Dakota)

Footnotes

1. See Assigned Risk Fund Retention Clause (3)(A) on p. 14 in the 2018 Standard Reinsurance Agreement, http://www.rma.usda.gov/pubs/ra/sraarchives/18sra.pdf

2. Support choice $[w_l, w_h]$ is arbitrary so long as θ values are calibrated to be within the interval.

3. For instance, commence with all probability mass on the line $x_1 \equiv x_2$, in which case both definitions apply. Notice the asymmetry in definitions, where LTD averages from below and RTI averages from above. Then carefully relocate probability mass into set $\{(x_1, x_2): x_1 > \hat{x}_1,$ $x_2 \leq \hat{x}_2\}$ such that one definition applies but the other does not.

4. These definitions, while standard, are controversial as they require that convergence to the tail occur along the diagonal. Furman et al. (2016) generalize to a larger set of conversion paths.

5. Seminal work in on drought vulnerability has emerged from the U.S. National Drought Mitigation Center where Wilhelmi and Wilhite (2002) is illustrative. They proposed a drought vulnerability index that sums soil attribute metrics which contribute to the likelihood of water deficit. Their work is similar to ours in that the vulnerability index is i) constructed to take soil quality and water availability as substitutes; ii) regional with connotation that beyond a certain threshold a drought event happens in the sense that adverse yields are systemic to the region.

6. Not all sampled counties have complete yield records over our sample period, 1950-2014.

7. Iowa counties are Calhoun, Pocahontas, Webster, Humboldt, Wright and Hamilton. South Dakota counties are Kingsbury, Brookings, Sanborn, Miner, Lake, Moody, Davison, Hanson, McCook and Minnehaha.

8. This is Eqn. (5), p. 294, Fang and Madsen (2013).

9. See a detailed discussion of PG copula properties in Fang and Madsen (2013).

10. The bootstrap standard errors are calculated following the method in Patton (2012).

11. We recognize that adaptation will occur such that technological change will be endogenous to climate change. For example, seed companies may invest more in developing drought-tolerant seed. However, the possible impacts of climate change are diverse. Accounting for the processes involved in such adaptation is beyond our work's scope and we are not aware of efforts to account for the totality of these possible effects.

12. Weather projections data are available from several climate models. Our experience with the CNRM model leads us to believe that its weather outputs for the Dakotas are representative of the near-median weather projections from many climate models. It also contains no missing values of generated weather variables for this area.

13. See USGS Geo-data Portal,

http://cida.usgs.gov/gdp/client/#!catalog/gdp/dataset/54dd5e31e4b08de9379b38f3

14. The Iowa counties are as in footnote 7. The South Dakota counties are Kingsbury, Brookings, Lake, Moody, Hanson, McCook, Minnehaha, Sully, Hand, Beadle, Jerauld and Brule. Relative to those in footnote 7, some counties in South Dakota had to be replaced because they had incomplete historical yield records over 1984-2014.