THREE ESSAYS ON INSURANCE AND OPTIONS

IN AGRICULTURAL ECONOMICS

By

EUNCHUN PARK

Bachelor of Science in Food and Resource Economics
Korea University
Seoul, South Korea
2010

Master of Science in Food and Resource Economics
Korea University
Seoul, South Korea
2013

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
May, 2017

THREE ESSAYS ON INSURANCE AND OPTIONS

IN AGRICULTURAL ECONOMICS

Dissertation Approved:

Dr. B Wade. Brorsen

Dissertation Adviser

Dr. Brian Adam

Dr. Rodney Jones

Dr. Shu Yan

Outside Committee Member

Name: EUNCHUN PARK

Date of Degree: MAY, 2017

Title of Study: THREE ESSAYS ON INSURANCE AND OPTIONS IN AGRICULTURAL ECONOMICS

Major Field: AGRICULTURAL ECONOMICS

Abstract: Rating of insurance policies depends on the probability of events in the tail of the distribution. A modeling tool based on extreme value theory to measure such tail-related risk could potentially improve insurance rating. It is also widely agreed that crop yield distributions are spatially correlated. Considering spatial correlation may provide more precisely rated policies. In this context, the first essay provides two innovations in rating area yield crop insurance. One is to provide a method to fit the tail of crop yield distributions using the Generalized Pareto Distribution (GPD), a member of the family of extreme value distributions that models only the tail of the distribution. The second innovation is to estimate parameters of the distribution using a Bayesian Kriging approach that provides spatial smoothing of GPD parameters. Our results demonstrate that the proposed model dominates the existing method in out-of-sample performance and substantially mitigates regional inequalities in crop insurance loss ratios. Specifically, the new model shows considerable improvement in rating deeper tail probability.

In the second essay, we extend the Kriging method to a spatial smoothing based on a climate space, which is composed of climatological measures. We compare the spatial smoothing from the climate space and a general physical space (longitude-latitude space) to evaluate the performance of each method. We test the performance for county level yearly corn yield data from six U.S. states. Spatial smoothing from climate space dominates the results from the physical space in out-of-sample prediction and mitigates regional inequalities in crop insurance loss ratios. The climate space notably outperforms the physical space in Colorado that has various climates due to its varying topography.

The third essay develops a new method for pricing calendar spread options (CSO) that uses a calendar spread as its underlying asset. Previous studies have suggested CSO pricing models based on a joint process of two futures prices. However, CSO for storable commodities have distinct features due to physical aspects of the underlying assets. To address such aspects, we introduce a new CSO pricing model based on the theory of storage. Our model incorporates three stochastic processes and allows non-zero correlation structure among the processes to reflect the dynamics of the calendar spread. Option prices from the previous models are estimated to evaluate the performance of the new model. The new model proposed here outperforms existing models.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER I

USING BAYESIAN KRIGING FOR SPATIAL SMOOTHING IN CROP INSURANCE RATING

Abstract

Rating of insurance policies depends on the probability of events in the tail of the distribution. A modeling tool based on Extreme Value Theory to measure such tail-related risk could potentially improve insurance rating. It is also widely agreed that crop yield distributions are spatially correlated. Considering spatial correlation may provide more precisely rated policies. In this context, this research provides two innovations in rating area yield crop insurance. One is to provide a method to fit the tail of crop yield distributions using the Generalized Pareto Distribution (GPD), a member of the family of extreme value distributions that models only the tail of the distribution. The second innovation is to estimate parameters of the distribution using a Bayesian Kriging approach that provides spatial smoothing of GPD parameters. The proposed model provides estimates of the spatial structure across regions such as the maximum distance of the spatial effect. Our results demonstrate that the proposed model dominates the existing method in out-of-sample performance and substantially mitigates regional inequalities in crop insurance loss ratios. Specifically, the new model shows considerable improvement in rating deeper tail probability.

Key words: Area yield insurance rating, Bayesian Kriging, Bayesian spatial smoothing, extreme value theory, Gaussian spatial process, spatial correlation

Introduction

Adverse selection and moral hazard are common concerns when designing insurance products. Area-yield insurance can avoid the problems of moral hazard and transactions costs arising from individual crop insurance policies since the indemnity payment is based on average yields of the county rather than the actual yields of the individual insured. A poorly rated product, however, does not protect against adverse selection and regional inequities since a producer who has sufficient information about the county's crop yield distribution could avoid/purchase overpriced/underpriced area-yield insurance.

Previous literature (Coble, Dismukes, and Glauber 2007; Harri et al. 2011) shows that incorrectly rated premiums may create significant economic loss to the Federal Crop Insurance Corporation (FCIC) under the Standard Reinsurance Agreement (SRA). Federally managed area-yield insurance has been offered by the Risk Management Agency (RMA) for major crops in the United States. Area Yield Protection (AYP)[1] currently offers insurance based on county level yield data from the National Agricultural Statistics Service (NASS) or other sources of data (RMA, 2013). In addition, revenue-based, area-level insurance known as Area Revenue Protection (ARP) and the harvest price exclusion version of ARP (ARP-HPE) use the same crop yield modeling methods as AYP. Further, two new area-based insurance products: the Stacked Income Protection Plan (STAX), which is available only for cotton and the Supplemental Coverage Option (SCO), which is available for cotton and other crops, were made available with the 2014

farm bill (P.L. 113-79). Crop yield modeling methods for these two new products are also similar to the methods used for AYP.

Several studies have developed methods to calculate premium rates for area yield crop insurance. Skees, Black, and Barnett (1997) propose a premium rating method that is still the base model of current U.S. area crop insurance. They used a two-knot, linear spline function to fit technological changes of crop yield and calculate premium rates using residuals of the regression. Goodwin and Ker (1998), Ker and Goodwin (2000), Ker and Coble (2003), and Harri et al. (2011) propose extensions. These proposed methods are based on a two-step procedure. They first fit the temporal process of crop yield using historical observations of each county, and then estimate conditional crop yield density from the predicted yield level and heteroscedasticity adjusted residuals. Current policies are rated with the Harri et al. (2011) approach and some proprietary adjustments. Harri et al. (2011) use a hybrid approach where parameters are estimated at the county level but are restricted by bounds determined at the district level. Other past studies have considered a wide variety of crop yield distributions such as the beta distribution (Nelson and Preckel 1989; Tirupattur, Hauser, and Chaherli 1996), the log-normal distribution (Tirupattur, Hauser, and Chaherli 1996; Jung and Ramezani 1999; Stokes 2000), and normal (Just and Weninger 1999). In addition to these distributions, Sherrick et al. (2004) also consider the logistic and Weibull distributions. Others have suggested a nonparametric kernel density approach (Ker and Goodwin 2000), or a semi-parametric approach (Ker and Coble 2003).

Crop yields of a given county tend to be spatially correlated with the yields of nearby counties because weather, geological features, and other hidden features that

could potentially affect crop yields in one county are usually similar in neighboring counties (Annan, Tack, and Harri 2014; Du et al. 2015). Considering this spatial correlation can potentially lead to more precise yield distribution estimates and thus more accurately estimated premium rates. Although the method suggested by Harri et al. (2011) estimates parameters at the county level with spatial restrictions at the district level, the method uses only the historical yield data from the county whose density is being estimated. Of considerable interest is accounting for spatial correlation to improve crop insurance rating. Ozaki et al. (2008) introduce a crop insurance rating method for Brazilian data based on a Bayesian hierarchical approach that models temporal and spatial autocorrelation of crop yield. They assume normality of crop yield and use a random effect to control spatial correlation of the crop yield. However, their spatio-temporal model only reflects the spatial effects from neighboring counties and their approach only considers spatial effects on the mean whereas our approach considers spatial smoothing in all parameters. Another approach to spatial smoothing is based on Bayesian Model Averaging (BMA). Woodard (2016) uses BMA to get a weighted average of county and district level parameters. Ker, Tolhurst, and Liu (2015) estimate a posterior density for each county using only its own data. The posterior density of the target county is then a weighted average of its own posterior density and densities from other counties. This method does not require any knowledge in density similarity and can be applied with both parametric and non-parametric estimators. The work using BMA, although very different[2] from the work presented here, indicates the interest in using spatial information to model crop yield distributions and crop insurance premium rates.

A familiar issue in the estimation of crop yield distributions is that the number of time-series observations is very limited. Particularly, estimating the tail of the distribution is difficult since rare events usually have limited observations. One potential solution is to include observations from the other counties. This approach in turn presents the challenge of directly considering spatial correlation as part of the estimation method. We address the problem by using a Bayesian hierarchical structure. Unlike BMA, our Bayesian Kriging method directly incorporates the spatial smoothing procedure into the model. Therefore, posterior densities are estimated jointly using data for all counties[3]. We estimate parameters of the yield distribution for each county with a specified functional form of the spatial covariance matrix in the process layer of the hierarchical structure.

Our research develops a method for determining more accurate area yield insurance premium rates based on extreme value theory and Bayesian Kriging. The contribution of our research to the literature on the estimation of crop yield distributions and crop insurance premium rates is the introduction of two innovations. One is to use a Bayesian Kriging method for spatial smoothing to estimate crop yield distributions across counties. To our knowledge, our research is the first to use such a method for spatial smoothing to develop a crop insurance rating model. The second innovation is to only estimate the tail of the distribution since that is all that matters in rating crop insurance. The tails of crop yield distributions are assumed to follow a Generalized Pareto Distribution (GPD)[4], which is a member of the extreme value distribution family. GPD models only the tail of the distribution that generates indemnity payments. County-level historical wheat yields (1970-2014) from NASS are used for the estimation. We assume that GPD parameters (scale and shape) of the counties are spatially correlated and the

spatial processes for the parameters are structured under a Bayesian hierarchical framework. Metropolis-Hastings (MH) steps within a Gibbs sampler are used to update posterior densities. Maximum likelihood estimates are used as candidates for the MH step to increase acceptance rate and speed convergence in the Markov Chain Monte Carlo (MCMC) procedure. Our model produces spatially smoothed GPD parameters within the MCMC procedure. We verify spatial correlation of GPD parameters from the posterior density of Kriging parameters, and we fit the Gaussian spatial process to visualize and verify the structure of spatial correlation. A main finding is that estimated premiums from the RMA model[5] result in considerable regional inequalities compared to the method proposed here. The RMA model more poorly rates the premiums when measuring deeper tail probability. We compare out of sample performance of the two approaches by assuming a representative insurance company that chooses whether to retain or cede polices under the Standard Reinsurance Agreement (SRA). We verify statistical significance of performance through bootstrapping. The new model outperforms the RMA model and mitigates regional inequalities in loss ratios.

Generalized Pareto Distribution

Extreme value theory is used for analysis of low probability events. The theory states that the tail of a loss distribution can be approximated by a GPD. The extreme value family can be written in a simple form involving three parameters: location (ς), scale ($\sigma > 0$), and shape (ξ). A positive shape parameter ($\xi > 0, \text{Fréchet type}$) implies a heavy tail distribution, a negative shape parameter ($\xi < 0, \text{Weibull type}$) implies a bounded tail, and a zero-value shape parameter ($\xi = 0, \text{Gumbel type}$) implies a light tail distribution.

Suppose X is a random variable. Now consider the conditional distribution of X given that it exceeds u. If F is a cumulative density function (CDF) for X, the probability of X exceeding x, given X is greater than the threshold u, can be approximated by the GPD[6]:

$$(1) \qquad \mathrm{P}(X > x | X > u) = \frac{1 - F(x)}{1 - F(u)} = \begin{cases} 1 - (1 + \frac{\xi(x-u)}{\sigma})^{-1/\xi} & \xi \neq 0 \\ 1 - \exp^{-(x-u)/\sigma} & \xi = 0 \end{cases}$$

where scale parameter $\sigma > 0$, and shape parameter $-\infty < \xi < \infty$.

The GPD only uses observations above (or below) a threshold and fits a model to those tail data only. GPD allows three basic forms of the tails. As with Generalized Extreme Value (GEV) distribution, a positive shape parameter ($\xi > 0, \text{Fréchet type}$), implies a heavy tail (i.e. Student's *t* distribution), a negative shape parameter ($\xi < 0, \text{Weibull type}$) implies a bounded tail (i.e. beta distribution), and a zero-value shape parameter ($\xi = 0, \ \text{Gumbel type}$) implies a light tail distribution (i.e. exponential distribution).

Another advantage of using a GPD for the tail of the distribution is that the quantiles have a closed form. Using equation (1) above, the quantiles ($X_q = F^{-1}(q)$) can be defined as

$$(2) \qquad X_q = \begin{cases} u + \frac{\sigma}{\xi}\left[(\zeta_u/q)^{\xi} - 1\right] & \xi \neq 0 \\ u + \sigma \ln(\frac{\zeta_u}{q}) & \xi = 0 \end{cases}$$

where $\zeta_u = 1 - F(u)$ denotes the probability of exceeding the threshold u. Therefore, once we obtain the parameters of the distribution, a probability that a variable exceeds some specific threshold can be directly computed. We must choose a threshold level u in order to fit the GPD. The threshold selection is one difficulty in fitting the GPD and thus

finding an optimal approach to select the threshold is still an active research area. One may think that it is natural to obtain the threshold u by maximum likelihood along with the other parameters. However, this approach will not be stable because the number of observations changes as u is changed. This fact will lead to a discontinuous or unbounded likelihood function. Consistent with the way RMA sets thresholds of current policies, we set the coverage level of each county i as the threshold to estimate the premium rates for different coverage levels[7]. Therefore, each county's threshold level u_i is the ith county's predicted yield $\hat{y}_i$ multiplied by coverage rate λ (i.e. 70 percent or 90 percent) of the insurance, where $u_i = \lambda\hat{y}_i$. Then the observations of each county i that are below the threshold for the county u_i, are used to fit the GPD parameters.

Bayesian Modeling Framework

Our spatial smoothing procedure is a Kriging approach. Kriging is a geostatistical spatial interpolation method. With our approach, the interpolated measures are obtained by estimating the parameters of a Gaussian spatial process with an explicit functional form of spatial covariance. The spatial correlation varies by distance as well as other measures of similarity. The GPD parameters for one county are based on a weighted average of its own data and the data from other counties, but the weights decrease with distance. Kriging based on Bayesian inference has been actively developed in a large variety of disciplines due to its advantage to incorporate model uncertainty associated with the Gaussian spatial process. Recent statistical literature (Cooley, Nychka, and Naveau 2007, Gelman et al. 2004, Woodworth 2004, Ntzoufras 2011) offers various adjustments of the

Bayesian Kriging method, such as spatial smoothing in extreme weather events, disease incidence, and mortality rates prediction.

Our Bayesian Kriging model has a Bayesian hierarchical structure. The Bayesian hierarchical structure is defined when a prior distribution of the general Bayesian model (priors for the parameters of the likelihood function, i.e. GPD parameters σ and ξ) can also be assigned to additional prior parameters, say hyper parameters. Our model assumes that the parameters of the likelihood function are spatially correlated. Hence, the model incorporates spatial smoothing that reflects such spatial correlation by adding one more prior layer (process layer) between the likelihood density and the prior density from the basic Bayesian model. Therefore, the values of likelihood function parameters (GPD parameters) of each county are assumed to be affected by additional covariates that characterize each county (i.e., historical yield level, longitude, latitude) and Kriging parameters for spatial smoothing.

Hierarchical structure

The model has three hierarchical layers. First, in the likelihood layer, the tail of crop yield distribution of each county follows a GPD. Second, the process layer models the spatial process of the GPD parameters. This second layer determines the parameters of the GPD. The parameters of each county are determined by covariates that reflect the spatial characteristics of the county. The spatial smoothing of the parameters is performed assuming a Gaussian spatial process with an explicit functional form of spatial covariance matrix. The third layer consists of the prior distributions for the hyper parameters, called

hyper priors, for the covariates of the process layer and Kriging parameters (sill and range) in the spatial covariance matrix. The three hierarchical layers can be specified as,

$$\begin{aligned} &\text{Likelihood layer: } \boldsymbol{Y}|\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2} \sim p_1(\boldsymbol{Y}|\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2}) \\ &\text{Process layer: } \boldsymbol{\Omega_1}|\boldsymbol{\Omega_2} \sim p_2(\boldsymbol{\Omega_1}|\boldsymbol{\Omega_2}) \\ &\text{Prior layer : } \boldsymbol{\Omega_2} \sim p_3(\boldsymbol{\Omega_2}) \end{aligned} \tag{3}$$

where p_j is the density associated with each layer of the hierarchies, $\boldsymbol{Y}$ is a matrix of yearly crop yields that spans all counties ($i = 1, \dots, N$) and years ($t = 1, \dots, T$), $\boldsymbol{\Omega_1}$ is a matrix of the GPD parameters that spans all counties ($i = 1, \dots, N$) so that $\boldsymbol{\Omega_1} = [\boldsymbol{\sigma}, \boldsymbol{\xi}]$, and $\boldsymbol{\Omega_2}$ is a vector of hyper parameters (coefficients for covariates and Kriging parameters), $\boldsymbol{\Omega}_2 = [\beta_0, \beta_h, \delta, \theta_\sigma, \rho_\sigma, \theta_\xi, \rho_\xi]'$.

When we factorize the conditional density of the likelihood layer $p_1(\boldsymbol{Y}|\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2})$ using Bayes' theorem, the prior distribution for the likelihood layer density, say $p(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2})$, can be separated into two components:

$$p(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2}) = p_2(\boldsymbol{\Omega_1}|\boldsymbol{\Omega_2}) p_3(\boldsymbol{\Omega_2}) \tag{4}$$

with the conditional distribution of GPD parameters given hyper parameters $p_2(\boldsymbol{\Omega_1}|\boldsymbol{\Omega_2})$ and prior distribution of hyper parameters $p_3(\boldsymbol{\Omega_2})$.

Then the joint posterior distribution of the Bayesian hierarchical model is

(5) $$p(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2} | \boldsymbol{Y}) = \frac{p(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2}, \boldsymbol{Y})}{p(\boldsymbol{Y})} = \frac{p_1(\boldsymbol{Y}|\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2}) p(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2})}{\int_{\boldsymbol{\Omega_1}} \int_{\boldsymbol{\Omega_2}} p_1(\boldsymbol{Y}|\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2}) p(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2}) d\boldsymbol{\Omega_2} d\boldsymbol{\Omega_1}},$$

and thus the joint posterior distribution $p(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2} | \boldsymbol{Y})$ is proportional to the multiplication of the likelihood $p_1(\boldsymbol{Y}|\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2})$ and $p(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2})$,

(6) $$p(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2} | \boldsymbol{Y}) \propto p_1(\boldsymbol{Y}|\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2}) p(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2}).$$

Next, plug equation (4) into the right-hand side of equation (6), which gives

(7) $$p(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2} | \boldsymbol{Y}) \propto p_1(\boldsymbol{Y}|\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2}) p_2(\boldsymbol{\Omega_1}|\boldsymbol{\Omega_2}) p_3(\boldsymbol{\Omega_2}).$$

Therefore, the joint posterior density of the model $p(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2} | \boldsymbol{Y})$, which is our target density, is proportional to the multiplication of the three layers of the model. The full conditional posterior density for each parameter can be fitted using MH within a Gibbs sampling algorithm.

Likelihood layer

A GPD likelihood function forms the first hierarchy of our model. Let z_{it} denote the yearly crop yield at county $i = 1, \dots, N$ at year $t = 1, \dots, T$. Since our interest is in minima of crop yield rather than maxima, we use the negative of the yield observations, where $y_{it} = -z_{it}$ for any i and t. Given that y_{it} falls below the threshold u_i for county i,

we assume that the tail of the crop yield of each county can be fit by a GPD whose parameters depend on the county's location.

Let $\phi_i = \log \sigma_i$ and ξ_i represent the log-transformed scale and shape parameters of county i. The log-transformation allows the parameter ϕ to take on both positive and negative values. Differentiating the cumulative distribution function (1) gives the probability density function (or likelihood) of the likelihood layer,

$$(8) \qquad p_1(\boldsymbol{Y}|\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2}) = \prod_{i=1}^{N} \prod_{t=1}^{T} \left[\frac{1}{\exp \phi_i} \left[1 + \frac{\xi_i y_{it}}{\exp \phi_i} \right]^{-\frac{1}{\xi_i}-1} \right]^{I(y_{it}<u_i)}$$

where the indicator function, $I(y_{it} < u_i) = \begin{cases} 1 & \text{if } y_{it} < u_i \\ 0 & \text{otherwise} \end{cases}$, denotes the yield data of county i that fall below the threshold u_i, $\boldsymbol{\Omega}_1 = [\boldsymbol{\phi},\ \boldsymbol{\xi}]$ and $\boldsymbol{\Omega}_2 = [\beta_0, \beta_h, \delta, \theta_\sigma, \rho_\sigma, \theta_\xi, \rho_\xi]'$, which are the matrix of GPD parameters and vector of hyper-parameters.

Process layer

The key part of the hierarchical structure is the process layer since it contains the spatial smoothing parameters. The Kriging of the GPD parameters (ϕ_i and ξ_i) is based on a Gaussian spatial process[8]. The Kriging parameters (sill and range) determine the degree and distance of spatial correlation when performing the spatial smoothing. Since we are in a Bayesian framework, the GPD parameters $\phi_i$and ξ_i are random variables. The GPD parameters are then determined by other parameters ($\beta_0, \beta_h, \delta, \theta_\sigma, \rho_\sigma, \theta_\xi, \rho_\xi$) and the priors on these parameters are called hyper priors. The hyper priors are assumed to be

independent and the spatial processes for each GPD parameter are assumed independent. Thus, the model for the log-transformed scale parameters $\phi_i = \log \sigma_i$ is

$$
\begin{aligned}
\phi_i &= \mu_i + \varepsilon_i, \\
\mu_i &= \beta_0 + \sum_{h=1}^{H} \beta_h Z_{hi}, \\
\boldsymbol{\phi} &\sim MVGP(\boldsymbol{\mu}, \Sigma_\phi), \\
\Sigma_\phi &= \psi(D_{ij}; \theta_\phi, \rho_\phi), \\
\varepsilon_i &\sim MVN(0, \Lambda),
\end{aligned}
\tag{9}
$$

where μ_i is the deterministic part of the Gaussian spatial process, Z_{hi} contains covariates of the process at county i, $\boldsymbol{\phi}$ is the vector of log-transformed scale parameters, $\boldsymbol{\phi} = [\phi_1, \ldots, \phi_N]'$, and is assumed to follow a multivariate Gaussian spatial process with spatial covariance matrix, $\Sigma_\phi = \psi(D_{ij}; \theta_\phi, \rho_\phi)$[9], that is a function of Euclidean distance (D_{ij}) between counties i and j, sill parameter ρ_ϕ, and range parameter θ_ϕ, and ε_i is a non-spatial error component that follows $\varepsilon_i \sim MVN(0, \Lambda)$, where Λ is a diagonal matrix with diagonal elements of υ^2 and all other elements are zero. The MCMC process generates Bayesian Kriging for ϕ_i from the Gaussian spatial process and the effect of spatial correlation is smoothly interpolated by the spatial covariance matrix form $\psi(D_{ij}; \theta_\phi, \rho_\phi)$[10].

Similarly, the latent process of the shape parameters of GPD, ξ_i, is

$$\xi_i = \delta + \epsilon_i$$

$$\boldsymbol{\xi} \sim MVGP(\boldsymbol{\delta}, \Sigma_\xi),$$

(10)

$$\Sigma_\xi = \varphi(D_{ij}; \theta_\xi, \rho_\xi),$$

$$\epsilon_i \sim MVN(0, \Psi),$$

where δ is a constant, $\boldsymbol{\xi}$ is a vector of shape parameters $\boldsymbol{\xi} = [\xi_1, \dots, \xi_N]'$, and is assumed to follow a multivariate Gaussian process with the explicitly assumed spatial covariance matrix, $\Sigma_\xi = \varphi(D_{ij}; \theta_\xi, \rho_\xi)$, that is a function of Euclidean distance(D_{ij}) between counties i and j, sill parameters ρ_ξ, and range parameter θ_ξ, and ϵ_i non-spatial error component that follows $\epsilon_i \sim MVN(0, \Psi)$ where Ψ is a diagonal matrix with diagonal elements of ν^2.

From the two spatial processes of GPD parameters ϕ and ξ, the vector of log-transformed scale parameter $\boldsymbol{\phi}$ and shape parameter $\boldsymbol{\xi}$ given $\beta_0, \beta_h, \rho_\phi, \theta_\phi, \delta, \rho_\xi$, and θ_ξ follows

$$\boldsymbol{\phi} \mid \beta_0, \beta_h, \rho_\phi, \theta_\phi \sim MVGP(\boldsymbol{\mu}, \Sigma_\phi)$$

(11)

$$\boldsymbol{\xi} \mid \delta, \rho_\xi, \theta_\xi \sim MVGP(\boldsymbol{\delta}, \Sigma_\xi)$$

where $\boldsymbol{\mu} = \boldsymbol{\beta}_0 + \sum_{h=1}^{H} \boldsymbol{\beta}_h \boldsymbol{Z}_h$, $\Sigma_\phi = \psi(D_{ij}; \theta_\phi, \rho_\phi)$ and $\Sigma_\xi = \varphi(D_{ij}; \theta_\xi, \rho_\xi)$.

Since we assume the spatial processes for each GPD parameter are independent, the second part of equation (7) is

$$p_2(\boldsymbol{\Omega_1}|\boldsymbol{\Omega_2}) = \frac{1}{\sqrt{(2\pi)^N|\Sigma_\phi|}}\exp\left[-\frac{1}{2}(\boldsymbol{\phi}-\boldsymbol{\mu})'{\Sigma_\phi}^{-1}(\boldsymbol{\phi}-\boldsymbol{\mu})\right] \times \frac{1}{\sqrt{(2\pi)^N|\Sigma_\xi|}}\exp\left[-\frac{1}{2}(\boldsymbol{\xi}-\boldsymbol{\delta})'{\Sigma_\xi}^{-1}(\boldsymbol{\xi}-\boldsymbol{\delta})\right] \quad (12)$$

where $\boldsymbol{\mu}$ is a vector of mean values of ϕ_i defined by equation (9), $\boldsymbol{\delta}$ is a vector of mean values of ξ_i defined by equation (10), Σ_ϕ and Σ_ξ are the covariance matrices for the GPD parameters ϕ_i and ξ_i, $\boldsymbol{\Omega_1}$ is the matrix of the GPD parameters, $\boldsymbol{\Omega_1} = [\boldsymbol{\phi}, \boldsymbol{\xi}]$, for all counties $(i = 1, \dots, N)$, and $\boldsymbol{\Omega_2}$ is a vector of the hyper parameters, $\boldsymbol{\Omega_2} = [\beta_0, \beta_h, \delta, \theta_\phi, \rho_\phi, \theta_\xi, \rho_\xi]'$.

To develop crop insurance rating models for crops with a temporal trend (i.e., technological progress), we would need to include a trend variable in the process layer of the Bayesian hierarchical structure. However, including a trend variable brings substantial technical challenges under the non-normality assumption for the likelihood layer[11]. One simple way to avoid the complexity in modeling is to estimate the trend of the crop yield outside the hierarchical procedure in a manner similar to Harri et al. (2011). Since the wheat yields have no significant trend, we do not explore the issue of trend here. But, if the model were extended to corn or cotton that have strong yield trends, the issue of trend would need to be addressed.

Prior layer

In the prior layer, we specify a prior (i.e., hyper-prior) for the vector of hyper-parameters $\boldsymbol{\Omega_2}$, which characterizes the GPD parameters in the process layer. Each parameter in the

layer is assumed independent of the others. Since we do not have any prior information about a relationship between GPD parameters and covariates in the process layer, we impose large prior ranges for the covariate parameters β_0 and β_h. For all the models, the prior for all covariates parameters β_0 and β_h is $Uniform(-10000, 10000)$. Setting the priors of the Kriging parameters sill (ρ_ϕ, ρ_ξ) and range $(\theta_\phi, \theta_\xi)$, which describe the spatial structure of the scale and shape parameter of the GPD is more difficult. Bayesian statistics literature (Berger, DeOivelira, and Sanso 2001; Banerjee, Carlin, and Gelfand 2004; Cooley, Nychka, and Naveau 2007) points out that improper priors for the Kriging parameters may result in improper posterior distributions. Banerjee, Carlin, and Gelfand (2004) suggest choosing informative priors for Kriging parameters as the safest way to avoid improper posteriors. Therefore, we use empirical information to construct proper priors for the Kriging parameters based on the maximum likelihood estimation. We explain the choice of prior distributions of the sill parameters (ρ_ϕ, ρ_ξ). We first estimate scale and shape parameters for each county independently using maximum likelihood. A histogram of obtained scale $(\phi_{i,MLE})$ and shape $(\xi_{i,MLE})$ parameters for each county is illustrated in Figure 1. We then fit an empirical variogram[12], $\hat{\gamma}(D_{ij})$, suggested by Cressie (1993), using the obtained MLE parameters for each county to find the proper range of prior distribution of the sill parameter. We finally impose a prior of $IG(0.01, 5)$ for ρ_ϕ and $IG(0.01, 2)$ for ρ_ξ. To find the prior distributions for range parameters $(\theta_\phi, \theta_\xi)$, we use prior geographical knowledge of empirical data. Since the spatial effect is measured using Euclidean distance based on longitude / latitude coordinate space, the nearest and farthest distance between the pair of locations in the empirical dataset are used for the prior distribution of the range parameter. Therefore, we impose a prior of

$gamma(\ 7{,}71, 0.23)$ for the range parameters both in scale and shape parameters (θ_ϕ and θ_ξ). With the priors as above, the third layer in equation (7) is

$$p_3(\boldsymbol{\Omega_2}) = p(\beta_0)p(\beta_h)p(\delta)p(\theta_\phi)p(\theta_\xi)p(\rho_\phi)p(\rho_\xi). \tag{13}$$

Since $p_1(\boldsymbol{Y}|\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2})$, $p_2(\boldsymbol{\Omega_1}|\boldsymbol{\Omega_2})$, and $p_3(\boldsymbol{\Omega_2})$ in equation (7) are obtained from the three layers in equations (8), (12), and (13), we now have posterior distributions for the parameters by multiplying these three layers. The overall schematic of the model structure to derive the distribution of crop yields is illustrated in Figure 2.

Markov Chain Monte Carlo Procedure

The estimation uses Metropolis-Hastings (MH) steps within a Gibbs sampler. In the MH algorithm, random parameter values are drawn from a candidate density and then accepted or rejected with the accepted values included in the posterior density. Similar to Cooley, Nychka, and Naveau (2007), we use the conditional density obtained from maximum likelihood estimates of the GPD parameters as the candidate density for the posterior distribution. Let $\widehat{\boldsymbol{\phi}}$ be a vector of MLE estimates for the vector of GPD scale parameters $\boldsymbol{\phi}$. From the asymptotic property of MLE's, we have

$$\sqrt{T}(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \xrightarrow{d} MVN\left(0, \lim_{T\to\infty}\left[\frac{1}{T}I(\boldsymbol{\phi})\right]^{-1}\right) \tag{14}$$

where $I(\boldsymbol{\phi})$ is the information matrix of MLE estimates. Given the process layer density in equation (12), we obtain the joint distribution of $\widehat{\boldsymbol{\phi}}$ and $\boldsymbol{\phi}$ as

$$\begin{pmatrix} \widehat{\boldsymbol{\phi}} \\ \boldsymbol{\phi} \end{pmatrix} = MVN\left(\begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \end{pmatrix}, \begin{bmatrix} \Sigma_\phi + I(\boldsymbol{\phi})^{-1} & \Sigma_\phi \\ \Sigma_\phi & \Sigma_\phi \end{bmatrix} \right). \tag{15}$$

We then construct the conditional distribution

$$\boldsymbol{\phi} \mid \widehat{\boldsymbol{\phi}} \sim MVN\left(\boldsymbol{\mu} + \Sigma_\phi\left(I(\boldsymbol{\phi})^{-1} + \Sigma_\phi\right)^{-1}\left(\widehat{\boldsymbol{\phi}} - \boldsymbol{\mu}\right),\ \ \Sigma_\phi - \Sigma_\phi{}'\left(I(\boldsymbol{\phi})^{-1} + \Sigma_\phi\right)^{-1}\Sigma_\phi\right) \tag{16}$$

that is used as the candidate density in the MH step. By the central limit theorem, MLE estimates will be close to the Bayesian posterior, and thus the sampling distribution of the MLE should be a good candidate distribution for this part of the posterior. Cooley, Nychka, and Naveau (2007) find this approach significantly increases the acceptance rate of MH steps. After updating the GPD parameters, Gibbs sampling is further used to update the other parameters. A variety of R-packages provide MCMC updating algorithms for Bayesian hierarchical structure. We mainly employ SpatialExtremes, extRemes, and spBayes packages to construct the estimation procedure. These packages provide algorithms for Bayesian spatial smoothing with a normal distribution and a Generalized Extreme Value (GEV) distribution. However, these packages do not provide a function to estimate the Bayesian spatial smoothing model with GPD. Therefore, we combine functions in these packages and modify the likelihood and process layer

procedures of the functions and construct a new R function for our proposed model with GPD.

Model Selection

We model county-level winter wheat yield data from National Agricultural Statistics Service (NASS). The data contain 1970-2014 annual yields (bushels per acre) for 77 counties in Oklahoma. Counties with missing observations are discarded. Therefore, 39 counties' yields are included in the final dataset. A single state is considered because RMA has generally wanted to tell producers that no data from another state affects their premium. Also, the algorithm is sufficiently slow enough that estimating the model for the whole United States at once is impractical with current computer resources. Since RMA would estimate separate models for each state, it would be practical for RMA to use the method proposed here. The box plot of the historical county level yields is presented in Figure 3.

Deviance Information Criterion (DIC) suggested by Spiegelhalter et al. (2002) is used to evaluate the goodness of fit of each model. DIC has substantial advantages for Gaussian likelihoods and is particularly convenient to compute from posterior samples (Finley, Banerjee, and Carlin 2007). This criterion is the sum of the Bayesian deviance and the effective number of parameters (a penalty for model complexity). The deviance is the negative of twice the log-likelihood,

(17) $$D(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2}) = -2\log p_1(\boldsymbol{Y}|\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2})$$

where the $p_1(\boldsymbol{Y}|\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2})$ is the likelihood function from the likelihood layer. The Bayesian deviance is the posterior mean, $\overline{D} = E[D(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2})]$, and an effective number of parameters is given by $P_D = \overline{D}(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2}) - D(\overline{\boldsymbol{\Omega}}_1, \overline{\boldsymbol{\Omega}}_2)$, where $\overline{\boldsymbol{\Omega}}_1$ and $\overline{\boldsymbol{\Omega}}_2$ are the posterior means of the model parameters. Therefore, DIC is given by, $DIC = \overline{D} + P_D$. Lower DIC values indicate preferred models.

We use the DIC to select from alternative covariates specifications and two different spatial covariance functional forms (exponential and Matern), providing nine alternative models. In the process, we draw 100,000 iterations for MCMC chains and burn-in the first 20,000 observations to get the posterior distribution of each parameter. We fit both trace plots and autocorrelation plots for all posterior densities. All densities show sufficiently fast convergence and no significant autocorrelation. Table 1 presents the models tested and their corresponding DIC values. We start with a model with no spatial smoothing (Model 0), and expand the model with several covariates in the process layer of the log-transformed scale parameter (ϕ_i), including historical average yield ($\bar{y}_i$), longitude (lon_i), and latitude (lat_i) of the county. Model 5 with a constant coefficient and exponential covariance function dominates other models with multiple covariates under DIC. Therefore, adding other covariates does not improve the model. Model 5 is selected as the main model. Two examples of estimated posterior densities of the GPD parameters are presented in Figure 4. Figure 4 illustrates posterior densities of two counties, Beaver and Oklahoma, which have the highest and lowest premium rates in 2014. Both counties have negative posterior values of shape parameters and have similar ranges of the posterior values from -0.58 to -0.01. All counties, including Beaver and Oklahoma, have similar ranges of the posterior values from -0.59 to -0.01 and the

negative shape posterior values indicating that the tail distribution of crop yields is a Weibull type, which implies a bounded tail distribution (not unexpected since yields cannot be negative). Posterior values of the scale parameters differ substantially between the two counties. The scale parameter of GPD indicates the thickness of the tail of the distribution. The scale parameter of Beaver County is greater than that of Oklahoma County, which indicates the tail of Beaver County's density is thicker than the tail of Oklahoma County's density.

One of the most important advantages of our model is that it provides estimates of the spatial correlation structure. Figure 5 illustrates posterior densities of Kriging parameters. The posterior values of the Kriging parameters are used to identify the structure of spatial correlation among GPD parameters of the counties. To illustrate the spatial correlation structure, we use an F-Madogram (Cooley, Naveau, and Poncet, 2006). Generally, when some observations y_i (in our case, GPD parameters ϕ_i and ξ_i) follow a Gaussian spatial process with a spatial covariance function ψ, a variogram (γ_{ij}) suggested by Cressi (1993) is used to illustrate the structure of spatial correlation. A variogram is formed as $\gamma_{ij} = \sigma^2\{1 - \psi(D_{ij};\ \theta, \rho)\}$, where γ_{ij} is the variogram between regions i and j, $y_i$and y_j are observations in location i and j, σ^2 is the variance of the Gaussian spatial process, D_{ij} is the Euclidean distance between two locations, and θ and ρ are Kriging parameters (range and sill). Therefore, the variogram shows the variance of the difference between the observations (i.e., GPD parameters) at two different locations (i.e., counties). Since a Gaussian spatial process is stationary and isotropic, the spatial correlation between two locations, does not depend on the actual coordinates of the locations, but it depends only on the Euclidean distance between two locations.

However, if our interest is in extreme values, the variogram cannot be a useful tool. Cooley, Naveau, and Poncet (2006) propose a modified madogram called the F-madogram $v_F(D_{ij})$, which is a useful quantity to evaluate the structure of spatial correlation in a Gaussian random process with an extreme value distribution. Like the variogram, the F-madogram is interpreted as the degree of the difference between the parameters of two different counties. We plot the F-Madogram for the scale and shape parameters from the posterior values of Kriging parameters (sill and range). Figure 6 illustrates the F-Madograms for the scale and shape parameter. A zero F-Madogram $v_F(D_{ij})$ indicates perfect correlation, and a larger value of $v_F(D_{ij})$ indicates a smaller spatial correlation. The sill parameter ρ involves the maximum limit of F-Madogram, which represents the maximum difference between the parameters in the different counties, and range parameter θ involves the distance at which the F-Madogram flattens out, which represents the distance that spatial correlation still remains. X-axis h in Figure 6 is the Euclidean distance between two locations (km), and Y-axis $v_F(D_{ij})$ represents the F-madogram. Therefore, an effect of spatial correlation on the parameters (scale and shape) at different locations remains at approximately 200km and the limit of the F-Madogram value is approximately 0.17, indicating the maximum difference of the parameters (scale and shape) obtained[13]. We use R-package 'SpatialExtremes' to obtain the F-Madogram of GPD parameters.

Out of sample comparison

The next step is calculating premium rates using the selected model and evaluating out of sample performance. We evaluate our model compared to the model suggested by Harri

et al. (2011). The first step is to calculate the premium rates of each county. The premium rate for county i, $prem_i$, for the area yield insurance contract that guarantees the coverage level, say $\lambda\hat{y}_i$, (Ker and Coble, 2003) is

$$prem_i = \frac{P(y_i < \lambda\hat{y}_i)(\lambda\hat{y}_i - E[(y_i | y_i < \lambda\hat{y}_i)])}{\lambda\hat{y}_i} \tag{18}$$

where λ is the coverage rate, $\hat{y}_i$ is the predicted yield in county i, and the expectation and probability measures are taken using the conditional predictive posterior distribution.

Premium rates are calculated for both 70 and 90 percent coverage levels and compared to rates with the current rating method suggested by Harri et al. (2011). As noted, we refer to the Harri et al. (2011) model as the RMA model. Figure 7 and 8 illustrate interpolation of 70 and 90 percent premium rates for 2014 from each model onto longitude-latitude space, respectively. The left figure has premium rates from the proposed model, and the right figure has premium rates from the RMA model. The figure demonstrates that the premium rates from the new model show a smoother surface than that of the rates from the RMA model. Further, we calculate the out-of-sample loss ratio for each model to compare the performance of the two models. The loss ratio is given by

$$lossratio_i = \frac{\sum_{t=1}^{T} \max[\lambda\hat{y}_{it} - y_{it},\ 0]}{\sum_{t=1}^{T} prem_{it}\hat{y}_{it}} \tag{19}$$

where λ is coverage rate, $\hat{y}_{it}$ is the predicted yield of county i at year t, y_{it} is actual yield for county i at year t, and $prem_{it}$ is premium rate of county i at year t, which is obtained from equation (18).

The premium gains (denominator) and indemnity losses (numerator) of equation (19), under the proposed and RMA (Harri et al, 2011) model, are obtained using actual yields of each county from 2000 to 2014. Average, variance, maximum, and minimum loss ratio across counties for each model are presented in Table 2. The loss ratio of fairly rated crop insurance should equal one. The average loss ratio across the counties under 70 and 90 percent coverage rates of the new model are 2.10 and 1.19, respectively. Both of the averages from the new model are closer to one than the RMA model (2.66 and 1.43), which indicates the premiums from the new model are more fairly rated than the RMA model. The fourth column in Table 2 shows that the new model has a smaller variation of loss ratio across counties both in 70 and 90 percent coverage rates. The reduction in the variance of the loss ratio (8.14 to 2.40) is greater under the 70 percent coverage rate. Specifically, for counties with a high loss ratio under the RMA model, such as Alfalfa and Pottawatomie counties, the loss ratio becomes substantially closer to one under the new model. More details of county level loss ratios in Oklahoma from the RMA model and the new model are presented in Table 3. Figures 9 and 10 project the county level loss ratio under the 70 and 90 percent coverage rates from the new model and the RMA model onto longitude and latitude space. Both in 70 and 90 percent coverage rates, the RMA model results in large loss ratios in east central Oklahoma, whereas the new model has a relatively equally distributed loss ratio. The new model offers more flexible shapes of the tail of the density via GPD and produces spatially

smoothed parameters of the tail density. The new model, in particular, clearly reveals its accuracy in the premium calculation under the 70 percent coverage rate. The RMA model fits relatively well in the 90 percent coverage rate, but it tends to increasingly inaccurately measure the premiums as the coverage rate moves close to the tails.

Next, similar to Harri et al. (2011), we assume a representative insurance company that can choose whether to retain or cede policies to the Federal Crop Insurance Corporation (FCIC) under the Standard Reinsurance Agreement (Coble, Dismukes, and Glauber 2007). The company is assumed to estimate premium rates using the proposed model and compare them with RMA rates. If the rates are higher than the RMA, then policies are ceded to the FCIC since the RMA rates are underestimated, and therefore we expect a loss. Whereas, if the rates are smaller than the RMA, policies are retained since the RMA rates are overestimated, and therefore a profit is expected from retaining the policy. We repeat the out-of-sample procedure over the fifteen years from 2000 to 2014 and calculate the loss ratio of the retained and ceded policies. We then conduct a statistical test under the null hypothesis that the loss ratios of the two types of policies are identical. The ceded to retained ratio is calculated by dividing the loss ratios of ceded and retained policies. Under the null hypothesis, the ceded to retained ratio should equal one. A non-parametric bootstrap is used to calculate statistical significance[14]. The "*p*-value" in Table 4 is the type 1 error estimated from the bootstrapping method under the null hypothesis that the loss ratios for the ceded and retained policies are identical. The results in Table 4 corroborate that the incorrect premium rate calculation may result in a significant economic loss under the Standard Reinsurance Agreement (SRA). The ceded to retained ratios are 4.20 in 70 percent and 2.18 in 90 percent coverage rate, which

implies that the loss ratio of ceded policies is much higher than that of retained policies. The *p*-value confirms the statistical significance.

Conclusions

We provide a new approach to rate area-yield crop insurance. Two innovations are introduced. One is to use a Bayesian Kriging method for spatial smoothing of all parameters to estimate crop yield distributions across counties. The model uses spatially smoothed parameters of the yield distribution. The second innovation is to only estimate the tail of the distribution (using GPD) since that is all that matters in rating crop insurance. GPD allows flexible forms of the tail of the distribution and therefore more accurately measures the tail probability.

Our Bayesian Kriging model has notable advantages over other types of spatial smoothing methods. The model produces spatially smoothed parameters of the crop yield density within the MCMC procedure and thus does not require additional steps for spatial smoothing. The weights of the spatial smoothing are measured by Euclidean distance between the counties and the Kriging parameters. In this context, the model is less ad hoc than other types of the spatial smoothing methods when imposing weights or selecting a boundary for the spatial effect. Further, the new model can illustrate the spatial correlation structure from the posterior densities of the Kriging parameters. This can provide important information about the spatial correlation. Our results demonstrate some important implications for crop insurance rating. Spatial correlations of crop yield tail distribution (GPD) parameters remain up to a distance of approximately 200 km. This result may be used for future studies as a reference when setting the proper range of the

effect in considering spatial correlation. In addition, our results verify the advantage of using GPD in measuring tail probability. The flexible tail shape of GPD contributes to considerable improvement in insurance rating, and the improvement is greater when measuring deeper tail probability.

Several model specifications are examined to identify the quality of the candidate models, and the performance of the selected new model is compared to the performance of the RMA model. The loss ratios of the two models indicate that policies are more fairly rated using the new model than the RMA model. Also, the new model mitigates regional inequalities in loss ratios. In fact, premium rates from Bayesian Kriging could lead to more precise risk measurement and therefore significantly reduce inequalities in loss ratios across counties.

An important extension of this research is to develop a comprehensive model for other crops, including crops with a temporal trend (i.e. corn or cotton). Adding a trend variable into the Bayesian hierarchical structure brings a substantial increase in the scope of our work, and of course, this extensive modeling will lead to a significant increase in computational complexity. While our model does not directly include the temporal trend in our model structure, we can estimate the trend of crop yields outside of the model procedure using several types of trend estimation methods similar to Harri et al. (2011). Another way to incorporate trend would be to assume normality instead of GPD. One possible way to increase computational speed is with Hamiltonian Monte Carlo (HMC). HMC can be used to create efficient candidates for the Metropolis-Hastings algorithm by adopting Hamiltonian dynamics, and can sometimes substantially reduce the time for convergence. In this regard, future research should explore an extensive model that

covers both spatial and temporal changes with a GPD distribution. Another possible extension of the model is to develop a model where spatial correlation can vary by soil type, climate, elevation, etc. Today the availability of abundant and accurate data on those factors may enable developing an advanced model to estimate the crop yield distribution and measure tail risks. One of the great strengths of the Bayesian Kriging method is that it could be widely adopted not only in crop insurance rating but also in any area of research that involves spatial correlation. Therefore, future work can extend both the application and development of the model to other research areas and issues related to spatial correlation.

[1] Area protection plans were first established in 1993 under the name Group Risk Plan (GRP).

[2] Our Bayesian Kriging model has multiple advantages over the Bayesian Model Averaging (BMA) approach, but there may be instances where the BMA approach could be preferred. Our approach has an explicit functional form for spatial correlation. Our approach yields estimates of spatial correlation parameters not provided by BMA. Bayesian methods are used because they allow estimating the GPD distribution with more precision. The BMA approach is more of a nonparametric approach to spatial smoothing. BMA is less restrictive than our approach, but also less precise if our restrictions are at least close to being true. As implemented, the BMA approach has used contiguity rather than a function of distance as in our model. We would also argue that our approach may be more intuitive and easier to explain to RMA and to producers.

[3] To estimate the target county's density, for example, the model not only uses the target county's historical yields but also uses all other counties' observations. The level of spatial effects (i.e., weight of the spatial smoothing) from other counties' distributions to the target county is smoothly interpolated by the Euclidean distance between the counties and Kriging parameters in the spatial covariance function.

[4] We fit the tail of the crop yield distribution using GPD. Note that the Bayesian spatial smoothing approach can also be applied to other density functions such as the normal or Generalized Extreme Value distribution.

[5] We refer to the Harri et al. (2011) model as the RMA model. The exact RMA model is proprietary and not known. We compare our model to the Harri et al. (2011) model on which the RMA model is based. The RMA is believed to do some heuristic adjustments including some spatial smoothing.

[6] Unlike Generalized Extreme Value (GEV) distribution, GPD does not have the location parameter (ς). Instead, GPD has an explicitly given threshold level u.

[7] Since a Generalized Pareto Distribution (GPD) only fits the tail of a crop yield distribution using observations over an explicitly given threshold, it cannot fit an entire yield distribution. Therefore, we use a normal distribution to estimate expected crop yield for each county. Note that parameters of the normal

distribution are obtained from the same Bayesian spatial smoothing procedures with GPD parameters and therefore the predicted yield for each county from the estimation are spatially smoothed as well.

[8] In the Gaussian spatial process, any subset of the field locations has a multivariate normal distribution. The covariance between any two locations (i.e., counties) is determined by a covariance function (or kernel) of the Gaussian spatial process evaluated at the spatial points of two locations..

[9] Note that the spatial covariance matrix Σ_ϕ is $N \times N$ matrix. If the covariance matrix has exponential type, then $\Sigma_\phi = \rho_\phi e^{-D_{ij}/\theta_\phi} = \rho_\phi \begin{bmatrix} 1 & & e^{-D_{1N}/\theta_\phi} \\ \vdots & \ddots & \vdots \\ e^{-D_{N1}/\theta_\phi} & & 1 \end{bmatrix}$.

[10] When the model draws from the candidate density for GPD parameters in the MCMC procedure, Bayesian Kriging is applied. The model generates K MCMC draws and each random draw creates an $N \times 1$ (N is the number of locations) vector $\boldsymbol{z}= [z_1, \dots, z_N]'$. Note that for the kth MCMC draw, $\boldsymbol{z}_k \sim N(0,1)$, and therefore $\sum_{i=1}^{N} z_{ik} \approx 0$. Next, the model conducts a Cholesky decomposition for $\Sigma_\phi = \mathbf{A}\mathbf{A}'$, where $\mathbf{A}$ is a lower triangular matrix. Then using the equation $\boldsymbol{\phi}_k = \boldsymbol{\mu}_k + \mathbf{A}\boldsymbol{z}_k$, the model draws the kth posterior value $\boldsymbol{\phi}$, say $\boldsymbol{\phi}_k$, from the kth posterior value of vector $\boldsymbol{\mu}_k$ and the kth random draw vector $\boldsymbol{z}_k$, and accepts or rejects the draw under the Metropolis-Hastings algorithm. We express this Gaussian spatial process as $\boldsymbol{\phi} \sim MVGP(\boldsymbol{\mu}, \Sigma_\phi)$.

[11] Simply including a trend variable as an additional covariate in the process layer may seem to be the simplest way to reflect the temporal trend of crop yield. However, only time-invariant covariates can be included in the current model with the non-normal likelihood layer. Therefore, in this type of Bayesian setting, adding a Gaussian temporal process can be a way to consider the temporal trend of crop yield.

[12] Since the sill parameter determines the maximum variogram value, we use the empirical variogram to find the prior information of the spatial structure of the GPD parameters. For some observations (i.e., GPD parameters) Z_i in location $i = 1, \dots, N$, empirical variogram can be defined as

$$\hat{\gamma}(D_{ij}) \coloneqq \frac{1}{2|M(D_{ij})|} \sum_{(i,j)\in M(D_{ij})} (Z_i - Z_j)^2$$

where $M(D_{ij})$ is the number of possible pairs of location i and j, D_{ij} is Euclidean distance.

[13] An F-madogram ($\nu_{F_{ij}}$) is similar to a variogram. Unlike a varioram, the F-madogram transforms the observations using a Fréchet marginal and calculates the expectation of absolute difference of the transformed values between two locations, $\nu_{F_{ij}} = \frac{1}{2}\mathrm{E}\left[\left|\exp(-\frac{1}{y_i}) - \exp(-\frac{1}{y_j})\right|\right]$. The 0.17 of F-madogram indicates that the maximum value of F-madogram within all possible pairs of two locations in the dataset.

[14] Similar to Harri et al. (2011), we first pool insurance rates for the counties from the RMA method and the proposed method. Then each pool of premium rates from the two methods has 39 times 15 observations, where 39 is the number of counties and 15 is the number of years. Using the pool of premium rates, we draw 10,000 samples with 15 observations from the non-parametric bootstrapping method. By comparing the premium rates from the two approaches, we group the rates under the ceded and retained policies, and calculate the loss ratio under the retained and ceded policies. From this procedure, we have the distribution of the loss ratio under the null hypothesis that premium rates from the two methods are identical. Finally, we compare the bootstrapped samples with actual loss ratio and obtain the *p*-value.

Table I-1. Deviance Information Criterion (DIC) for Models of Oklahoma County Wheat Yield

Spatial Effect	Model	Specification	$\bar{D}$	p_D	DIC
No spatial smoothing	Model 0	$\phi_i = \beta_0 + \varepsilon_i$	11517	40	11557
Matern	Model 1	$\phi_i = \beta_0 + \varepsilon_i$	11442	41	11483
	Model 2	$\phi_i = \beta_0 + \beta_1 \bar{y}_i + \varepsilon_i$	11440	43	11483
	Model 3	$\phi_i = \beta_0 + \beta_1 lon_i + \beta_2 lat_i + \varepsilon_i$	11441	45	11486
	Model 4	$\phi_i = \beta_0 + \beta_1 \bar{y}_i + \beta_2 lon_i + \beta_3 lat_i + \varepsilon_i$	11439	48	11487
exponential	Model 5	$\phi_i = \beta_0 + \varepsilon_i$	11441	40	11481
	Model 6	$\phi_i = \beta_0 + \beta_1 \bar{y}_i + \varepsilon_i$	11440	44	11484
	Model 7	$\phi_i = \beta_0 + \beta_1 lon_i + \beta_2 lat_i + \varepsilon_i$	11439	45	11484
	Model 8	$\phi_i = \beta_0 + \beta_1 \bar{y}_i + \beta_2 lon_i + \beta_3 lat_i + \varepsilon_i$	11440	48	11488

Table I-2. Estimated Loss Ratio under the New Model and RMA (Harri et al, 2011) Model

Model	Coverage Rate (%)	Mean Loss Ratio	Variance of Loss Ratio	Max Loss Ratio	Min Loss Ratio
RMA	70	2.66	8.14	15.72	0.00
	90	1.43	0.36	3.07	0.47
New model	70	2.10	2.40	5.35	0.05
	90	1.19	0.21	2.21	0.35

Table I-3. County Level Loss Ratios of New Model and RMA (Harri et al, 2011) Model

County	New model		RMA model	
	70%	90%	70%	90%
Alfalfa	3.62	1.25	5.57	2.47
Beaver	0.33	0.68	0.77	0.90
Blaine	2.48	1.29	2.25	1.48
Caddo	0.58	0.83	0.04	0.75
Canadian	1.54	1.08	4.93	1.67
Cimarron	0.76	0.82	6.88	2.39
Cleveland	4.51	1.34	6.33	2.07
Comanche	2.24	1.65	1.00	1.03
Cotton	3.88	1.90	1.49	1.27
Craig	0.35	0.35	1.01	0.71
Custer	2.28	1.22	1.37	1.45
Dewey	3.57	1.75	3.14	1.84
Garfield	5.34	1.48	5.12	1.41
Garvin	0.50	0.56	1.07	0.87
Grady	1.68	1.43	0.33	1.21
Grant	3.51	1.44	1.46	1.07
Greer	1.14	1.00	1.31	1.16
Harper	1.11	1.27	0.67	1.02
Kay	3.86	1.84	1.63	1.77
Kingfisher	1.61	0.88	2.37	1.76
Kiowa	0.62	1.08	0.37	1.06
Logan	5.08	1.71	1.20	1.65
Major	0.38	0.84	0.22	0.69
Mayes	0.43	0.52	2.38	1.10
McClain	2.35	0.61	0.86	0.48
Noble	5.25	2.21	5.06	1.94
Oklahoma	0.22	0.67	3.51	1.77
Osage	1.12	1.04	2.17	1.21
Ottawa	2.46	1.18	3.92	1.77
Pawnee	2.33	1.77	0.14	0.74
Payne	2.55	1.66	1.92	2.13
Pottawatomie	0.48	0.94	15.72	3.08
Roger Mills	1.00	0.80	4.71	1.84
Stephens	0.06	0.72	0.00	0.62
Tillman	2.48	1.64	1.85	1.42
Wagoner	3.30	1.82	3.94	2.56
Washita	1.28	0.66	4.24	1.25
Woods	4.08	1.54	1.76	1.49
Woodward	1.29	0.81	0.83	0.78

Table I-4. Loss Ratio for Ceded and Retained Policies

Coverage Rate	Loss ratio of Ceded Policies	Loss ratio of Retained Policies	Ceded to Retained Ratio	Percent of Retained Policies	*p*-value
70%	5.34	1.27	4.20	0.38	0.017
90%	2.44	1.12	2.18	0.26	0.046

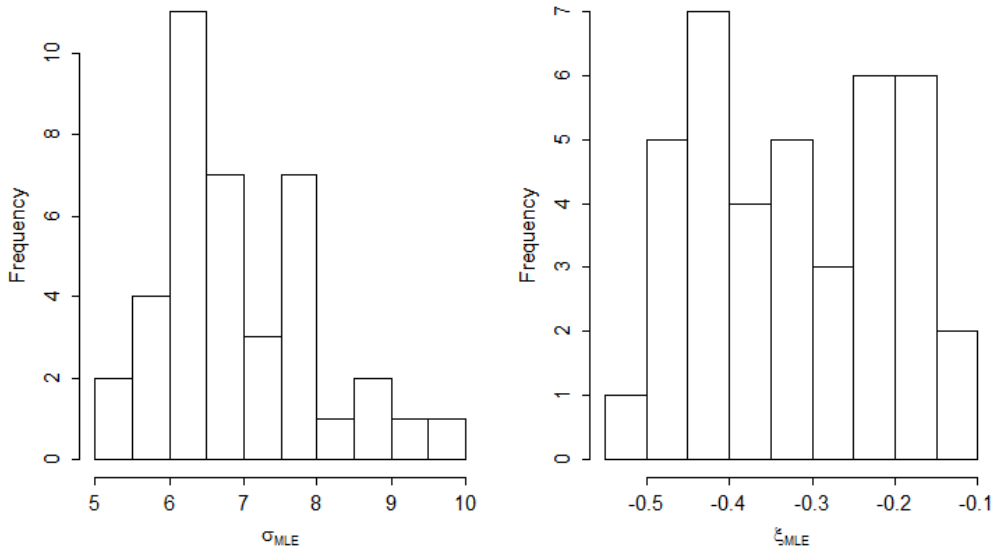

Figure I-1. Histogram of MLE estimates for parameters of GPD.

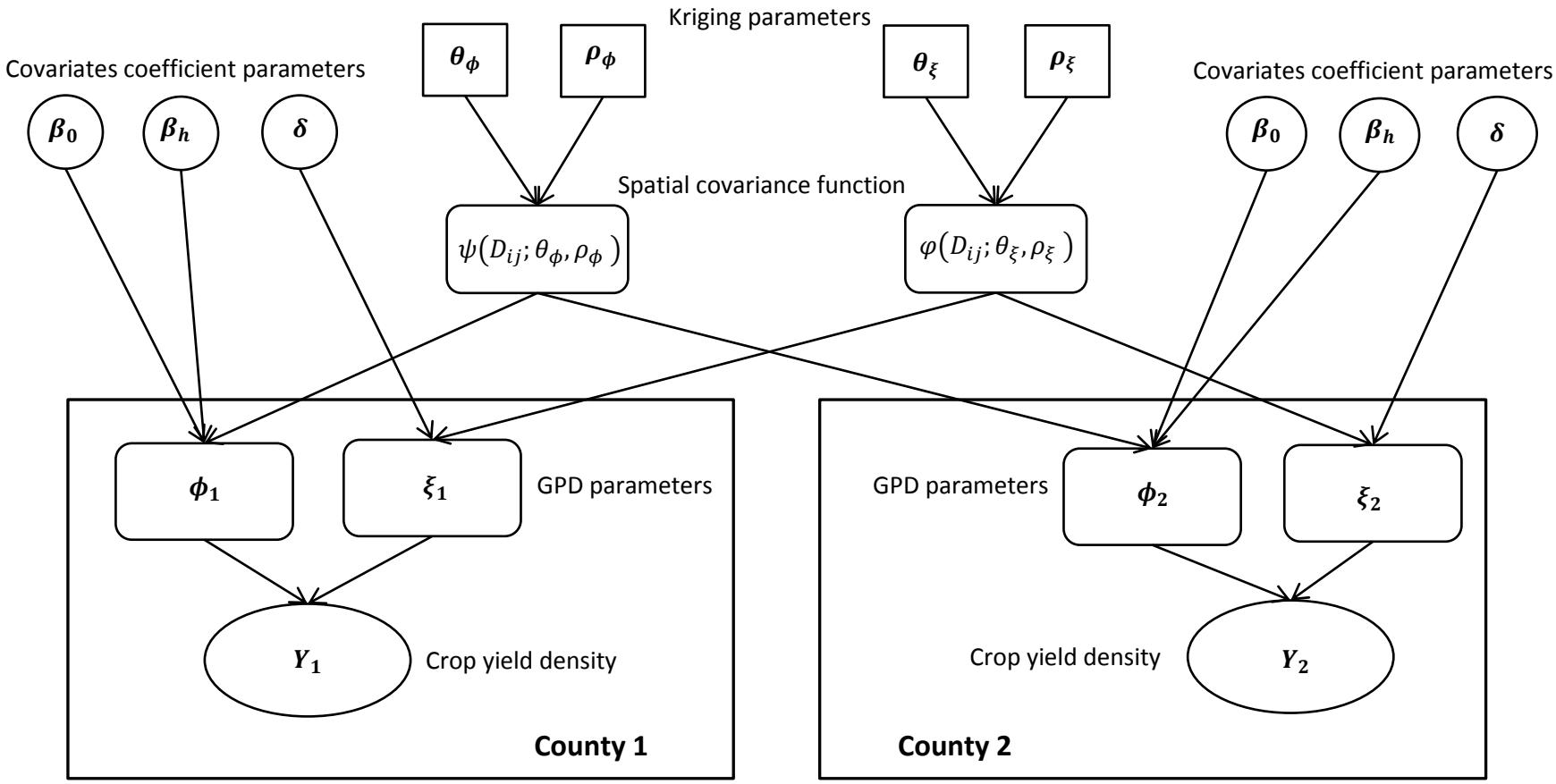

Figure I-2. Schematic of the Bayesian hierarchical structure for the spatial smoothing

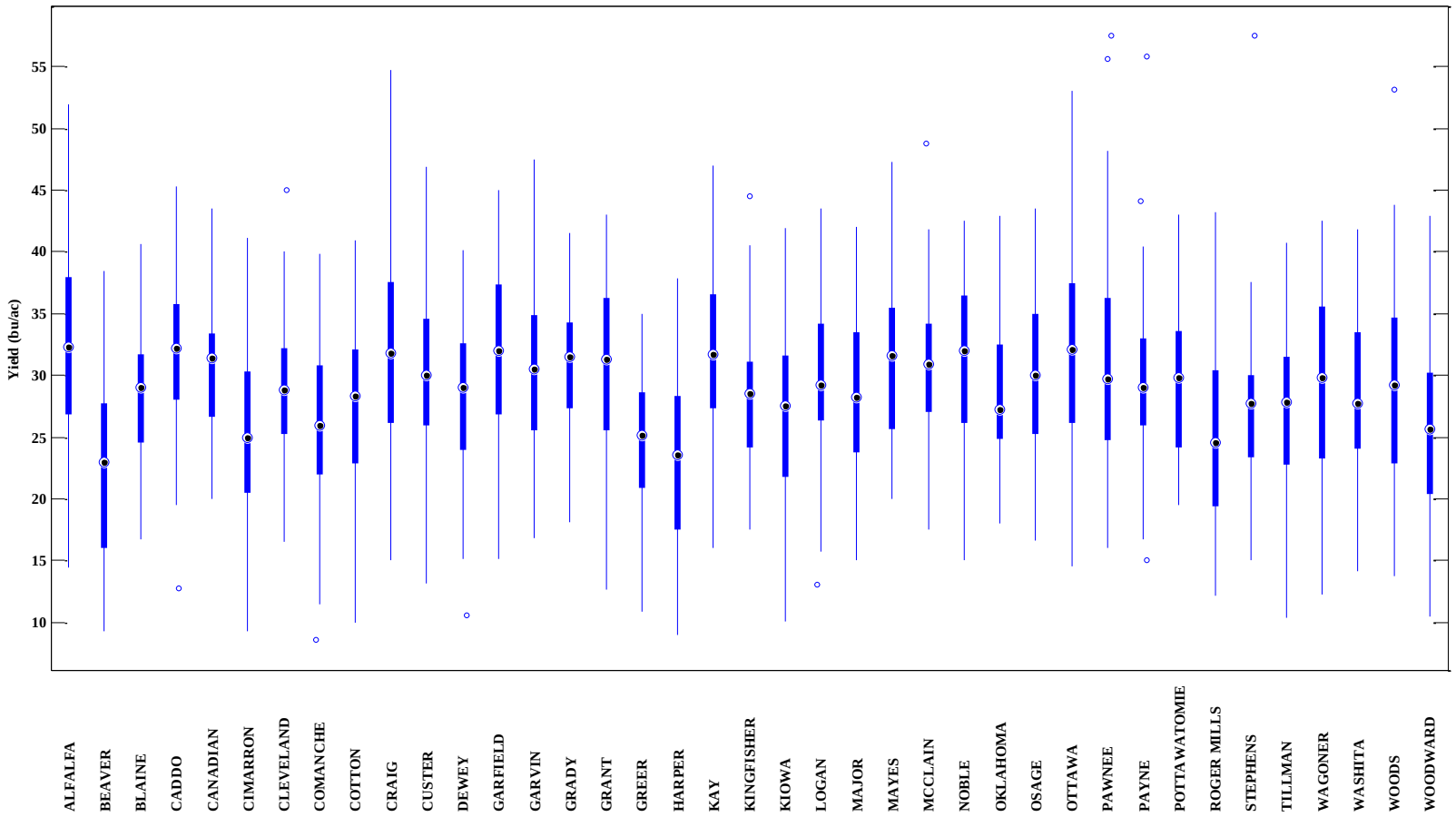

Figure I-3. Box plot of county level wheat yield in Oklahoma (1970 – 2014).

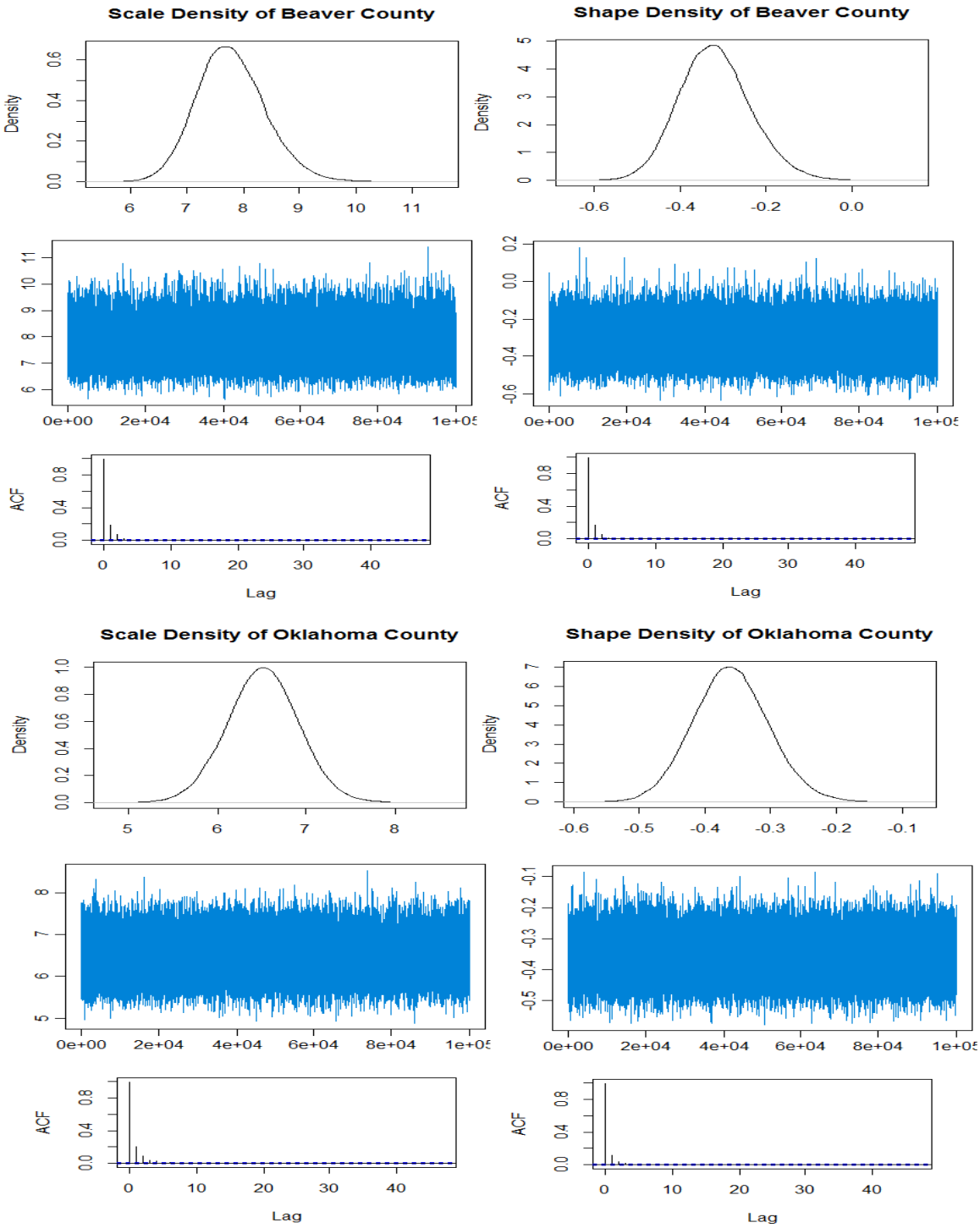

Figure I-4. Posterior densities of scale and shape parameters for Beaver (left) and Oklahoma (right) county.

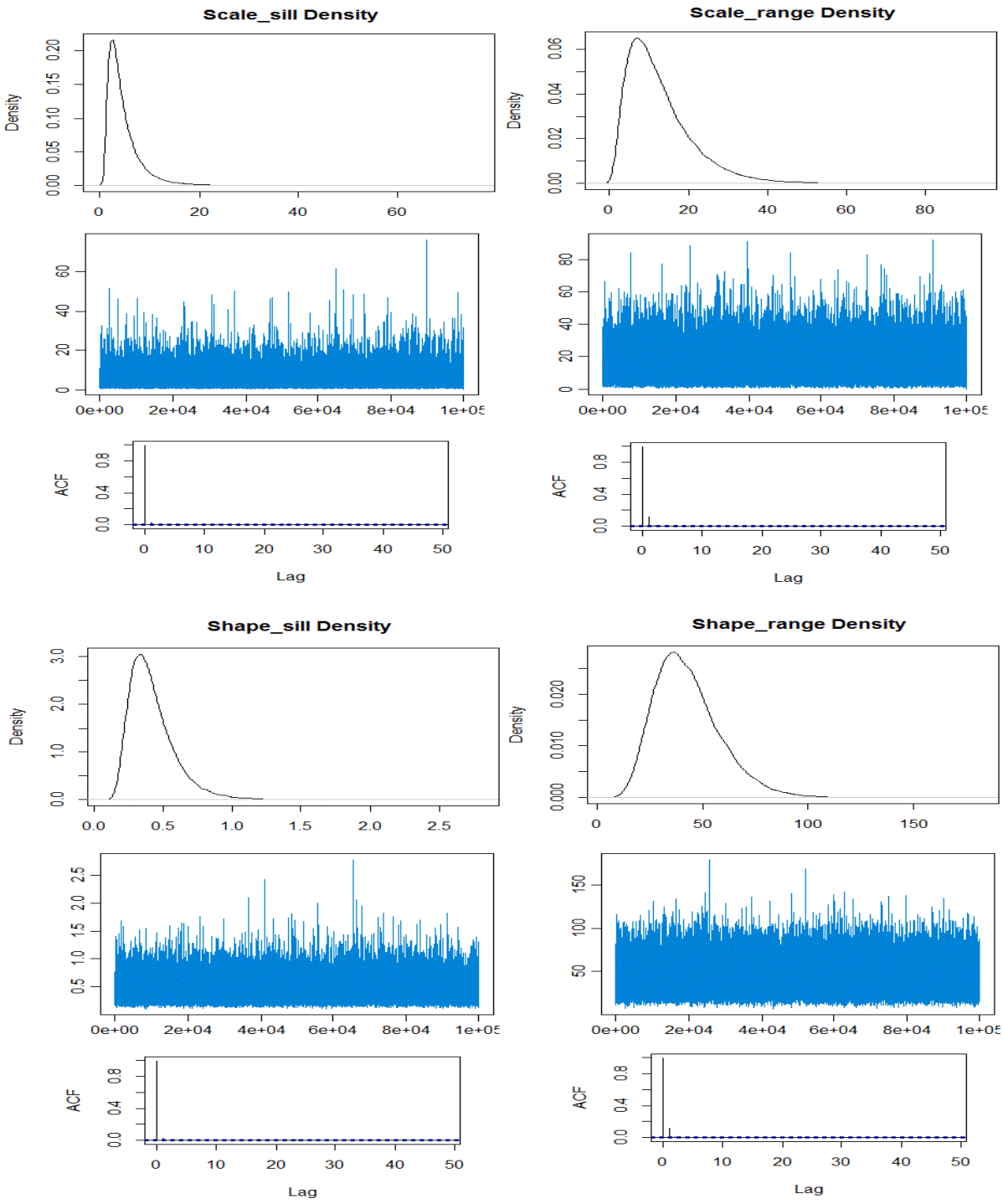

Figure I-5. Posterior densities of Kriging parameters

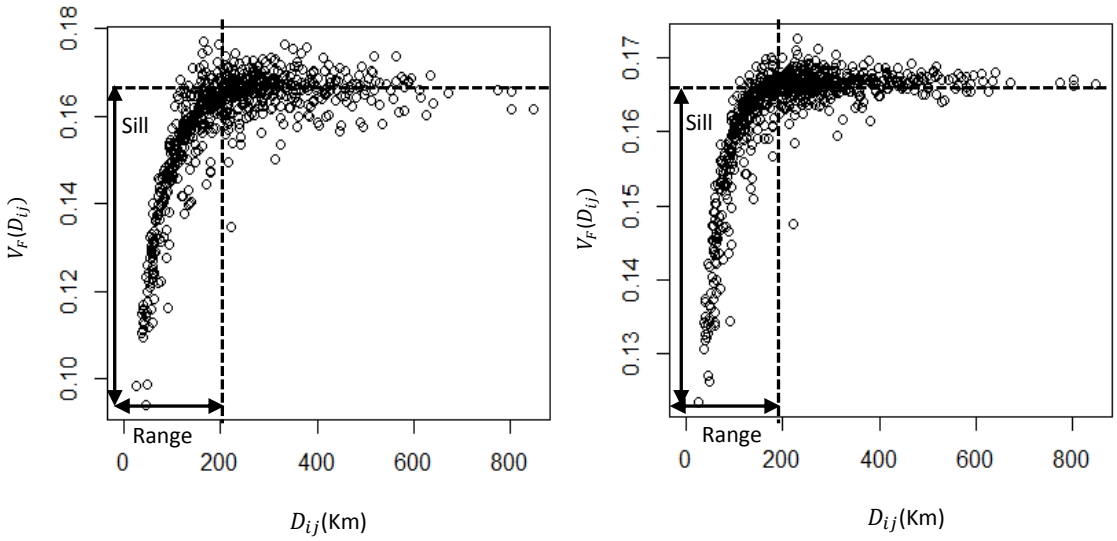

Figure I-6. F-Madogram of scale (left) and shape parameters (right)

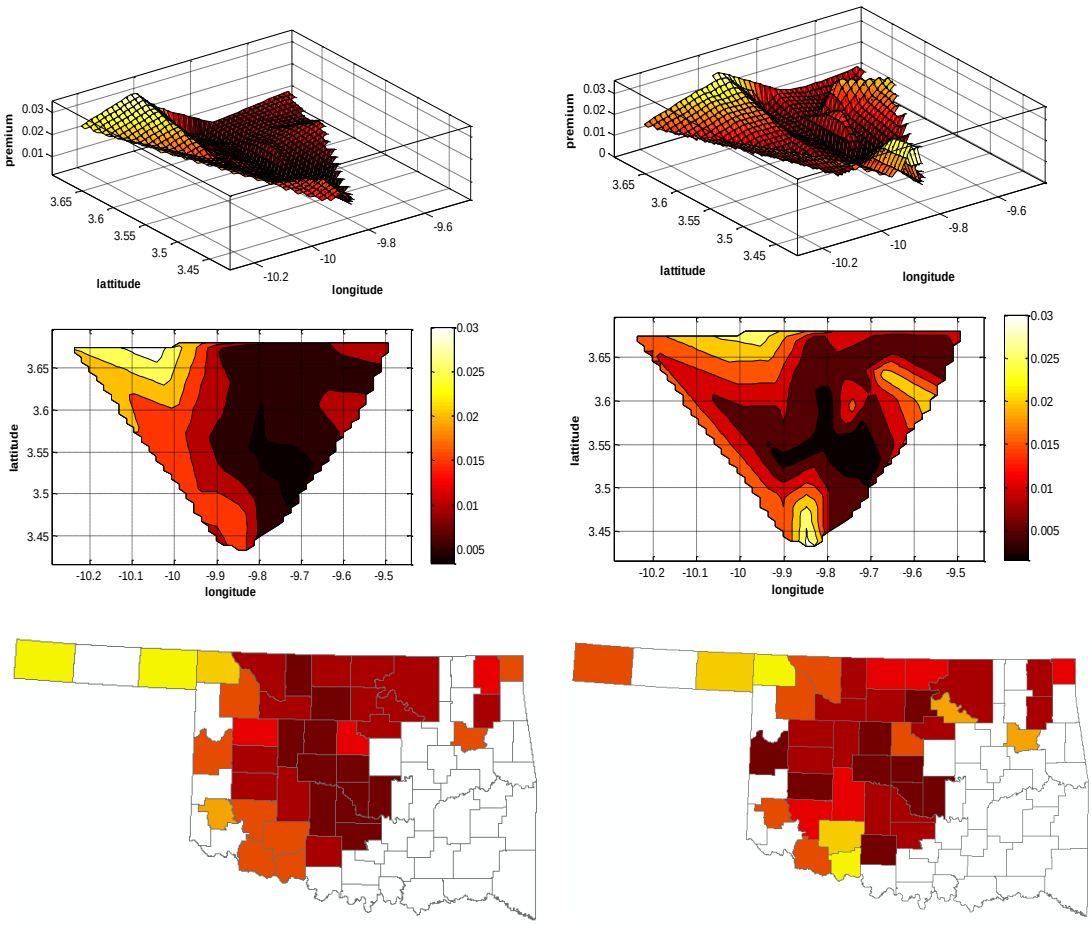

Figure I-7. 70% Premium rates from new model (left) and RMA (Harri et al, 2011) model (right)

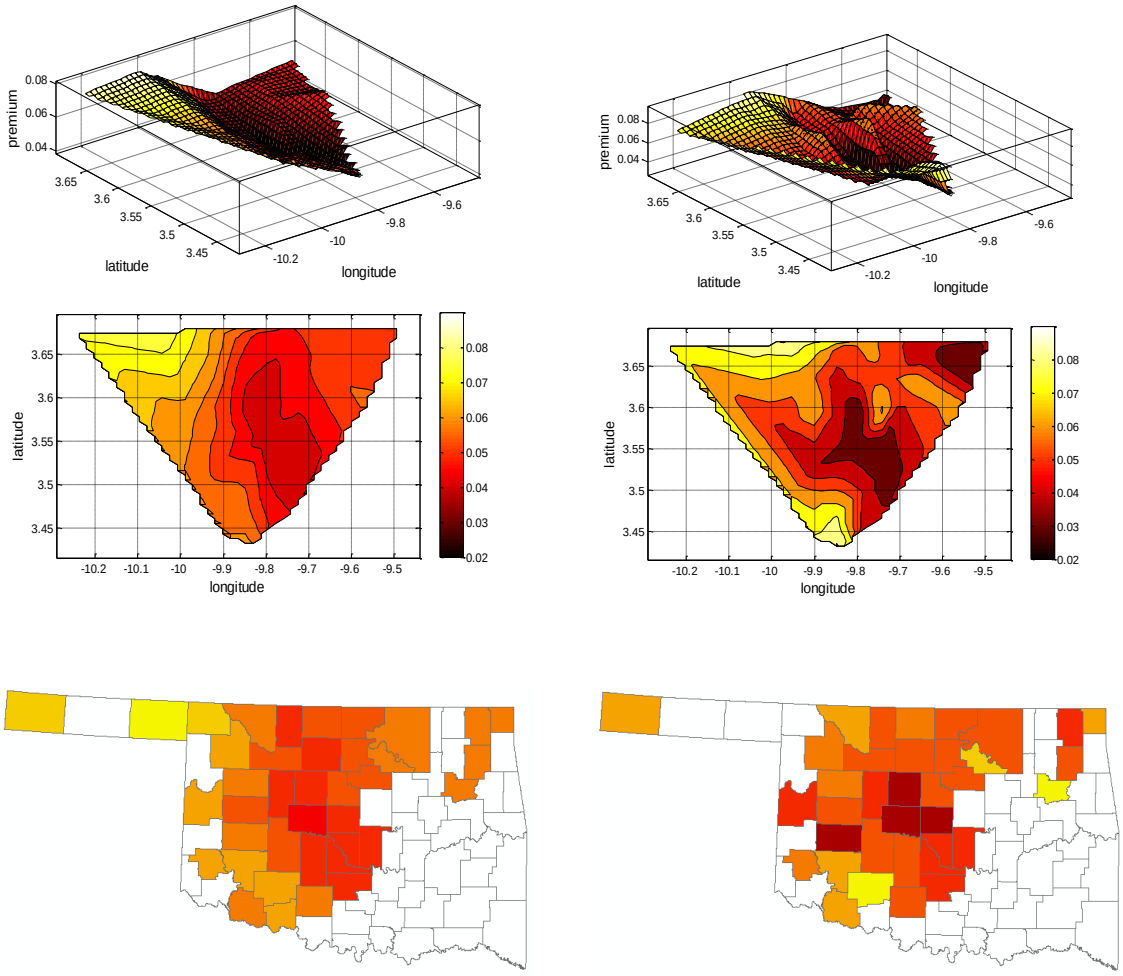

Figure I-8. 90% Premium rates from new model (left) and RMA (Harri et al, 2011) model (right)

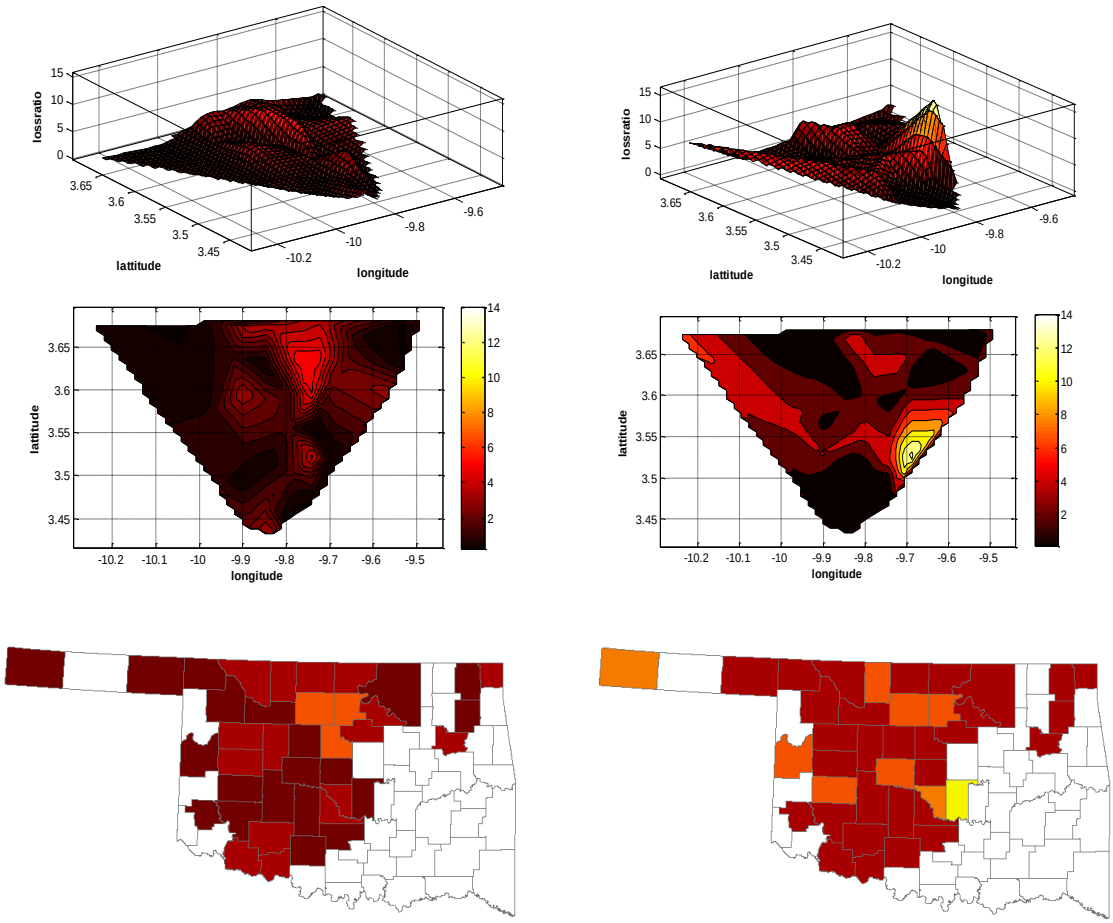

Figure I-9. Loss ratio (70% coverage rate) of new model (left) and RMA (Harri et al, 2011) model (right).

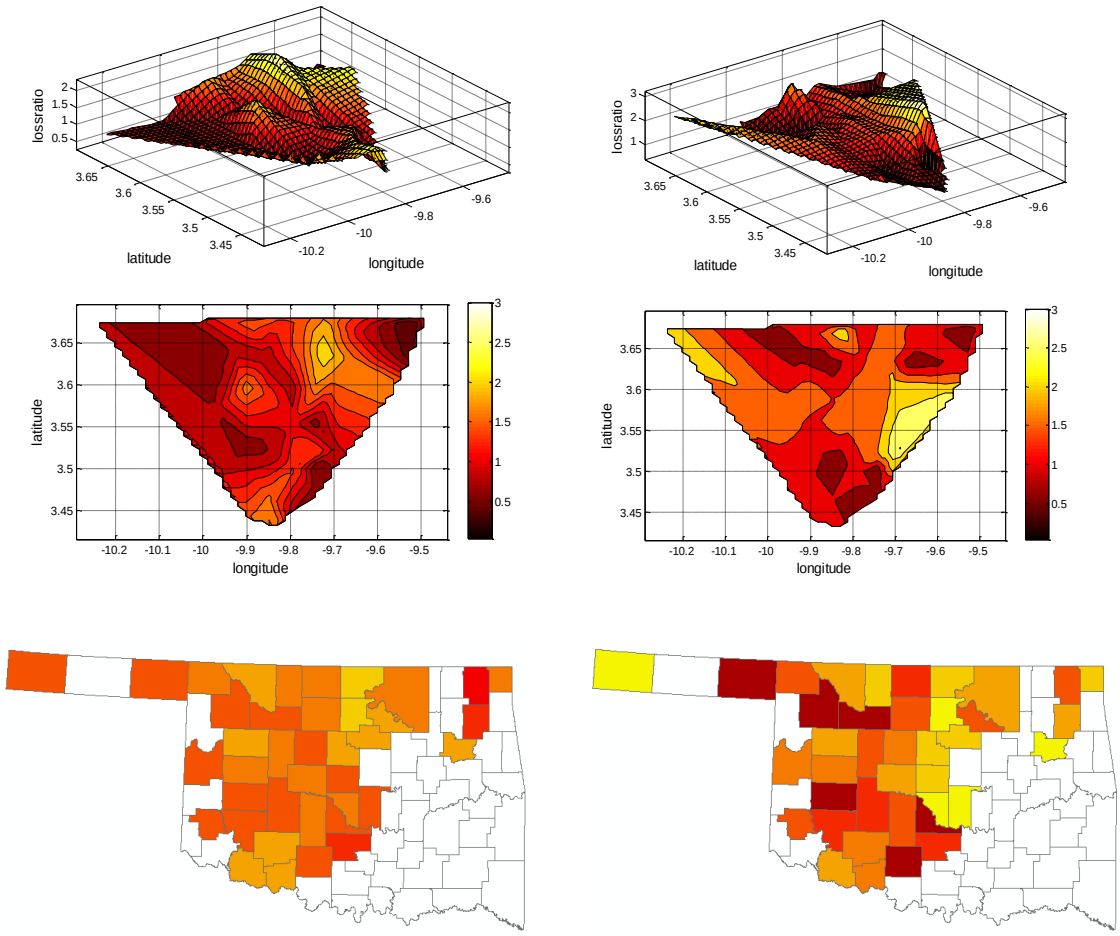

Figure I-10. Loss ratio (90% coverage rate) of new model (left) and RMA (Harri et al, 2011) model (right).

REFERENCES

Annan, F., J. Tack, A. Harri, and K.H. Coble. 2013. Spatial Pattern of Yield Distributions: Implications for Crop Insurance. *American Journal of Agricultural Economics* 96(1):253–68.

Banerjee, S., B. Carlin, and A. Gelfand. 2004. *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, FL: Chapman & Hall/CRC Press.

Berger, J., V. DeOiveira, and B. Sanso. 2001. Objective Bayesian Analysis of Spatially Correlated Data. *Journal of the American Statistical Association* 96:1361–74.

Coble, K.H., T.O. Knight, R.D. Pope, and J.R. Williams. 1996. Modeling Farm-Level Crop Insurance Demand with Panel Data. *American Journal of Agricultural Economics* 78:439–47.

Coble, K.H., R. Dismukes, and J.W. Glauber. 2007. Private Crop Insurers and the Reinsurance Fund Allocation Decision. *American Journal of Agricultural Economics* 89:582–95.

Cooley, D., D. Nychka, and P. Naveau. 2007. Bayesian Spatial Modeling of Extreme Precipitation Return Levels. *Journal of the American Statistical Association* 102:824-40.

Cooley, D., P. Naveau, and P. Poncet. 2006. Variograms for Spatial Max-stable Random Fields. *Dependence in Probability and Statistics* 187:373-90.

Cressie, N.A.C. 1993. *Statistics for Spatial Data*. Wiley Series in Probability and Statistics. New York: John Wiley & Sons Inc.

Du, X., C.L. Yu., D.A. Hennessy, and R. Miao. 2015. Geography of Crop Yield Skewness. *Agricultural Economics* 46(4):463–73.

Finley, A.O., S. Banerjee, and B.P. Carlin. 2007. spBayes: A Program for Multivariate Point-Referenced Spatial Modeling. *Journal of Statistical Software* 19:4.

Gelman, A., J.H.S. Carlin, and D. Rubin. 2014. *Bayesian Data Analysis*. Vol. 2. Boca Raton, FL, USA: Chapman & Hall/CRC.Goodwin, B.K., and A.P. Ker. 1998. Nonparametric Estimation of Crop Yield Distributions: Implications for Rating Group-Risk (GRP) Crop Insurance Contracts. *American Journal of Agricultural Economics* 80:139–53.

Harri, A., K.H. Coble, A.P. Ker, and B.J. Goodwin. 2011. Relaxing Heteroscedasticity Assumptions in Area-Yield Crop Insurance Rating. *American Journal of Agricultural Economics* 93(3):707–17.

Jung, A.R., and C.A. Ramezani. 1999. Valuing Risk Management Tools as Complex Derivatives: An Application to Revenue Insurance. *Journal of Financial Engineering* 8:99–120.

Just, R.E., and Q. Weninger. 1999. Are Crop Yields Normally Distributed? *American Journal of Agricultural Economics* 81:287–304.

Ker, A.P., T.N. Tolhurst, and Y. Liu. 2015. Bayesian Estimation of Possibly Similar Yield Densities: Implications for Rating Crop Insurance Contracts. *American Journal of Agricultural Economics* 98(2):360–82

Ker, A.P., and K.H. Coble. 2003. Modeling Conditional Yield Distributions. *American Journal of Agricultural Economics* 85:291–304.

Ker, A.P., and B.K Goodwin. 2000. Nonparametric Estimation of Crop Insurance Rates Revisited. *American Journal of Agricultural Economics* 82(2):463–78.

Nelson, C.H., and P.V. Preckel. 1989. The Conditional Beta Distribution as a Stochastic Production Function. *American Journal of Agricultural Economics* 71(2):370–8.

Ntzoufras. I. 2011. *Bayesian Modeling Using WinBUGS*. Vol. 698. John Wiley & Sons.

Ozaki, V., S.K. Ghosh, B.K. Goodwin, and R. Shirota. 2008. Spatio-Temporal Modeling of Agricultural Yield Data with an Application to Pricing Crop Insurance Contracts. *American Journal of Agricultural Economics* 90(4):951–61.

Risk Management Agency. 2013. Area Risk Protection Insurance Policy. http://www.rma.usda.gov/policies/2014/14-arpibp.pdf. Accessed, September 2016.

Sherrick, B.J., F.C. Zanini, G.D. Schnitkey, and S.H. Irwin. 2004. Crop Insurance Valuation under Alternative Yield Distributions. *American Journal of Agricultural Economics* 86:406–19.

Skees, J. R., J.R. Black, and B.J. Barnett. 1997. Designing and Rating an Area Yield Insurance Contract. *American Journal of Agricultural Economics* 79:430–8.

Spiegelhalter, D., N. Best, B. Carlin, and A. van der Linde. 2002. Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society* 64: 583–639.

Stokes, J.R. 2000. A Derivative Security Approach to Setting Crop Revenue Coverage Insurance Premiums. *Journal of Agricultural and Resource Economics* 25:159–76.

Tirupattur, V., R.J. Hauser, and N.M. Chaherli. 1996. Crop Yield and Price Distributional Effects on Revenue Hedging. Working paper series: *Office of Futures and Options Research* 96-05:1–17.

Woodard, J.D. 2016. Determining Optimal Data Aggregation: An Application of Out-of-Sample Mixture Models. Paper presented at the 2016 Agricultural and Applied Economics Association Annual Meeting in Boston, Massachusetts, July 31 – Aug 2.

Woodworth, G. 2004. Biostatistics: A Bayesian Introduction. New York: Wiley.

CHAPTER II

SPATIALLY SMOOTHED CROP YIELD DENSITY ESTIMATION:

PHYSICAL DISTANCE VS CLIMATE SIMILARITY

Abstract

Crop yield density tends to be spatially correlated because nearby areas share similar climate and agronomic characteristics. Many crop insurance studies have pointed out that the spatial yield correlation should be considered to provide more precise premium rating. Bayesian Kriging for spatial smoothing offers a promising way to use such spatial correlation when estimating crop yield densities. This article contributes to agricultural economics literature by providing a spatial smoothing method based on a climate space, which is composed of climatological measures. We compare the spatial smoothing from the climate space and a general physical space (longitude-latitude space) to evaluate the performance of each method. We use loss ratio of crop insurance to test the performance for county level yearly corn yield data from six U.S. states. Spatial smoothing from climate space dominates the results from the physical space in out-of-sample prediction and mitigates regional inequalities in crop insurance loss ratios. The climate space notably outperforms the physical space in Colorado that has varying climate due to its varying topography.

Key words: Bayesian hierarchical structure, Bayesian spatial smoothing, Bayesian Kriging, climate space, crop insurance, crop yield similarity, physical space, spatial correlation.

Introduction

Yields of a given crop tend to be spatially correlated due to agronomical and climatological similarity of nearby areas. A risk inherent in crop yield, therefore, can also be spatially correlated due to such spatial correlation in crop yields. As a result, shortfalls of crop yield in a particular region, such as a county or district, usually tend to be correlated with shortfalls of yield in the neighboring regions. Previous literature (Annan et al. 2014; Du et al. 2015; Goodwin 2015; Ker, Tolhurst, and Liu 2015; Woodard 2016) has discussed this issue and pointed out that considering spatial correlation could more accurately measure downside yield risk and thus reduce adverse selection in crop insurance.

Several statistical methods have been suggested in agricultural economics literature to reflect the spatial correlation in estimating crop yield density. For example, Goodwin and Ker (1998) use pooled observations from surrounding counties. Ozaki et al. (2008) use a spatial weighting matrix in an attempt to consider spatial correlation in crop yield. They impose uniform weights on parameter estimates from surrounding counties but impose zero weights beyond the surrounding counties. Ozaki and Silva (2009) propose a skewed normal multivariate conditional yield distribution for spatial smoothing. However, similar to Ozaki et al. (2008), they do not consider correlation beyond surrounding counties. Current area-based crop insurance programs are rated with

a model suggested by Harri et al. (2011). To reflect the spatial correlation, the model imposes a district level restriction on the county level parameters. Other studies consider spatial correlation using Bayesian Model Averaging (BMA). Ker, Tolhurst, and Liu (2015) estimate a posterior density for each county using observations of each county and then take Bayesian averaging of its own posterior density and densities from other counties. Woodard (2016) employs BMA to get a weighted average of county and district level parameters. More recently, Park, Brorsen, and Harri (2016) suggest Bayesian Kriging as a method for spatial smoothing. Their Bayesian Kriging approach produces spatially smoothed parameter estimates that vary smoothly over space. Their weight for smoothing is determined by a physical distance in longitude-latitude space (i.e., physical space).

In addition to the Bayesian Kriging method for spatial smoothing under the traditional way of using physical space, we offer an alternative spatial smoothing under a space with climatological coordinates, which is a climate space. The climate space uses temperature and precipitation as coordinates rather than latitude and longitude. We use an average number of days in months (July and August) with maximum temperature greater than or equal to 90°F (DT90) and a total average precipitation amount (mm) for the months from May to August (TPCP) as the two axes. In this application, our focus is on how the distribution of crop yield varies over two different types of spaces (physical and climate space). We then compare the performances of the estimates from these two spaces using a loss ratio under a crop insurance program.

The primary goal of the article is to suggest a new method for actuarially accurate crop insurance rating considering the spatial correlation of the crop yield densities. We

extend the Bayesian Kriging spatial smoothing method to use climate space instead of physical distance. We are not aware of any literature in agricultural economics that uses spatial smoothing based on climate space. We evaluate and compare the performances of the spatial smoothing from physical space and from climate space.

We choose corn as a crop for evaluating the performance of each model. We utilize annual county level yield data from the National Agricultural Statistics Service (NASS) for Iowa, Illinois, Nebraska, Minnesota, Indiana, and Colorado from 1955 to 2014. We find that climate space performs better or at least similar to physical space in every state in the dataset. Specifically, in Colorado and Nebraska, climate space substantially mitigates regional inequalities of loss ratio for crop yield densities.

In the following section, we discuss a theoretical framework for Bayesian hierarchical structure of our Bayesian Kriging method. In the empirical application section, we explain the dataset used for empirical estimation and describe two different types of smoothing spaces. We then introduce premium calculating procedures for evaluating the performance of the models from the different smoothing spaces. The last section has conclusions.

Theoretical Framework

We use a Kriging method for spatial smoothing. Kriging is a geostatistical spatial interpolation method that has been actively employed in a broad variety of disciplines. The method assumes that spatial correlation (i.e., density similarity) varies smoothly and decreases with the distance between locations. Note that regardless of which space

(physical or climate) is used for the spatial smoothing, the estimation procedure is identical.

Overview of the Bayesian hierarchical structure

The Kriging method here is estimated under a Bayesian hierarchical structure. A Bayesian hierarchical model can be specified when Bayesian modeling structure can be written in multiple levels (i.e., hierarchies). In a Bayesian hierarchical framework, therefore, a prior distribution of the general Bayesian model can also be structured as additional prior parameters, called hyper-priors.

We consider two types of specifications for the process layer: a Gaussian spatial process type (GP) and an auto-regressive type with Gaussian spatial process[1] (AR). GP only considers spatial correlation of the crop yield distribution based on the Gaussian spatial process whereas AR takes into consideration both spatial and temporal correlation (spatio-temporal) using the Gaussian spatial process and the auto-regressive process. Both GP and AR can be represented in the Bayesian hierarchical structure with three layers: likelihood layer, process layer, and prior layer. In the likelihood layer of the hierarchy, the crop yield distribution for each county is assumed to follow a normal distribution. Second, the process layer models the spatial and temporal structure for parameters of the crop yield distribution. In this layer, we only model the hierarchical structure of mean parameter μ of the crop yield distribution[2]. The process layer has both deterministic and stochastic effects. The deterministic part of the process at each county is determined by a set of explanatory variables of the county and the stochastic part will operate the spatial and temporal smoothing process. The third layer of the hierarchy

consists of the prior density for the coefficients of the explanatory variables and Kriging parameters to conduct spatial smoothing, which are called hyper priors. The hierarchy we use can be structured as,

$$\begin{aligned} \boldsymbol{Y} \mid \boldsymbol{\mu}_t, \boldsymbol{\Theta} &\sim p_1(\boldsymbol{Y} \mid \boldsymbol{\mu}_t, \boldsymbol{\Theta}) \\ \boldsymbol{\mu}_t \mid \boldsymbol{\Theta} &\sim p_2(\boldsymbol{\mu}_t \mid \boldsymbol{\Theta}) \\ \boldsymbol{\Theta} &\sim p_3(\boldsymbol{\Theta}) \end{aligned} \tag{1}$$

where p_1, p_2, and p_3 are the density associated with each layer of the hierarchy, likelihood layer, process layer, and prior layer, respectively, $\boldsymbol{Y}$ is a matrix of crop yields that spans all counties ($n = 1, \dots, N$) and all years ($t = 1, \dots, T$), $\boldsymbol{\mu}_t$ is a vector of the mean parameters of the likelihood function at year t that contains all counties, where $\boldsymbol{\mu}_t = [\mu_{1t}, \dots, \mu_{Nt}]'$, and $\boldsymbol{\Theta}$ is a vector of hyper parameters, where $\boldsymbol{\Theta} = [\beta_1, \dots, \beta_K, \omega, \theta, \rho, \sigma^2]'$.

By Bayes' theorem, the joint posterior distribution of the model is

$$p(\boldsymbol{\mu}_t, \boldsymbol{\Theta} \mid \boldsymbol{Y}) \propto p_1(\boldsymbol{Y} \mid \boldsymbol{\mu}_t, \boldsymbol{\Theta}) p_2(\boldsymbol{\mu}_t \mid \boldsymbol{\Theta}) p_3(\boldsymbol{\Theta}). \tag{2}$$

Therefore, the joint posterior density of the model $p(\boldsymbol{\mu}_t, \boldsymbol{\Theta} \mid \boldsymbol{Y})$ is proportional to the multiplication of the three layers of the hierarchy, which will be specified in the following subsections.

Likelihood layer

A likelihood function of the crop yield distribution forms the first layer of the model. Both GP and AR assume that the crop yield of each county follows a normal distribution. Then, the first layer of the model, the likelihood layer is

$$p_1(\boldsymbol{Y}|\,\boldsymbol{\mu}_t, \boldsymbol{\Theta}) \;=\; \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{(\boldsymbol{y}_t - \boldsymbol{\mu}_t)'(\boldsymbol{y}_t - \boldsymbol{\mu}_t)}{2\sigma^2} \tag{3}$$

where $\boldsymbol{y}_t$ denotes a vector of crop yield at year t that spans all counties, $\boldsymbol{y}_t = [y_{1t}, \dots, y_{Nt}]'$, $\boldsymbol{\mu}_t$ is a vector of the mean parameter at year t that includes all counties, and $\boldsymbol{\Theta}$ is a vector of hyper parameters, $\boldsymbol{\Theta} = [\beta_1, \dots, \beta_K, \omega, \theta, \rho, \sigma^2]'$.

Process layer

In the second layer of the hierarchy, we model the spatial process of mean parameters μ for each GP and AR accounts for spatial / spatial and temporal correlations relevant to crop yield distribution. Since we assume a Gaussian spatial process, mean parameters of all counties are assumed to be multivariate normally distributed[3]. Spatial and temporal smoothing for the parameter $\boldsymbol{\mu}_t$ is conducted from the stochastic part of the process via Gaussian spatial process and first order autoregressive, AR(1), process. The level of spatial dependence is measured from a spatial covariance matrix with Kriging parameters, which captures the detailed spatial structure for the mean parameters $\boldsymbol{\mu}_t$.

First, the Gaussian spatial process (GP) allows spatial correlation of crop yield. The GP is specified as,

$$\boldsymbol{\mu}_t = \boldsymbol{\psi}_t + \boldsymbol{\eta} + \boldsymbol{\varepsilon}_t,$$

$$\boldsymbol{\psi}_t = \boldsymbol{X}_t\boldsymbol{\beta}$$

(4) $$\boldsymbol{\eta} \sim MVGP(\Sigma),$$

$$\Sigma_\eta = \rho_\eta e^{-D_{ij}/\theta_\eta},$$

$$\boldsymbol{\varepsilon}_t \sim MVN(\mathbf{0},\ \sigma^2\boldsymbol{I}),$$

where $\boldsymbol{\mu}_t$ is a vector of mean parameter for crop yield distribution at year t that contains all counties, $\boldsymbol{\mu}_t = [\mu_{1t}, \dots, \mu_{Nt}]'$, $\boldsymbol{\psi}_t$ deterministic part of the mean structure, $\boldsymbol{X}_t$ is a T by K matrix of explanatory variables that determine the mean structure at time t such as historical moving average yield level of each county and trend variable, $\boldsymbol{\beta}$ is a K by 1 vector of coefficients of explanatory variables $\boldsymbol{\beta} = [\beta_1, \dots, \beta_K]'$, $\boldsymbol{\eta}$ is the spatial random effects, $\boldsymbol{\eta} = [\eta_1, \dots, \eta_N]'$ that is assumed to follow a multivariate Gaussian process with exponential type spatial covariance matrix, $\Sigma = \rho e^{-D_{ij}/\theta}$, which is a function of Euclidean distance (D_{ij}) between counties i and j, sill parameter ρ, and range parameter θ, and $\boldsymbol{\varepsilon}_t$ is a non-spatial error component. In empirical part, we obtain the parameter estimates under the two different types of spatial smoothing structure: traditional physical space and climate space. Therefore, the distance D_{ij} between two counties in the spatial covariance matrix will be differently measured in accordance with which spatial space is used.

The second type of specification is an auto-regressive model with Gaussian spatial process (AR). The AR is defined as,

(5) $$\boldsymbol{\mu}_t = \omega\boldsymbol{\mu}_{t-1} + \boldsymbol{\psi}_t + \boldsymbol{\eta} + \boldsymbol{\varepsilon}_t,$$

$$\boldsymbol{\psi}_t = \boldsymbol{X}_t\boldsymbol{\beta}$$

$$\boldsymbol{\eta} \sim MVGP(\Sigma),$$

$$\Sigma = \rho e^{-D_{ij}/\theta},$$

$$\boldsymbol{\varepsilon}_t \sim MVN(\mathbf{0},\ \sigma^2\boldsymbol{I}),$$

where $\boldsymbol{\mu}_{t-1}$ is a vector of lagged mean parameter for crop yield distribution at year $t-1$ that contains all counties, $\boldsymbol{\mu}_t = [\mu_{1t-1}, \dots, \mu_{Nt-1}]'$, ω denotes the temporal correlation parameter under the first order auto regressive process that is assumed to be in the interval, $-1 < \omega < 1$, and all other parameters are identical to the GP model. Obviously, for $\omega = 0$, AR is exactly the same as the GP model.

The AR addresses temporal correlation of the crop yield together with its spatial correlation. Some factors that affect crop yield realizations (i.e., climate) tend to be correlated, both spatially and temporally, and thus adjacent counties would experiences spatial correlations of crop yields over multiple periods of time. The AR specification reflects these spatio-temporal aspects of the crop yield densities.

From the process for the mean parameters, the vector of the parameter $\boldsymbol{\mu}_t$ given the parameters , ω, θ, ρ, and σ^2 follows

(6) $$\boldsymbol{\mu}_t \mid \beta, \omega, \theta, \rho, \sigma^2 \sim MVGP(\boldsymbol{\psi}_t,\ \Sigma).$$

Then the process layer densities for GP and AR model can be specified as equation (7) and (8), respectively,

(7) $$p_2(\boldsymbol{\mu}_t|\boldsymbol{\Theta}) = \frac{1}{\sqrt{(2\pi)^N|\Sigma|}}\exp\left[-\frac{1}{2}(\boldsymbol{\mu}_t - \boldsymbol{\psi}_t)'\Sigma^{-1}(\boldsymbol{\mu}_t - \boldsymbol{\psi}_t)\right]$$

(8) $$p_2(\boldsymbol{\mu}_t|\boldsymbol{\Theta}) = \frac{1}{\sqrt{(2\pi)^N|\Sigma|}}\exp\left[-\frac{1}{2}(\boldsymbol{\mu}_t - \omega\boldsymbol{\mu}_{t-1} - \boldsymbol{\psi}_t)'\Sigma^{-1}(\boldsymbol{\mu}_t - \omega\boldsymbol{\mu}_{t-1} - \boldsymbol{\psi}_t)\right]$$

where $\boldsymbol{\psi}_t$ is a vector of the deterministic part of the mean process at time t defined by equations (4) and (5), and $\boldsymbol{\Theta} = [\beta, \omega, \theta, \rho, \sigma^2]'$ is a vector of hyper-parameters, Σ is spatial covariance parameter, where $\Sigma = \rho e^{-D_{ij}/\theta}$. Since we are in the Bayesian framework, the parameters of the crop yield process are treated as random variables. Therefore, we impose independent priors for the hyper parameters ($\beta, \omega, \theta, \rho, \sigma^2$) in the following prior layer.

Prior layer

The third layer of the hierarchy has the priors for the hyper-parameters $\boldsymbol{\Theta}$, which are parameters for explanatory variables, Kriging parameters, and variance parameter in the process layer. Since the model assumes the parameters in the prior layer are independent, a multiplication of each prior for the hyper parameters forms the prior layer. For convenience, we group the hyper parameters into three different types depending on their role in the process layer: coefficient, variance, and Kriging parameters. First, all coefficient parameters in the process layer $\beta_1, \dots, \beta_k$ and ω are given normal priors. We impose $N(0, 10^4)$ priors for each of the coefficient parameters. For the variance parameter σ^2, we impose general inverse gamma priors $IG(0.1, 0.1)$ same as Ozaki et al. (2008). However, imposing priors for the Kriging parameters (ρ, θ), which describe the spatial structure of the Gaussian spatial process, is more problematic than the other priors. There is a large Bayesian statistics literature (Berger, DeOivelira, and Sanso 2001; Banerjee, Carlin, and Gelfand 2004; Cooley, Nychka, and Naveau 2007) regarding

consistency of proper priors for the Kriging parameters that argues improper priors for such parameters may induce significant improper posteriors. Some statistics literature (Banerjee, Carlin, and Gelfand 2004; Sahu, Gelfand, and Holland 2006; Cooley, Nychka, and Naveau 2007) suggest an empirical Bayes method in which the Kriging priors are estimated from the empirical data to avoid improper priors. In this regards, we use the empirical information to find the priors of the Kriging parameters. Since the sill parameter ρ determines the maximum level of the variogram, which is a function describing the degree of spatial correlation of a stochastic spatial process, an empirical variogram is a general way to collect prior information about the sill parameter. Therefore, we first estimate the mean parameter of each county using maximum likelihood. Then using the estimated MLE parameters for each county, we fit the empirical variogram.[4] The results of the empirical variogram are used to impose inverse gamma prior for the sill parameter ρ since the value of sill parameter determines maximum of variogram. Two parameters of the inverse gamma prior are obtained from maximum likelihood by using the empirical variogram values.

The next step is to find the prior distributions for the range parameter θ. We use prior empirical distance information of the empirical data to impose prior for the range parameter since the range parameter θ determines maximum distance of the spatial effect. Two parameters of gamma prior for the range parameter θ are imposed based on the previous empirical distance information and maximum likelihood estimation.

With the priors as above, the third layer in equation (2) can be expressed as

$$p_3(\boldsymbol{\Theta}) = p(\beta_k)p(\omega)p(\rho_\eta)p(\theta_\eta)p(\sigma^2). \tag{9}$$

Joint posterior distribution

We now have densities for each hierarchy, $p_1(\boldsymbol{Y} \mid \boldsymbol{\mu}_t, \boldsymbol{\Theta})$, $p_2(\boldsymbol{\mu}_t \mid \boldsymbol{\Theta})$, and $p_3(\boldsymbol{\Theta})$ from the previous sections. The joint posterior distributions for our model can be obtained by multiplying these three layers. The logarithm of the joint posterior distributions of the GP is

$$\log p(\boldsymbol{\mu}_t, \boldsymbol{\Theta} \mid \boldsymbol{Y}) \propto -\frac{NT}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(\boldsymbol{y}_t - \boldsymbol{\mu}_t)'(\boldsymbol{y}_t - \boldsymbol{\mu}_t) - \frac{\sum_{t=1}^{T}\log|\Sigma|}{2} \tag{10}$$
$$-\tfrac{1}{2}\textstyle\sum_{t=1}^{T}(\boldsymbol{\mu}_t - \boldsymbol{\psi}_t)'\Sigma^{-1}(\boldsymbol{\mu}_t - \boldsymbol{\psi}_t) + \log(p_3(\boldsymbol{\Theta})).$$

Likewise, the logarithm of the joint posterior distributions of the AR is

$$\log p(\boldsymbol{\mu}_t, \boldsymbol{\Theta} \mid \boldsymbol{Y}) \propto -\frac{NT}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(\boldsymbol{y}_t - \boldsymbol{\mu}_t)'(\boldsymbol{y}_t - \boldsymbol{\mu}_t) - \frac{\sum_{t=1}^{T}\log|\Sigma|}{2} \tag{11}$$
$$-\tfrac{1}{2}(\boldsymbol{\mu}_t - \omega\boldsymbol{\mu}_{t-1} - \boldsymbol{\psi}_t)'\Sigma^{-1}(\boldsymbol{\mu}_t - \omega\boldsymbol{\mu}_{t-1} - \boldsymbol{\psi}_t) + \log\big(p_3(\boldsymbol{\Theta})\big).$$

Empirical Application

Our study uses county-level yearly corn yield data from NASS. The data contains 1955-2014 annual yields (bushels per acre) for Iowa, Illinois, Nebraska, Minnesota, Indiana, and 1963-2009 for Colorado. Counties with missing observations are discarded. Therefore, the final dataset includes 99 counties for Iowa, 77 for Illinois and Nebraska, 68 for Minnesota, 75 for Indiana, and 18 counties for Colorado.

The states of Iowa, Illinois, and Nebraska have been the first, second and third largest corn producers in the United States. In 2015, Iowa produced around 2.5 billion

bushels of corn, Illinois produced 2.01 billion bushels of corn, and Nebraska produced 1.7 billion bushels of corn (2015 state agriculture overview, NASS). Minnesota and Indiana are fourth and fifth largest corn producer in the United States, respectively. Colorado is the fourteenth largest corn producer in the United States. Colorado is located where the Great Plains of North America connects with the Rocky Mountains. Colorado is included because it has more varying climatic conditions than the other states.

A coordinate of the physical space consists of longitude and latitude of each county and thus the spatial smoothing from the physical space uses Euclidean distance between the locations on the physical space (i.e., physical distance). On the other hand, a coordinate of each point in the climate space is given by its climatological quantities. Therefore, spatial smoothing based on climate space uses the Euclidean distance between the locations on the climate space, which reflects climatological similarity between the locations[5]. From this climate space, we construct the spatial structure that relates the parameters of the crop yield density to the climate characteristics of the locations.

Several empirical studies have tried to determine the relationship between climate factors and crop yields. Schlenker and Roberts (2009) estimate the impact of climate change on agricultural output using panel data. They find that temperature above a threshold level has a negative impact on crop yields since it increases heat exposure and water stress. They use temperature and precipitation as the main climate variables for their research. Other related studies (Hendricks and Peterson 2012; Lobell et al. 2013; Dell, Jones, and Olken 2014) employ temperature and precipitation as explanatory variables. In accordance with the previous literature, we choose temperature and precipitation of each location as the climate coordinates.

The data for the climate space are collected from the Global Historical Climatology Network Database (GHCND) under the National Oceanic and Atmospheric Administration (NOAA). GHCND includes 18 meteorological variables including temperature (monthly means, extremes, and number of days that exceed a threshold), precipitation (total, mean, extremes), and snowfall, snow depth, and some other elements for each weather station. The climatological quantities for our climate space are a set of collected data over an extended period of time. Therefore, an average number of days in months (July and August) with maximum temperature greater than or equal to 90°F (DT90) and an average precipitation amount (mm) for the months from May to August (TPCP) from 1955 to 2014 are used as the climate coordinates. We take an average of the climate quantities from the counties with multiple weather stations. Counties with no weather stations (5 counties in all dataset) are discarded from our dataset.

In the climate space, counties with similar climate features are grouped together even when their locations are physically distant. Colorado has diverse geographical features, including mountainous terrain, vast plains, desert canyons, and mesas. For that reason, we expect that Colorado has diverse climate conditions as well so that county locations in Colorado on the two different spaces are substantially different. Figure 1 translates county locations in Colorado from the physical space (longitude/latitude) to the climate space (DT90/TPCP). The x-axis of the climate space is DT90 and the y-axis is TPCP. In contrast to Colorado, counties in corn-belt states such as Iowa tend to be grouped together in the climate space as well as in the physical space. The reason may be that the counties in Iowa have similar geographical features and thus nearly located counties have similar climate characteristics as well. Therefore, we expect that distances

among the counties in Iowa on the physical space and the climate space are closely related.

To verify the difference between these two states, we calculate correlations between the distances from two different spaces[6]. As with our expectation for the two states, the correlations between two types of distances are 0.47 in Iowa and -0.09 in Colorado. This fact may result in considerable differences in the spatial smoothing estimation from the two different spaces, and yield more substantial difference in Colorado than in Iowa.

Posterior predictive distribution

As suggested by Ozaki et al. (2008), we compute premium rates from posterior predictive values. The posterior predictive distribution $p(\boldsymbol{y}^* | \boldsymbol{Y})$ is obtained by integrating over the parameters with respect to the joint posterior distributions,

$$(12) \qquad p(\boldsymbol{y}^* | \boldsymbol{Y}) = \int_{\boldsymbol{\mu}} \int_{\boldsymbol{\Theta}} p_1(\boldsymbol{y}^* | \boldsymbol{\mu}, \boldsymbol{\Theta}) p_2(\boldsymbol{\mu}^* | \boldsymbol{\mu}, \boldsymbol{\Theta}) p(\boldsymbol{\mu}, \boldsymbol{\Theta} | \boldsymbol{Y}) d\boldsymbol{\Theta} d\boldsymbol{\mu},$$

where $p_1(\boldsymbol{y}^* | \boldsymbol{\mu}, \boldsymbol{\Theta})$ is the density of the likelihood layer, $p_2(\boldsymbol{\mu}^* | \boldsymbol{\mu}, \boldsymbol{\Theta})$ is the density of the process layer, $p(\boldsymbol{\mu}, \boldsymbol{\Theta} | \boldsymbol{Y})$ is the posterior density of the model obtained from Markov Chain Monte Carlo (MCMC) procedure[7], $\boldsymbol{\mu}^*$ and $\boldsymbol{y}^*$ are vector of posterior predicted mean parameters and crop yields, respectively. Note that $\boldsymbol{\mu}^*$ and $\boldsymbol{y}^*$ can denote either or both of a new location (county) and a new time point (year). Conceptual steps for prediction are as follows. First, vectors of random samples $\boldsymbol{\mu}$ and $\boldsymbol{\Theta}$ are drawn from the posterior density $p(\boldsymbol{\mu}, \boldsymbol{\Theta} | \boldsymbol{Y})$. Then the Bayesian spatial smoothing (Kriging) for $\boldsymbol{\mu}^*$ is applied from the process layer $p_2(\boldsymbol{\mu}^* | \boldsymbol{\mu}, \boldsymbol{\Theta})$ by updating conditional distribution for

$\boldsymbol{\mu}^*$ given the current values of $\boldsymbol{\mu}, \boldsymbol{\Theta}$ in each iteration of the MCMC algorithm. Finally, the vector of predicted yields $\boldsymbol{y}^*$ is drawn from the likelihood layer density $p_1(\boldsymbol{y}^* | \boldsymbol{\mu}, \boldsymbol{\Theta})$.

One important advantage of using the Kriging method is to get densities of counties with insufficient historical observations (i.e., missing observation) or even no observations. Since Kriging identifies a spatial structure for explaining a variation of densities across space, we can predict densities of locations with no historical observations using the Kriging parameters and posterior predictive formula in equation (12)[8].

The prediction quality of the model is evaluated by calculating the Predictive Model Choice Criteria (PMCC) suggested by Gelfand and Ghosh (1998), which is defined as,

$$PMCC = \sum_{t=1}^{T} E[(\boldsymbol{y}_t^* - \boldsymbol{y}_t)'(\boldsymbol{y}_t^* - \boldsymbol{y}_t)] + var(\boldsymbol{y}_t^*), \tag{13}$$

where $\boldsymbol{y}_t^*$ denotes a vector of predicted yield level at year t, and $\boldsymbol{y}_t$ is a vector of actual yields at year t. The first term of the PMCC represents the goodness of fit of the model, and the second term represents a penalty of model complexity. Several specifications for deterministic parts of GP and AR are tested as potential candidates for the empirical analysis. A model with lower PMCC is chosen as the preferred model. We test multiple different specifications for the process layer structure with a moving average term for historical yields (from last five years to ten years), simple linear trend, and quadratic linear trend. Among these alternatives, the model including five years of moving average and simple linear trend minimizes PMCC and thus is selected as our main model. The

main model for the GP can be expressed according to the following structure for the process layer,

$$\boldsymbol{\mu}_t = \beta_0 + \beta_1 \overline{\boldsymbol{y}}_t + \beta_2 \boldsymbol{t} + \boldsymbol{\eta} + \boldsymbol{\varepsilon}_t, \tag{14}$$

and for the AR model,

$$\boldsymbol{\mu}_t = \omega \boldsymbol{\mu}_{t-1} + \beta_0 + \beta_1 \overline{\boldsymbol{y}} + \beta_2 \boldsymbol{t} + \boldsymbol{\eta} + \boldsymbol{\varepsilon}_t, \tag{15}$$

where $\overline{\boldsymbol{y}}_t$ is a vector of current five years average of yields, $\overline{\boldsymbol{y}} = [\,\bar{y}_1, \ldots, \bar{y}_N]'$, and $\boldsymbol{t}$ is a vector of trend variable $\boldsymbol{t} = [\,t_1, \ldots, t_N]'$.

Several R-packages provide MCMC algorithms for Bayesian Kriging estimation. We mainly use spTimer, spBayes, and SpatialExtremes packages. These packages are used to estimate posterior distributions, posterior predictive distributions, and PMCC of the models. We run 20,000 iterations for MCMC chains and burn-in the first 5,000 observations to avoid an autocorrelation problem of the posterior values. We check for all parameters the graphical diagnostics of convergence using trace plots and autocorrelation plots. All posterior densities achieve fast convergence with no significant autocorrelation. To save space, Table 1 presents only the averages and standard deviations of the posterior parameter values for the four states in the dataset. Two leading states in the corn-belt area, Iowa and Illinois, show similar posterior parameter values. Both states have very small levels of temporal correlation parameter ω (0.01 for Iowa and 0.02 for Illinois) and thus the parameter of the current five year average yields (β_1) and trend parameter (β_2) differ little between the GP and the AR. In contrast, we find a notable level of temporal correlations (ω) in Nebraska and Colorado (0.37 for Nebraska and 0.57 for Colorado).

The results indicate that the two leading states in the corn-belt area (Iowa and Illinois) have a more steadily increasing trend than Nebraska and Colorado.

Out of sample performance

Providing actuarially sound premiums is an essential task to RMA. Crop producers will turn down insurance contracts when the premiums are overrated and may result in an adverse selection problem. Likewise, underrated premiums may induce insurance losses to agencies. The premium rate of crop insurance represents expected payouts as a proportion of total liability. The premium rate can be structured as

$$prem_i = \frac{P(y_i < \lambda\hat{y}_i)(\lambda\hat{y}_i - E[(y_i | y_i < \lambda\hat{y}_i)])}{\lambda\hat{y}_i} \tag{16}$$

where λ is the coverage level, $0 < \lambda < 1$, $\hat{y}_i$ is an expected crop yield at county i.

The premium rates can be estimated using the posterior predictive distributions and the premium rate formula equation (16). The posterior predictive distribution for each county can be obtained from the formula in equation (12). We then calculate the 90 percent coverage premium rates from equation (16). The expected yield $\hat{y}_i$ in county i is to be the posterior mean of the predictive distribution for each county. The premium rates differ across the alternative model specification and use of the spatial smoothing spaces. For example, each state has premium rates from AR and GP model estimated under both the physical and the climate space. Table 2 presents each state's average premium rates across counties from 2000 to 2014. One interesting result is that Colorado shows notable difference in average premium rates between using the physical space and the climate space compared to the other states. Figure 2 and 3 illustrate estimated premium rates of Iowa and Colorado from the two types of spaces, respectively. The left map presents the

premiums from the physical space, and the right map presents the premiums from climate space. In accordance with our expectation, premium estimates in Iowa from the two spaces show no substantial difference, which reflects Iowa having similar spatial structure both climate and physical distance. In contrast to Iowa, premiums in Colorado show meaningful differences in spatial structure. Specifically, we find that premiums of western region counties (Delta, Mesa, and Montrose) and southern region counties (Baca and Otero) are increased when we use climate space.

We use a loss ratio under the corresponding crop insurance program as a tool for evaluating out of sample performance of the models under the two different smoothing spaces. The loss ratio is given by

$$lossratio_i = \frac{\sum_{t=1}^{T} \max[\lambda \hat{y}_{it} - y_{it},\ 0]}{\sum_{t=1}^{T} prem_{it} \hat{y}_{it}}, \tag{17}$$

where λ is coverage level, $\hat{y}_{it}$ is a predicted yield of county i at year t, y_{it} is an actual yield for county i at year t, and $prem_{it}$ is the premium rate of county i at year t, which is obtained from equation (16).

The premium gains and indemnity losses from equation (17) are calculated using actual yields and estimated premiums of each county from 2000 to 2014. Average, variance, maximum, and minimum loss ratio across counties for the six states are presented in Table 3. The loss ratio of fairly rated crop insurance should equal one. Thus, a model with average loss ratio close to one and with a small variance across counties (i.e., regional equality) might be a preferred model. Our results demonstrate that considering the temporal correlation of crop yield (AR model) results in notable improvement in measuring the premiums in every state in our dataset. The average loss

ratios from the AR model are close to one and have smaller variance in every state compared to the GP model.

Climate space performs better or equal to physical space in every state in our dataset. In Colorado and Nebraska, the climate space conspicuously provides greater performance in out of sample prediction. Both of the average loss ratios from the climate space are closer to one than the physical space in Colorado (4.75 and 4.17) and Nebraska (1.97 and 1.67), which indicates the premiums from the climate space are more fairly rated than the physical space. Perhaps most significant is the finding that climate space smoothing resolves the serious regional inequality problem in Colorado. Our results also show that climate space has a smaller variation of loss ratio across counties. The reduction in the variance of the loss ratio of Colorado is from 36.31 to 18.80 and Nebraska is from 12.34 to 7.46. Specifically, for counties in Colorado with a high loss ratio under physical space, such as Adams and Washington counties, the loss ratio becomes closer to one under climate space smoothing. Our results demonstrate that climate space smoothing more pertinently describes spatial structure of the crop yield density compared to the physical space smoothing especially for a state like Colorado where the climate is diverse. In case of Colorado, a closeness of physical distance between locations would not be a successful factor to explain crop yield similarity across space.

Conclusion

Federal crop insurance programs have been solidified by the Agricultural Act of 2014. In this regard, providing accurate premiums for insurance contracts has become again the

utmost important role of RMA. However, the current RMA model does not fully attempt to use historical data from other areas in estimating the premium for an area of interest. Therefore, the current model has problems both in reflecting spatial correlation and retaining enough observations to properly estimate the premiums. The Bayesian Kriging model proposed here suggests a promising way to solve such problems.

There are only a few examples of Bayesian Kriging models in the agricultural economics literature, and this article is the first to use climate space smoothing. Our results show that crop yield similarity across locations tends to be more affected by climate similarity than locational closeness. By conducting climate space smoothing, we can provide a better way to measure the crop yield densities especially for a geographically diverse region such as Colorado. Future research could consider additional measures of similarity such as soil type and slope of agricultural land.

The Kriging method could be adopted in any research area that involves spatial correlation. A prominent extension of the model, for example, would be for precision agriculture. Today the availability of accurate and abundant field monitoring data allows developing a crop yield response model that parameters are smoothed by site specific agrological characteristics such as soil type, water, etc. This application could provide better site specific Variable Rate Application (VRA) fertilizer prescriptions.

One of great strengths of the Kriging model would be density estimation for the counties where historical yield data is limited or there are no observations. Bayesian Kriging method defines and describes a spatial structure for a variation of densities across space. The method allows a density prediction of any locations (counties) with no historical observations by using a posterior predictive distribution. In this regard, the

method suggests a useful way to produce density measures for counties with no yield reported.

Because we are focusing on introducing and comparing the performance of the climate space smoothing method compared to the physical distance smoothing, our model treats crop yield density in a simple manner using the normality assumption. Although the normality assumption has the advantage of simplifying the MCMC structure and could easily include trend variable into our model specification, the assumption still has a shortcoming to adjust higher moment characteristics of crop yield density such as asymmetric skewness. Therefore, while it may require developing techniques that are not yet available in the statistics literature, future research should attempt to relax this distributional assumption.

The Kriging method for climate space smoothing proposed here clearly has the potential to offer significant efficiency gains in crop yield density estimation, where historical observations is limited and has varying climate conditions.

[1] The Gaussian spatial process is a stochastic process where any finite subcollection of random variable (i.e., mean parameter of any county) is multivariate normally distributed. The covariance of those random variables between any two locations (i.e., covariance between μ_i and μ_j) is determined by a Euclidean distance between two locations and spatial covariance matrixΣ.

[2] To avoid complexity of the Markov Chain Monte Carlo (MCMC) structure, we only assume the hierarchical structure for the mean parameter. Therefore, we directly impose general inverse-gamma type prior distribution to the variance parameter of the likelihood layer σ^2.

[3] Under the Bayesian framework, we regard the mean parameter μ as a random variable. When a random vector of mean equation, $\boldsymbol{\mu} = (\mu_1, \dots \mu_N)^T$ is assumed to follow a Gaussian spatial process, the mean parameters of the counties $\mu_1, \dots \mu_N$ are jointly normally distributed.

[4] We use the empirical variogram to impose proper priors for the sill parameter ρ. For MLE estimates of mean parameter $\hat{\mu}_i$ for county location $i = 1, \dots, N$, empirical variogram can be defined as

$$\hat{\gamma}(D_{ij}) \coloneqq \frac{1}{2M} \sum_{(i,j) \in M} (\hat{\mu}_i - \hat{\mu}_j)^2$$

where M is the number of all possible pairs of counties, D_{ij} is Euclidean distance between two counties i and j, $\hat{\mu}_i$ and $\hat{\mu}_j$ are MLE estimates for the mean parameter of county i and j, respectively.

[5] In the climate space, locations (counties) that have similar climate features are closely located, even though their locations may be distant on the traditional physical space.

[6] Since our dataset includes 99 counties for Iowa and 18 counties for Colorado, all possible pairs of distances among the counties for each state are 4,851 and 153, respectively. We first obtain the distances from two different types of space (physical and climate). We then normalize these distances and calculate correlations between the two types of distance.

[7] All models in this article are fitted using a Metropolis-Hastings (MH) within Gibbs sampling algorithm. As mentioned in the prior layer section, standard conjugate priors are assumed for all coefficients (normal) and variance (inverse gamma) parameters. However, prior densities for the Kriging parameters (sill and

range) are non-standard. Hence the MH algorithm is employed to draw samples from the Kriging parameters.

[8] Kriging defines a spatial interpolation function that determines spatial correlation structure across density parameters of locations. Therefore, we can generate an interpolated density for any specific point using the Kriging parameters (sill and range) and spatial information of the point (coordinates). In this article, we do not aim to produce a specific county's density with no historical observations. However, Kriging can produce densities with no observations by using the posterior predictive distribution in equation (12).

Table II-1. Average Posterior Parameter Values for Selected States

State / Smoothing Space		Physical Space		Climate Space	
Model Structure		GP (S.D)	AR (S.D)	GP (S.D)	AR (S.D)
Iowa	β_1	1.01 (0.02)	1.02 (0.03)	1.01 (0.02)	1.02 (0.03)
	β_2	1.75 (0.02)	1.78 (0.03)	1.76 (0.02)	1.78 (0.02)
	ω	-	0.01 (0.00)	-	0.01 (0.00)
	ρ	40.35 (3.25)	40.15 (3.25)	51.31 (4.12)	50.23 (4.19)
	θ	18.23 (2.65)	18.42 (2.95)	21.83 (3.65)	22.43 (3.11)
Illinois	β_1	1.03 (0.03)	1.01 (0.03)	1.02 (0.02)	1.01 (0.02)
	β_2	1.72 (0.02)	1.71 (0.03)	1.72 (0.02)	1.72 (0.02)
	ω	-	0.02 (0.00)	-	0.02 (0.00)
	ρ	36.35 (3.95)	36.15 (3.99)	45.31 (5.22)	45.23 (5.23)
	θ	19.38 (3.26)	19.22 (3.15)	24.13 (4.82)	24.15 (5.11)
Nebraska	β_1	0.78 (0.03)	0.63 (0.03)	0.99 (0.02)	0.63 (0.03)
	β_2	2.07 (0.02)	1.34 (0.02)	2.13 (0.02)	1.34 (0.02)
	ω	-	0.37 (0.01)		0.36 (0.02)
	ρ	42.11 (6.51)	42.19 (6.52)	61.31 (4.12)	62.23 (4.19)
	θ	15.12 (1.11)	14.21 (1.12)	20.69 (2.18)	20.66 (2.26)
Colorado	β_1	0.30 (0.02)	0.14 (0.02)	0.30 (0.02)	0.15 (0.02)
	β_2	1.83 (0.05)	0.74 (0.01)	1.81 (0.04)	0.74 (0.01)
	ω	-	0.56 (0.01)	-	0.58 (0.01)
	ρ	77.25 (7.42)	78.10 (7.44)	98.31 (11.12)	98.44 (11.19)
	θ	12.13 (1.00)	12.12 (1.00)	19.13 (3.65)	19.12 (3.61)

Table II-2. Average of 90 Percent Coverage Premium Rates across Counties

State / Smoothing Space		Physical Space		Climate Space	
Model Structure		GP	AR	GP	AR
Iowa	Premium Rate (%)	1.73	1.79	1.57	1.59
Illinois	Premium Rate (%)	1.52	1.54	1.54	1.55
Nebraska	Premium Rate (%)	1.38	1.27	1.28	1.31
Minnesota	Premium Rate (%)	1.59	1.57	1.63	1.59
Indiana	Premium Rate (%)	1.65	1.69	1.64	1.75
Colorado	Premium Rate (%)	1.48	1.62	2.41	2.53

Table II-3. Estimated Loss Ratio under the Physical Space and the Climate Space

State / Smoothing Space		Physical Space		Climate Space	
Model Structure		GP	AR	GP	AR
Iowa	Mean	0.94	0.94	1.05	1.00
	Variance	1.26	1.30	1.23	0.96
	Max	5.06	4.88	4.26	3.62
	Min	0.00	0.55	0.00	0.56
Illinois	Mean	2.06	2.03	2.03	2.00
	Variance	1.08	1.10	1.07	1.01
	Max	4.28	4.35	4.26	4.32
	Min	0.00	0.00	0.00	0.13
Nebraska	Mean	2.22	1.97	2.36	1.67
	Variance	7.55	4.30	9.54	3.15
	Max	14.39	10.28	12.34	7.46
	Min	0.00	0.00	0.05	0.31
Minnesota	Mean	0.37	0.41	0.36	0.42
	Variance	0.21	0.20	0.18	0.19
	Max	1.67	1.49	1.45	1.47
	Min	0.00	0.00	0.00	0.00
Indiana	Mean	0.86	0.92	0.91	0.98
	Variance	0.58	1.74	0.36	0.12
	Max	5.36	4.89	5.34	4.98
	Min	0.05	0.09	0.06	0.23
Colorado	Mean	6.97	4.75	5.86	4.17
	Variance	59.90	26.01	36.31	18.80
	Max	23.19	15.78	17.65	13.25
	Min	0.17	0.05	0.17	0.35

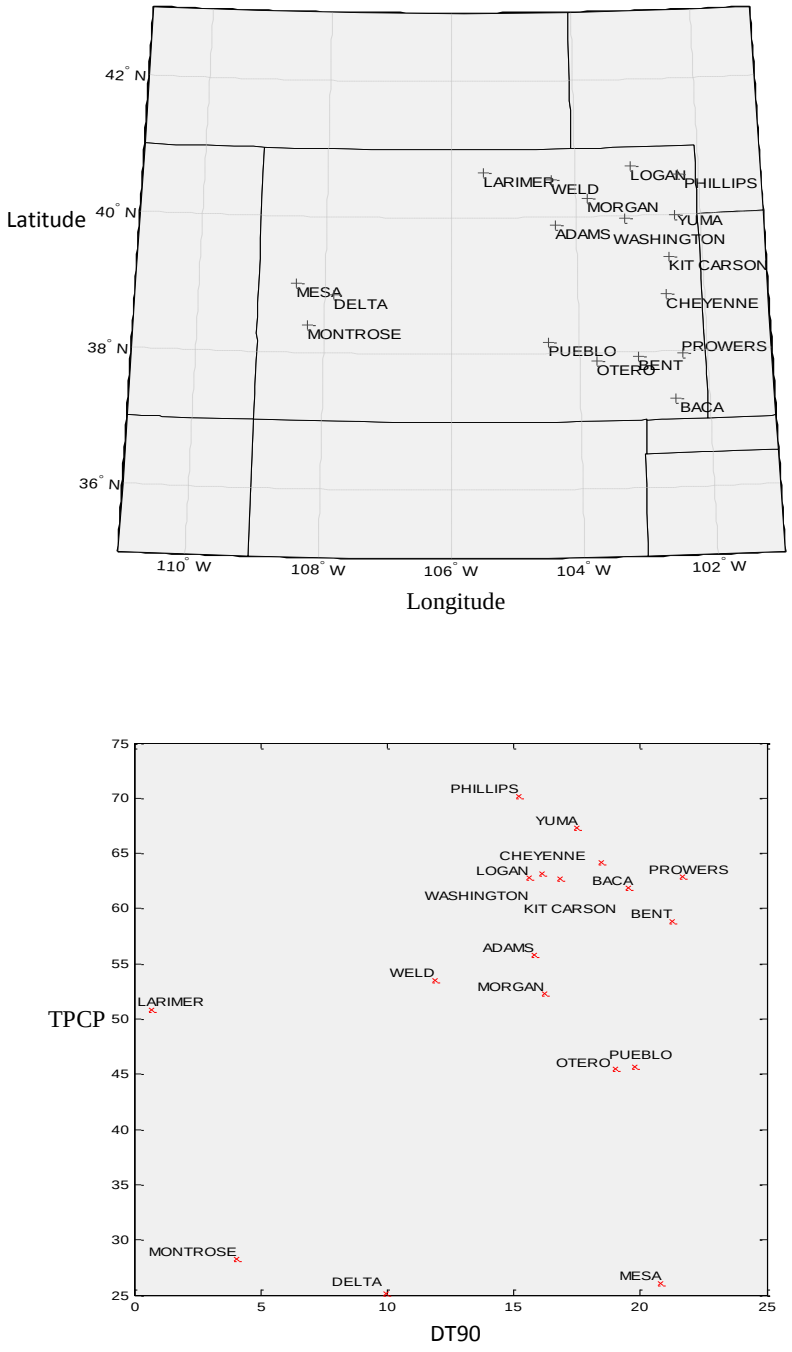

Figure II-1. Translation of counties of Colorado in physical space (above) to climate space (below)

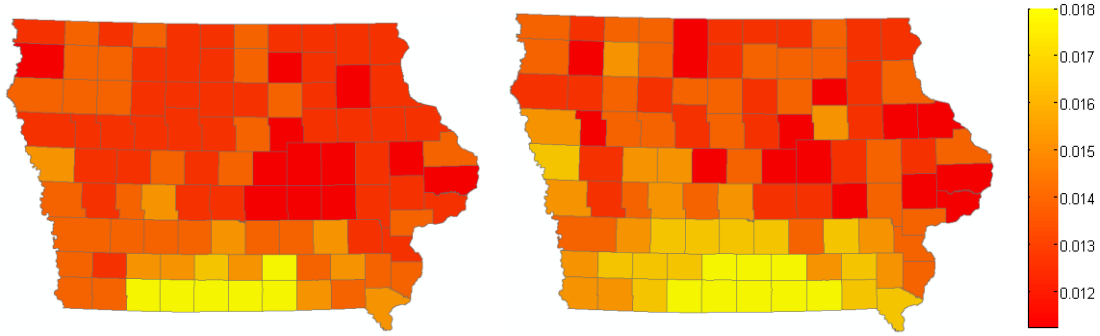

Figure II-2. Premiums of Iowa from physical space (left) and from climate space (right)

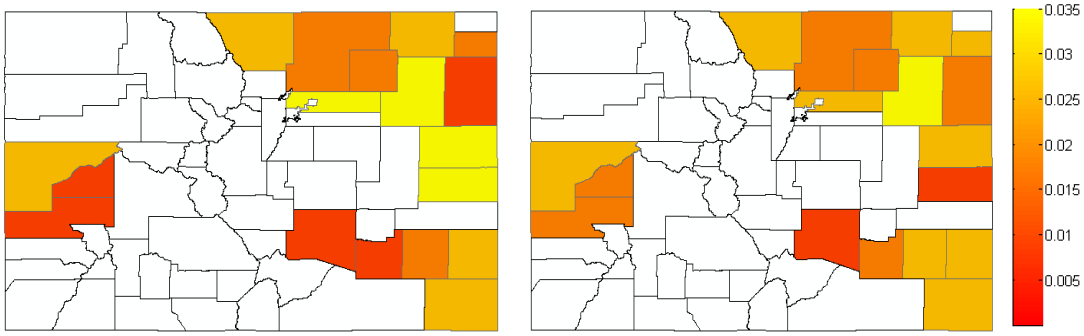

Figure II-3. Premiums of Colorado from physical space (left) and from climate space (right)

REFERENCES

Annan, F., J. Tack, A. Harri, and K.H. Coble. 2013. Spatial Pattern of Yield Distributions: Implications for Crop Insurance. *American Journal of Agricultural Economics* 96(1):253–68.

Babcock, B.A, and D.A. Hennessy. 1996. Input Demand under Yield and Revenue Insurance. *American Journal of Agricultural Economics* 78:416–27.

Banerjee, S., B. Carlin., and A. Gelfand. 2004. *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, FL: Chapman & Hall/CRC Press.

Berger, J., V. DeOiveira., and B. Sanso. 2001. Objective Bayesian Analysis of Spatially Correlated Data. *Journal of the American Statistical Association* 96: 1361–74.

Coble, K.H., T.O. Knight, R.D. Pope., and J.R. Williams. 1996. Modeling Farm-Level Crop Insurance Demand with Panel Data. *American Journal of Agricultural Economics* 78:439–47.

Coble, K.H., R. Dismukes., and J.W. Glauber. 2007. Private Crop Insurers and the Reinsurance Fund Allocation Decision. *American Journal of Agricultural Economics* 89:582–95.

Cooley, D., D. Nychka., and P. Naveau. 2007. Bayesian Spatial Modeling of Extreme Precipitation Return Levels. *Journal of the American Statistical Association* 102:824-40.

Cooley, D., P. Naveau., and P. Poncet. 2006. Variograms for Spatial Max-stable Random Fields. *Dependence in Probability and Statistics* 187:373-90.

Dell, M., B.F. Jones, and B.A. Olken. 2014. What Do We Learn from the Weather? The New Climate-Economy Literature. *Journal of Economic Literature* 52:740–98.

Goodwin, B.K., and A.P. Ker. 1998. Nonparametric Estimation of Crop Yield Distributions: Implications for Rating Group-Risk (GRP) Crop Insurance Contracts. *American Journal of Agricultural Economics* 80:139–53.

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):879–96.

Harri, A., K. H. Coble., A. P. Ker., and B. J. Goodwin, 2011. Relaxing Heteroscedasticity Assumptions in Area-Yield Crop Insurance Rating. *American Journal of Agricultural Economics* 93 (3): 707–17.

Hendricks, N. P., and J. M. Peterson. 2012. Fixed Effects Estimation of the Intensive and Extensive Margins of Irrigation Water Demand. *Journal of Agricultural and Resource Economics* 37(1): 1-19.

Ker, A. P., Tolhurst, T. N., and Liu, Y. 2015. Bayesian Estimation of Possibly Similar Yield Densities: Implications for Rating Crop Insurance Contracts. *American Journal of Agricultural Economics* 98(2):360–82

Lobell, D.B., G.L. Hammer, G. McLean, C. Messina, M.J. Roberts, and W. Schlenker. 2013. The Critical Role of Extreme Heat for Maize Production in the United States. *Nature Climate Change* 3:497–501.

Ozaki,V., B. K. Goodwin., and R. Shirota. 2008. Parametric and Nonparametric Statistical Modeling of Crop Yield: Implications for Pricing Crop Insurance Contracts. *Applied Economics* 40: 1151–64.

Ozaki, V.A., and R.S. Silva. 2009. Bayesian Ratemaking Procedure of Crop Insurance Contracts with Skewed Distribution. *Journal of Applied Statistics* 36 (4): 443–52.

Park, E., B.W Brorsen., and A. Harri., 2016. Using Bayesian Spatial Smoothing and Extreme Value Theory to Develop Area-Yield Crop Insurance Rating. Paper presented at the 2016 Agricultural and Applied Economics Association Annual Meeting in Boston, Massachusetts, July 31 – Aug 2.

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 Academy of Sciences* 106:15594–15598.

Woodard, J. D. 2016. Determining Optimal Data Aggregation: An Application of Out-of-Sample Mixture Models. Paper presented at the 2016 Agricultural and Applied Economics Association Annual Meeting in Boston, Massachusetts, July 31 – Aug 2.

CHAPTER III

MULTI-FACTOR STOCHASTIC VOLATILITY MODEL FOR PRICING CALENDAR SPREAD OPTIONS

Abstract

A calendar spread is the difference between the nearby futures price and a specified deferred futures month price on the same commodity. A calendar spread option (CSO) is an option that uses a calendar spread as its underlying asset. Previous studies have suggested CSO pricing models based on a joint process of two futures prices. However, CSO for storable commodities have distinct features due to physical aspects of the underlying assets. To address such aspects, we introduce a new CSO pricing model based on the theory of storage. Our model incorporates three stochastic processes and allows non-zero correlation structure among the processes to reflect the dynamics of the calendar spread. Option prices from the previous models are estimated to evaluate the performance of the new model. We use root mean squared error (RMSE) as tools for the evaluation. The new model proposed here outperforms the existing models in 'in the money' and 'out of the money.' The existing models tend to overestimate actual payoffs when 'at the money' and 'out of the money'.

Key words: calendar spread options, futures, stochastic volatility model, term structure effect, crude oil futures

Introduction

A calendar spread option (CSO) is an option that uses a calendar spread, which is the difference between a nearby futures price and a specified deferred futures month price on the same commodity, as an underlying asset. The Chicago Mercantile Exchange (CME) Group offers CSO on futures in storable commodities such as corn, wheat, soybean, soybean oil, soybean meal, cotton, and crude oil. Unlike a payoff of a standard vanilla option that depends only on the dynamics of one underlying futures contract, a payoff of CSO is calculated from the price dynamics of the two futures contracts. Therefore, pricing CSO must consider the joint stochastic processes of the two futures contracts.

Several models for pricing CSO have been suggested. Gibson and Schwartz (1990) introduce a two-factor model taking account of stochastic convenience yield. They assume a mean reverting convenience yield. Shimko (1994) suggests a closed-form approximation of the futures spread option model, based on the framework of Gibson and Schwartz. Poitras (1998) proposes a model that futures prices have a joint normal distribution. Hinz and Fehr (2010) suggest a model that assumes the calendar spread follows a shifted lognormal distribution. Seok et al. (2014) suggest a CSO pricing model for storable commodities that assumes convenience yield follows arithmetic Brownian motion. However, they use Monte Carlo to price the CSO and do not use a calibration procedure to obtain parameters of the solution. More recently, Schneider and Tavin (2016) suggest a model for pricing CSO based on the joint characteristic function of two futures. They provide an analytical solution from the joint characteristic process.

There are some important empirical features of commodity futures markets. First is non-constant (stochastic) volatility of futures contract, which is well known as term

structure (Samuelson effect) of volatility; a futures price tends to be more volatile as it approaches maturity. The non-constant volatility may also be observed as a volatility smile; implied volatilities tend to increase as strike prices get away from a current spot price. A second feature is that futures with different maturities are not perfectly correlated, but tend to move together. To address these features, the paper employs traditional theory of storage (Working 1949; Brennan 1958) to reflect these features of the commodity futures market in pricing the CSO for storable commodities. From the theory, a calendar spread can be expressed as a function of nearby futures, storage cost, interest cost, and convenience yield. We then define processes for the components of the calendar spread and suggest a formula for pricing CSO. Our model includes multi-stochastic factors based on the stochastic volatility framework proposed by Heston (1993) and Christoffersen et al. (2009). The model allows correlations among the stochastic factors to reflect empirical features of the commodity market. The model also provides an analytical solution to price CSO. We demonstrate that our model is more flexible than other CSO pricing models in reflecting the term structure effect and smile or smirk of the implied volatility. We expect that a more precise pricing model for CSO would allow market makers to offer lower bid-ask spreads and thus increase market volume of CSO.

We use the empirical market data for West Texas Intermediate (WTI) to calibrate the parameters of the analytical solution. Daily closing prices for 1-month CSO from Jan 4, 2016 to Feb 27, 2017 are obtained from the Datamine of CME group. T-bill rate from Federal Reserve Bank Report is used as the risk-free rate of the model. Using the dataset, we obtain the set of parameters that minimizes a relative difference between market value

of CSO and the theoretical value of CSO from the model. The new model with three stochastic factors outperforms existing models in out of sample performance especially when the strike of the option is placed ‘deep in the money’ or ‘deep out of money’. The results might be due to non-zero correlation and stochastic factors in the model that provides flexibility in controlling term structure effects in implied volatility. Gibson and Schwartz (1990) model shows the least preferred performance over the models. It relatively performs well in ‘in the money’ pricing, but overestimates actual payoffs in ‘at the money’ and ‘out of the money.'

The rest of the article proceeds as follows. In the theoretical framework section, we define the theory of storage and describe stochastic processes for each factor that composes the calendar spread. In the next section, we describe CSO structure and provide an analytical solution of CSO. In the calibration section, we explain the dataset we use and explain the calibration procedure. In the empirical analysis, we compare the performance of the new model with existing models using root mean squared error (RMSE).

Theoretical framework

The spread between futures and spot prices has been explained using the theory of storage. Following the theory, the spread between the futures and spot price is a function of the interest cost, storage cost, and convenience yield. The traditional theory of Kaldor (1939) and Brennan (1958) argues that net marginal cost of storage includes the financial opportunity cost of capital, which is interest cost, and direct warehousing or storage costs, and convenience yield earned from holding inventories. The benefits reflected in

convenience yield arise from the flexibility in stock uses since the company can provide immediately when the cost in production increases. Therefore, the benefits of holding inventories have been discussed in many articles and play a crucial role in explaining normal and inverted markets. Following the traditional framework, the spread between futures and spot price can be expressed as

(1) $$F(t,T) - S(t) = r(t,T-t)S(t) + W(t,T-t) - Y(T-t)$$

where $F(t,T)$ is futures price at time t for maturity T, $S(t)$ is spot price at current time t, $r(t,T-t)$ is interest rate, $W(t,T-t)$ is storage costs, and $Y(t,T-t)$ is convenience yield.

The spot price of the commodity is not readily observable. Therefore, several studies support using the futures price as a proxy for the spot price (Brennan 1958; Gibson and Schwartz 1990; Schwartz 1997; Hinz and Fehr 2010). In this regard, we replace nearby futures price as a proxy for the spot price, and the equation is changed into

(2) $$F_2(t,T_2) - F_1(t,T_1) = r(T_2 - T_1)F_1(t,\ T_1) + W(T_2 - T_1) - Y(t,T_2 - T_1)$$

where $F_1(t,T_1)$ is nearby futures price at t with T_1 maturity and $F_2(t,T_2)$ is distant futures price at t with T_2 maturity. Following the equation (2), the spread between futures and spot price is interest cost plus storage cost minus sum of convenience yield that includes the price of the risk arises from the stochastic volatility. Since we change the spot price to nearby futures price in equation (1) and (2), we can apply the formula to estimate the value of CSO. If the calendar spread put option is exercised, at the expiration date, a buyer of the option will receive a short position in the nearby futures and a long position in the distant futures. Therefore, the value of the European calendar spread put option with exercise price K is

(3) $$P_{cs} = \max(K - [F_1(T,T_1) - F_2(T,T_2)],\ 0) \qquad T \le T_1 \le T_2,$$

where T is maturity date of the European calendar spread put option P_{cs}.

We assume that both interest rate (r) and storage costs (W) are constant. Thus, we can only consider nearby futures price and convenience yield to derive the payoff of the calendar spread call option. By substituting equation (2) into (3), equation (3), the value of the European calendar spread put option P_{cs}, can be replaced with a bear spread valuation problem where a long call option of nearby futures (C_A) and a short call option of convenience yield with zero strike (C_B). Then equation (3) is changed into

(4) $$\begin{aligned} P_{cs} &= C_A - C_B, \\ P_{cs} &= \max\big(F(t,T) - \widetilde{K},\ 0\big) - \max(Y(t,T),\ 0), \end{aligned}$$

where $F(t,T) = r(t,T)F_1(t,T) + W(t,T)$ and $\widetilde{K} = -K$.

Long call option solution

Christoffersen et al. (2009) extend the Heston (1993) model to a multi-stochastic volatility factor setting. They show that the model can provide more flexible modeling of the volatility term structure than the Heston (1993) model. Therefore, it has a broader capacity to capture the empirical dynamics of calendar spread. Our model is based on the model introduced by Christoffersen et al. (2009). Unlike the previous model, we assume that convenience yield follows arithmetic Brownian motion. The model is structured as follows,

(5) $$\begin{aligned} dF(t) &= \mu F(t)dt + \sqrt{v(t)}F(t)dZ_1(t) + \sigma_1 F(t)dZ_2(t) \\ dv(t) &= k[\theta - v(t)]dt + \delta\sqrt{v(t)}dZ_3(t) \\ dY(t) &= \alpha dt + \sigma_2 dZ_4(t) \end{aligned}$$

$$dZ_1(t)dZ_3(t) = \rho_1 dt$$

$$dZ_2(t)dZ_4(t) = \rho_2 dt,$$

where $F(t)$ is nearby futures price introduced in equation (4) and is assumed to follow geometric Brownian motion with two stochastic volatility factors, the volatility factor $v(t)$ is assumed to follow Cox-Ingersoll-Ross (CIR) process, convenience yield $Y(t)$ follows arithmetic Brownian motion, and $Z_1(t)$, $Z_2(t)$, $Z_3(t)$ and $Z_4(t)$ are standard Wiener process. The model assumes that $Z_1(t)$ has correlation ρ_1 with $Z_3(t)$, and $Z_2(t)$ has correlation ρ_2 with $Z_4(t)$. However, $Z_3(t)$ and $Z_4(t)$, and $Z_1(t)$ and $Z_2(t)$ are not correlated. Therefore, the variance of the futures price is the sum of two uncorrelated factors, which are general volatility factor and convenience yield, and these factors are individually correlated with the futures price.

Our model structure allows dynamics of the calendar spread (i.e., asymmetry or term structure effect). The volatility process $v(t)$ controls smile or smirk of implied volatility. Also, the correlation parameters ρ_1 and ρ_2 capture non-zero correlations between return of $F(t)$ and volatility $v(t)$ and convenience yield $Y(t)$. An advantage of the model is that the model has an analytical solution. To simplify the notation, we change the stochastic processes in equation (5) into

$$dF = \mu_F dt + \sigma_{F1} dz_1 + \sigma_{F2} dz_2$$

$$dv = \mu_v dt + \sigma_v dz_3$$

$$dY = \mu_Y dt + \sigma_Y dz_4 \quad (6)$$

$$dz_1(t)dz_3(t) = \rho_1 dt$$

$$dz_2(t)dz_4(t) = \rho_2 dt.$$

Expanding the first long call option C_A using Ito's lemma gives,

$$dC_A = \left[\frac{\partial C_A}{\partial t} + \mu_F \frac{\partial C_A}{\partial F} + \mu_v \frac{\partial C_A}{\partial v} + \mu_Y \frac{\partial C_A}{\partial Y} + \frac{1}{2}(\sigma_{F1}^2 + \sigma_{F2}^2)\frac{\partial^2 C_A}{\partial F^2} + \frac{1}{2}\sigma_v^2 \frac{\partial^2 C_A}{\partial v^2} + \right.$$

(7) $$\left. \frac{1}{2}\sigma_Y^2 \frac{\partial^2 C_A}{\partial Y^2} + \rho_1 \sigma_{F1}\sigma_v \frac{\partial^2 C_A}{\partial F \partial v} + \rho_2 \sigma_{F2}\sigma_Y \frac{\partial^2 C_A}{\partial F \partial Y}\right] dt + \sigma_{F1}\frac{\partial C_A}{\partial F} dz_1 +$$

$$\sigma_{F2}\frac{\partial C_A}{\partial F} dz_2 + \sigma_v \frac{\partial C_A}{\partial v} dz_3 + \sigma_Y \frac{\partial C_A}{\partial Y} dz_4.$$

Under the complete market assumption of risk neutral measure, the form of the dynamic portfolio $M = C_A - \xi F - \psi U$, which has option $C_A = C_A(F, v, Y, t)$, and ξ units of the futures F, and ψ units of another option $U = U(F, v, Y, t)$, can be used for perfect hedge. Applying Ito's lemma to dM, the dynamic of the portfolio value M can be expanded in a similar manner to equation (7), and now we have,

(8) $$dM = dC_A - \xi dF - \psi dU.$$

To obtain risk neutrality, the coefficients of dz_1, dz_2, dz_3 and dz_4in portfolio expansion equation (8) must be zero. This implies the following two equations,

(9) $$\psi = \frac{\partial C_A/\partial v}{\partial U/\partial v} = \frac{\partial C_A/\partial Y}{\partial U/\partial Y} = \frac{\partial C_A}{\partial U}$$

(10) $$\xi = \frac{\partial C_A}{\partial F} - \psi \frac{\partial U}{\partial F}.$$

Under the risk neutrality, the instantaneous change of a value of the portfolio M should be equal to the return on a risk-free investment. Otherwise, there will be an arbitrage opportunity. Since the portfolio M must earn the same amount of risk-free rate r, we have

(11) $$dM = rMdt = r[dC_A - \xi dF - \psi dU]dt.$$

When we equate equation (8) and (11), and replace ψ and ξ from equation (9) and (10) gives a function that represents price of volatility risk (drift term of volatility process)

suggested by Heston (1993), which is assumed to be a linear function of v such that $\mu_v = f(F,v,t) = k[\theta - v] - \lambda(F,v,t) = k[\theta - v] - \lambda v$. Substitute $\mu_v = f(F,v,t)$ equation (8) and (11) gives the fundamental partial differential equation for the option is

$$
\begin{aligned}
& rF\frac{\partial C_A}{\partial F} + \frac{1}{2}(v+\sigma_1^2)F^2\frac{\partial^2 C_A}{\partial F^2} + \{k[\theta - v] - \lambda v\}\frac{\partial C_A}{\partial v} + \frac{1}{2}v\delta^2\frac{\partial^2 C_A}{\partial v^2} \\
& + \alpha\frac{\partial C_A}{\partial Y} + \frac{1}{2}\sigma_2^2\frac{\partial^2 C_A}{\partial Y^2} + \rho_1\delta v F\frac{\partial^2 C_A}{\partial F\partial v} + \rho_2\sigma_1\sigma_2 F\frac{\partial^2 C_A}{\partial F\partial Y} \\
& + \frac{\partial C_A}{\partial t} - rC_A = 0.
\end{aligned} \tag{12}
$$

To simplify the equation, assume that $x = \ln F$, then the partial derivatives are $\frac{\partial C_A}{\partial F} = e^{-x}\frac{\partial C_A}{\partial x}$, $\frac{\partial^2 C_A}{\partial F^2} = \frac{1}{F^2}\frac{\partial^2 C_A}{\partial x^2} - \frac{1}{F^2}\frac{\partial C_A}{\partial x}$, $\frac{\partial^2 C_A}{\partial F\partial v} = \frac{1}{F}\frac{\partial^2 C_A}{\partial x\partial v}$, and $\frac{\partial^2 C_A}{\partial F\partial Y} = \frac{\partial}{\partial x}\left(\frac{\partial C_A}{\partial Y}\right)\frac{\partial x}{\partial F} = \frac{1}{F}\frac{\partial^2 C_A}{\partial x\partial Y}$. Substituting the results into equation (12) gives

$$
\begin{aligned}
& \left(r - \frac{1}{2}(v+\sigma_1^2)\right)\frac{\partial C_A}{\partial x} + \{k[\theta - v] - \lambda v\}\frac{\partial C_A}{\partial v} + \alpha\frac{\partial C_A}{\partial Y} + \frac{1}{2}(v+\sigma_1^2)\frac{\partial^2 C_A}{\partial x^2} \\
& + \frac{1}{2}\delta^2 v\frac{\partial^2 C_A}{\partial v^2} + \frac{1}{2}\sigma_2^2\frac{\partial^2 C_A}{\partial Y^2} + \rho_1\delta v\frac{\partial^2 C_A}{\partial x\partial v} + \rho_2\sigma_1\sigma_2\frac{\partial^2 C_A}{\partial x\partial Y} \\
& + \frac{\partial C_A}{\partial t} - rC_A = 0.
\end{aligned} \tag{13}
$$

The difference between equation (12) and (13) is that the coefficient of the partial derivatives does not contain x. Therefore, now we can more easily solve the partial differential equation.

The European call option has the following three boundary conditions $C_A(F_T,v,Y,t) = \max(F_T - \widetilde{K},0)$, $C_A(0,v,Y,t) = 0$, and $\frac{\partial C_A}{\partial F_T}C_A(\infty,v,Y,t) = 1$. By

analogy with the Black-Scholes formula, the solution for the European option has the form

$$C_A(F_T, v, Y, t) = FP_1 - e^{-r(T-t)}\widetilde{K}P_2$$

(14)

$$C_A(x, v, Y, t) = e^x P_1 - e^{-r\tau}\widetilde{K}P_2$$

where P_1 and P_2 are cumulative density functions about the moneyness of the option at maturity, and $\tau = T - t$. From equation (14), we obtain the following partial derivatives such as $\frac{\partial C_A}{\partial x} = e^x P_1 + e^x \frac{\partial P_1}{\partial x} - e^{-r\tau}\widetilde{K}\frac{\partial P_2}{\partial x}$, and resubstitute these partial derivatives into equation (14) to give

(15) $$e^x[h(P_1)] - e^{-r\tau}\widetilde{K}[h(P_2)] = 0$$

where $h(P_j)$ is a polynomial function of P_j.

Then the solution for the equation (15) is $h(P_1) = h(P_2) = 0$, where

(16)
$$\begin{aligned} h(P_1) = &\left(r + \frac{1}{2}(v + \sigma_1^2)\right)\frac{\partial P_1}{\partial x} + \frac{1}{2}(v + \sigma_1^2)\frac{\partial^2 P_1}{\partial x^2} \\ &+ (k\theta - v(k + \lambda - \rho_1\delta))\frac{\partial P_1}{\partial v} + \frac{1}{2}\delta^2 v\frac{\partial^2 P_1}{\partial v^2} + (\alpha - Y\rho_2\sigma_2)\frac{\partial P_1}{\partial Y} \\ &+ \frac{1}{2}\sigma_2^2\frac{\partial^2 P_1}{\partial Y^2} + \rho_1\delta v\frac{\partial^2 P_1}{\partial x\partial v} + \rho_2\sigma_1\sigma_2\frac{\partial^2 P_1}{\partial x\partial Y} + \frac{\partial P_1}{\partial t} = 0 \end{aligned}$$

(17)
$$\begin{aligned} h(P_2) = &\left(r - \frac{1}{2}(v + \sigma_1^2)\right)\frac{\partial P_2}{\partial x} + \frac{1}{2}(v + \sigma_1^2)\frac{\partial^2 P_2}{\partial x^2} + (k\theta - v(k + \lambda))\frac{\partial P_2}{\partial v} \\ &+ \frac{1}{2}\delta^2 v\frac{\partial^2 P_2}{\partial v^2} + \alpha\frac{\partial P_2}{\partial Y} + \frac{1}{2}\sigma_2^2\frac{\partial^2 P_2}{\partial Y^2} + \rho_1\delta v\frac{\partial^2 P_2}{\partial x\partial v} + \rho_2\sigma_1\sigma_2\frac{\partial^2 P_2}{\partial x\partial Y} \\ &+ \frac{\partial P_2}{\partial t} = 0. \end{aligned}$$

However, the cumulative probability function P_j cannot be immediately solved in closed form. According to Heston (1993) and Christoffersen et al (2009), by alternatively using the Feynman-Kac theorem, it can be shown that its characteristic function $f_j(x, v, Y, T\,; \phi)$ must satisfy the same PDEs subject to the terminal condition $f_j(x, v, Y, \tau\,; \phi) = e^{i\phi x}$.

Assuming that the characteristic function has the solution

$$f_j(x, v, Y, \tau\,; \phi) = e^{C+Dv+GY+i\phi x} \tag{18}$$

where C, D, and G are the function of τ. Therefore, the characteristic function should satisfy equations (16) and (17), and from $f_j(x, v, Y, \tau\,; \phi) = e^{i\phi x}$, we get partial derivatives such as $\frac{\partial f}{\partial x} = i\phi f$ and $\frac{\partial f}{\partial \tau} = \left(\frac{\partial C}{\partial \tau} + \frac{\partial D}{\partial \tau} v + \frac{\partial G}{\partial \tau} Y\right) f$, and resubstitute these partial derivatives into equation (16) and (17), we have

$$\begin{aligned}\left[\tfrac{1}{2}\delta^2 D^2 + \left(\rho_1 \delta i\phi - b_j\right)D + u_j i\phi - \tfrac{1}{2}\phi^2 + \tfrac{\partial D}{\partial \tau}\right] v + \left[\tfrac{1}{2}\sigma_2^2 G^2 + \left(\rho_2 \sigma_2 i\phi - \right.\right. \\ \left.\left. c_j\right)D + u_j i\phi - \tfrac{1}{2}\phi^2 + \tfrac{\partial G}{\partial \tau}\right] Y + \left[ri\phi + k\theta D + \alpha G + \tfrac{\partial C}{\partial \tau}\right] = 0,\end{aligned} \tag{19}$$

for $j = 1, 2$, where $b_1 = k + \lambda - \rho_1 \delta$, $b_2 = k + \lambda$, $c_1 = -\rho_2 \sigma_2$, $c_2 = 0$, $u_1 = \frac{1}{2}$, $u_2 = -\frac{1}{2}$. This gives rise to a system for P_1 and P_2, and each function has three Riccati equations :

$$\begin{aligned}\frac{1}{2}\delta^2 D^2 + \left(\rho_1 \delta i\phi u_i - b_j\right)D + u_j i\phi - \frac{1}{2}\phi^2 + \frac{\partial D}{\partial \tau} = 0 \\ \frac{1}{2}{\sigma_2}^2 G^2 + \left(\rho_2 \sigma_2 i\phi u_i - c_j\right)G + u_j i\phi - \frac{1}{2}\phi^2 + \frac{\partial G}{\partial \tau} = 0 \\ ri\phi + k\theta D + \alpha G + \frac{\partial C}{\partial \tau} = 0\end{aligned} \tag{20}$$

subject to the initial value $C(0;\phi) = D(0;\phi) = G(0;\phi) = 0$. Now we can solve the Riccati equation with the initial value above. The solution is

$$C(\tau;\phi) = ri\phi\tau + \frac{k\theta}{\delta^2}\left[(b_j - \rho_1\delta i\phi + d_1)\tau - 2\ln\left(\frac{1-g_1e^{d_1\tau}}{1-g_1}\right)\right] + \frac{\alpha}{\sigma_2^2}\left[(c_j - \rho_2\sigma_2 i\phi + d_2)\tau - 2\ln\left(\frac{1-g_2e^{d_2\tau}}{1-g_2}\right)\right] \quad (21)$$

$$D(\tau;\phi) = \frac{b_j - \rho_1\delta i\phi + d_1}{\delta^2}\left[\frac{1-e^{d_1\tau}}{1-g_1e^{d_1\tau}}\right] \quad (22)$$

$$G(\tau;\phi) = \frac{c_j - \rho_2\sigma_2 i\phi + d_2}{\sigma_2^2}\left[\frac{1-e^{d_2\tau}}{1-g_2e^{d_2\tau}}\right] \quad (23)$$

where $g_1 = \frac{b_j-\rho_1\delta i\phi+d_1}{b_j-\rho_1\delta i\phi-d_1}$, $g_2 = \frac{c_j-\rho_2\sigma_2 i\phi+d_2}{c_j-\rho_2\sigma_2 i\phi-d_2}$, $d_1 = \sqrt{(\rho_1\delta i\phi - b_j)^2 - \delta^2(2u_i i\phi - \phi^2)}$, and $d_2 = \sqrt{(\rho_2\sigma_2 i\phi - c_j)^2 - {\sigma_2}^2(2u_i i\phi - \phi^2)}$.

The solution forms of the terms $D(\tau;\phi)$ and $G(\tau;\phi)$ are identical to the one-dimensional counterpart in Heston (1993), and $C(\tau;\phi)$is also identical to the sum of two terms of one-dimensional Heston (1993)'s solution plus $ri\phi\tau$.

From the Fourier inversion theorem, we have

$$P_1 = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty Re\left[\frac{e^{i\phi\ln\left(\frac{F(t)}{\tilde{K}}\right)f(v,Y,\tau,\phi+1)}}{i\phi F(t)e^{r\tau}}\right]d\phi, \quad (24)$$

$$P_2 = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty Re\left[\frac{e^{i\phi\ln\left(\frac{F(t)}{K}\right)f(v,Y,\tau,\phi)}}{i\phi}\right]d\phi. \quad (25)$$

By employing option pricing formula $C(F_T, v, Y, t) = FP_1 - e^{-r\tau}\tilde{K}P_2$, we have a solution for the call option,

$$C_A(F_T, v, Y, t) = \frac{1}{2}(F(t) - \tilde{K}e^{-r\tau})$$

(26)
$$+\frac{e^{-r\tau}}{\pi}\int_0^{\infty} Re\left[\frac{e^{i\phi \ln\left(\frac{F(t)}{\tilde{K}}\right)f(\phi+1)}}{i\phi} - \frac{e^{i\phi \ln\left(\frac{F(t)}{\tilde{K}}\right)f(\phi)}}{i\phi}\right]d\phi.$$

Short call option solution

The second part of the calendar spread put option (P_{CS}) in equation (4) is a short call option that has convenience yield (Y) as its underlying asset with zero strike price C_B. We assume that the convenience yield follows arithmetic Brownian motion with mean zero ($\alpha = 0$) and variance $\sigma_2^2\tau$. Then the value of a European call option is

(27)
$$C_B = \max(Y(t,T) - X,\ \ 0)$$

where X is the strike price of the option. According to the framework of Bachelier (1990), the second call option value for the convenience yield with zero strike price is

(28)
$$C_B = e^{-r\tau}\left[Y(t)\Phi\left(\frac{Y(t)}{\sigma_2\sqrt{\tau}}\right) + \sigma_2\sqrt{\tau}\phi\left(\frac{Y(t)}{\sigma_2\sqrt{\tau}}\right)\right]$$

where Φ and ϕ are standard normal cumulative density and probability density, respectively.

A solution of CSO

We already defined two call option values (C_A and C_B) that compose calendar spread put options of the model. Hence the complete analytical solution for the calendar spread put option P_{cs} in equation (4) can be obtained by combining the two call option values,

(29)
$$P_{cs} = C_A - C_B$$

$$= \frac{1}{2}\left(F(t) - \widetilde{K}e^{-r\tau}\right)$$

$$+ \frac{e^{-r\tau}}{\pi}\int_0^{\infty} Re\left[\frac{e^{i\phi \ln\left(\frac{F(t)}{\widetilde{K}}\right) f(\phi+1)}}{i\phi} - \frac{e^{i\phi \ln\left(\frac{F(t)}{\widetilde{K}}\right) f(\phi)}}{i\phi}\right] d\phi$$

$$- e^{-r\tau}\left[Y(t)\Phi\left(\frac{Y(t)}{\sigma_2\sqrt{\tau}}\right) + \sigma_2\sqrt{\tau}\phi\left(\frac{Y(t)}{\sigma_2\sqrt{\tau}}\right)\right].$$

Note that calendar spread call option value can be obtained using put-call parity. We use the analytical solution to calibrate parameters and compare our values with values from existing methods.

Empirical Framework

In this section, we use empirical data for crude oil (WTI) to calibrate parameters of the model. Our dataset includes daily New York Mercantile Exchange (NYMEX) CSO closing prices, WTI futures prices, T-bill rate for interest rate, and physical storage costs. The daily NYMEX 1-month CSO closing prices from Jan 4, 2016 to Feb 27, 2017 are obtained from the Datamine of CME group. We only use closing prices that are actually traded. Therefore, options with zero volume are discarded from the dataset. We use the 1-month CSO price (where $T_2 - T_1 = 1$ month) for calibration since it has good liquidity compared to 2, 3, 6, and 12-month CSOs. The calibrated parameters are a set of parameters that minimizes difference between market CSO values and the theoretical CSO values. To proceed with the calibration, we should additionally calculate a current convenience yield to get the second call option value (C_B), which is a part of an analytical solution for CSO. Based on the theory of storage in equation (2), the implicit convenience yield $Y(t)$ can be obtained by adding spread ($F_2(t, T_2) - F_1(t, T_1)$), interest costs ($r(T_2 -$

$T_1)F_1(t, T_1)$), and physical storage cost ($W(T_2 - T_1)$). T-bill rate from Federal Reserve Bank Report is used as a risk-free rate to determine the interest costs. Storage costs are difficult to obtain. We employ a monthly spot freight rate (for very large crude carriers, or VLCCs, steaming from the Middle East Gulf to the West) from Containerisation International (CI) as a proxy for storage costs since offshore storage in carriers would be a form of holding stocks. The spot freight rates are interpolated to daily values.

Calibration procedure

One difficulty of applying our new model is that there are unknown structural parameters which cannot be observed from market data. Several types of research have been devoted to finding a way of obtaining those parameters, yet there is no consensus. In this section, we introduce a numerical calibration procedure for our proposed model.

One possible challenge when estimating the parameters of the multi-factor stochastic model is that we must jointly estimate the structural parameters (k, θ, δ, σ_1, σ_2, ρ_1, ρ_2) with the spot volatility v. Several methods were introduced in previous literature. One general approach regards the spot volatility v as another parameter and reestimates the value daily(Bakshi et al. 1997). Other methods filter the volatility value from the empirical market data of underlying returns. This approach can be done using a Kalman filter (Carr and Wu 2007), Bayesian estimation (Jones 2003; Eraker 2004), and an efficient method of moments approach (Pan 2002).

In this article, we employ an iterative two-step calibration procedure suggested by Huang and Wu (2004) and Christoffersen et al. (2009) where the structural parameters and spot volatility are estimated using empirical CSO data. Let V_i^m denote the *i'*th market

value of CSO, and $V_i^*(\tau, F, r, Y, K; \Theta, \hat{v})$ denote the theoretical CSO price of the new model. Consider a sample of N CSOs. In the first step, the sum of squared errors between theoretical values and market CSO values given a set of structural parameters Θ will be minimized under the spot process values $\hat{v}$,

$$\hat{v} = \underset{v}{\operatorname{argmin}} \sum_{i=1}^{N} \left[\frac{V_i^m - V_i^*(\tau, F, r, K; \Theta, v)}{Vega_i} \right]^2, \quad i = 1, \dots, N. \tag{30}$$

where $Vega_i$ denotes vega of the option i computed using the implied volatility from the Black-Scholes formula and market CSO values.

In the second step, for given spot volatility values $\hat{v}$ obtained from equation (30), the sum of squared errors between theoretical values and market CSO values is minimized under the set of structural parameters $\widehat{\Theta}$,

$$\widehat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \sum_{i=1}^{N} \left[\frac{V_i^m - V_i^*(\tau, F, r, K; \Theta, \hat{v})}{Vega_i} \right]^2, \quad i = 1, \dots, N. \tag{31}$$

The iteration procedure between equation (30) and (31) is conducted until convergence is achieved in the objective function value in equation (31).

The overall calibration procedure proceeds as follows. First, risk free rates and physical storage rates are interpolated to daily values. Next, the implicit convenience yield is calculated using the theory of storage formula in equation (2). Then using equations (30) and (31), the set of structural parameters $\widehat{\Theta}$ and the spot volatility $\hat{v}$ that minimizes sum of squared errors is obtained. The daily 1-month CSO closing prices from Jan 4, 2016 to Feb 27, 2017 are used for the calibration. Since we discard zero volume

options from our dataset, the total number of market CSO prices N for the calibration is 1,901. MATLAB function 'lsqnonlin' is used for estimation.

Calibration of the model can be sensitive to the choice of the starting point. Therefore, we impose zero starting values for speed of reversion k and long run average θ in our first step of iteration in equation (3) to avoid problems associated with starting value sensitivity. Further, we find the calibration is sensitive to the volatility for convenience yield σ_2. Hence, we impose a starting value $\sigma_2 = 0.3257$, which is calculated from the implicit convenience yield under the assumption of arithmetic Brownian motion using the market data from January 4, 2016 to Feb 27, 2017. Likewise, we use empirical data to set the starting value for v using the same dataset with σ_2 (v=0.4661). The two correlation parameters ρ_1 and ρ_2 are bounded from -1 to 1, and all volatility parameters (θ, σ_1, σ_2, δ) and speed of adjustment parameter (k) have lower bound of zero. Table 1 presents the values of obtained parameters $\widehat{\Theta}$. The results show that the correlation parameters ρ_1 is positive, whereas ρ_2 is negative. Therefore, we find that when the nearby futures price increases, the volatility process becomes more volatile. On the other hand, negative correlation parameter ρ_2 induces that when the futures price increases, the convenience yield process becomes less volatile.

Prediction Tests

Two different previous models (Gibson and Schwarz 1990; Poitras 1998) are considered to compare the performance of the new model. The dataset from January 2, 1985 to Feb 27, 2017 is used for obtaining parameters of each model. The dataset is collected from

Energy Information Administration (EIA). Table 2 summarizes statistics for each calendar spread, implicit convenience yield, interest rates, and storage cost obtained.

First, Gibson and Schwarz (1990) assume the spot price follows geometric Brownian motion and the convenience yield follows a mean-reverting process. We replace spot price to the nearby futures price. Then the two joint stochastic processes for the model can be defined as

$$dF(t) = \mu F(t)dt + \sigma_{s1}F(t)dZ_1(t)$$

(32)
$$dY(t) = a[b - Y(t)]dt + \sigma_{s2}dZ_2(t)$$

$$dZ_1(t)dZ_2(t) = \rho.$$

All structural parameters of the model including the nearby futures drift μ, two volatility parameters σ_{s1} and σ_{s2}, and the mean reverting process parameters k and θ are calibrated from the same procedure and dataset as our propsed model. Since Heston (1993) provides an analytical solution for single stochastic volatility type model we have defined here for Gibson and Schwarz model, the solution is replaced into equation (30) and (31) to obtain the parameters. From the two equations for calibration, all structural parameters are obtained. Two parameters of the mean reverting process $a = 0.6895$ and $b = 0.0157$. All the other parameters are $\mu = 0.08$, $\sigma_{s1} = 0.23$, $\sigma_{s2} = 0.35$, and $\rho = -0.29$. We then calculate the value of the call option values from the analytical solution.

Poitras (1998) suggests a calendar spread option pricing model where two futures prices (F_1, F_2) follow arithmetic Brownian motion such that

(33)
$$dF_1(t) = \mu_1 dt + \sigma_{p1}dZ_1(t)$$

$$dF_2(t) = \mu_2 dt + \sigma_{p2}dZ_2(t).$$

Then the calendar spread between two futures follows

(34) $$d\big(F_1(t) - F_2(t)\big) = (\mu_1 - \mu_2)dt + \sigma_p dZ_p(t)$$

where $\sigma_p^2 = \sigma_{p1}^2 + \sigma_{p2}^2 - 2\sigma_{p12}$ is the variance of the spread process, and σ_s is the covariance between two futures process. Then the solution of the calendar spread call option C is

(35) $$C_P = e^{-r\tau}\left[(F_1(t) - F_2(t) - X)\Phi\left(\frac{F_1(t) - F_2(t) - X}{\sigma_p\sqrt{\tau}}\right) + \sigma_p\sqrt{\tau}\phi\left(\frac{F_1(t) - F_2(t) - X}{\sigma_p\sqrt{\tau}}\right)\right]$$

where τ is time to maturity, $F_1(t)$ and $F_2(t)$ are current price level of two futures, X is strike price, and Φ and ϕ are CDF and PDF of standard normal distribution, respectively. We estimate the parameter σ_p of the model using empirical dataset. The estimated value is $\sigma_p = 0.9747$.

We use root mean squared error (RMSE) as a measure of accuracy. The results are provided by three types of moneyness (at the money, in the money, and out of the money) with different trading days. First, the exercise price of at the money is the same as the calendar spread of the first trading day ($X = Spread$). Second, exercise price of in the money is the calendar spread of the first trading day minus one ($X = Spread - 1$), the exercise price for out of the money is specified as the spread of the first trading day plus one ($X = Spread + 1$). Also, the exercise prices and spreads given in the first trading day of each month (from January 2000 to October 2016) are used to compute option payoffs over the subsequent trading days. Five trading days are trading day 10, 20, 40, 80, and 120. The RMSE for model h is calculated from the equation

(36)
$$RMSE_h = \sqrt{\frac{1}{M}\sum_{i=1}^{N}\left[V_{ih} - \left(\hat{V}_{ih} \mid trading\ days = j\right)\right]^2},\ j = 10, 20, 40, 80, 120$$

where M is the number of payoffs, $M = 202$, V_i is actual payoffs, and $\hat{V}_i$ denotes predicted payoffs (values) of the CSO from each model.

Table 3 reports payoffs of the five different maturities in 'at the money'. Averages of actual payoffs for several trading days are calculated by comparing the spreads given in the first trading day of each month with spreads at the expiration day of each option and its strike prices. In all cases, actual payoffs of call options are greater than that of calls. Gibson and Schwarz and the new model show similar pricing for 'at the money' options. Both Gibson and Schwarz and the new model overestimate put option values compared to actual payoffs. Poitras model tends to underestimate option values especially for options near expiration. Since Poitras model assumes no cointegration between the two futures and constant volatility for the calendar spread process, the model may fail to reflect the volatility dynamics of the calendar spread process. Tables 4 and 5 report the payoffs of 'in the money' and 'out of the money' options. Gibson and Schwarz model overestimates put option values at 'in the money', whereas Poitras underestimates call options. The new model shows relatively good performance in pricing calls, however it overestimates put payoffs. In Table 5, both call and put option values from the Gibson and Schwarz model are overpriced. Poitras values are underpriced / overpriced in call / put option payoffs. The new model relatively performs well in pricing 'out of the money' options. Both call and put option values of the new model are close to actual payoffs. The RMSE values are reported as well to compare the performance of each model. The new

model mostly outperforms other models. For example, in Table 5, the new model presents smallest RMSE values at ‘out of money’ options in all expirations. Gibson and Schwartz (1990) model relatively performs well in ‘at the money’ pricing, but overestimates actual payoffs in ‘in the money’ and ‘out of money’.

Conclusions

The futures and options markets for crude oil are the most liquid and vigorously traded commodities contracts in the financial market. The term structure in the market is substantially affected by demand and supply of the crude oil, storage costs, convenience yield, and anticipation of production. Further, in contrast to other agricultural commodity markets where seasonality dominates the relationship between spot and futures price, the crude oil market has more continuous price evolution.

The new model proposed here describes the calendar spread between two futures based on the theory of storage. By doing so, the calendar spread process is converted to the processes of the nearby futures, volatility of the futures, and convenience yield based on the theory of storage. We assume that convenience follows arithmetic Brownian motion, nearby futures follows geometric Brownian motion, and volatility of nearby futures follows a CIR process. Further, our model allows correlations between the return of the nearby futures and its volatility and the convenience yield. Our specification of the model provides a flexible adjustment of the empirical features of the futures contracts.

In the empirical analysis, we calibrate parameters of the model using the analytical solution and market values of CSO. Closing prices of the 1-month CSO for

WTI futures are used as the market values. We test the performance of the new model by comparing with other existing pricing models and find that the new model with multi-stochastic factors outperforms the other two models. The reason may be the flexible specification of the model for adjusting dynamics of the underlying asset. The new model has clear advantages in controlling the dynamics of the calendar spread such as term structure effect and Samuelson effect.

The solution of the new model additionally requires a parameter calibration procedure, which can be sensitive to starting values. To get stable estimates, we employ an iterative two-step calibration method that differentiates procedures for the structural parameters and the spot processes. Further, we impose some upper / lower bound restrictions on the structural parameter estimates. However, future study would need to compare the results from the method used in this article with the results from other calibration approaches such as Bayesian estimation, Kalman filter, and particle filtering to verify robustness of the estimates.

We expect that our proposed CSO pricing model can be used both for market makers and the CME group. The new model has clear potential to lower bid-ask spreads. Also, the model is more flexible and accurate than the existing pricing models considered. Therefore, the model can be a promising alternative to provide reasonable settlement prices in the CSO market as well, which could increase the trade volume of the CSO market.

Table III-1. Calibrated Parameters of the New Model

k	θ	δ	σ_1	σ_2	ρ_1	ρ_2
0.52	0.03	0.39	0.32	0.24	0.17	-0.48

Table III-2. Summary Statistics of Interest Rate, Storage Costs, and Convenience Yield

Commodity	Sample Period	Mean (%, $/bbl)	Maximum (%, $/bbl)	Minimum (%, $/bbl)	S D (%, $/bbl)
T-BILL (3-month)	Jan 02, 1985 - Feb 27, 2017	4.22	11.14	0.00	2.80
Prime rate	Jan 02, 1985 - Feb 27, 2017	7.04	13.00	3.25	2.57
Storage cost	Jan 02, 1985 - Feb 27, 2017	0.24	0.40	0.15	0.08
Convenience yield	Jan 02, 1985 - Feb 27, 2017	1.72	15.57	-13.31	1.93
One-month spread	Jan 02, 1985 - Feb 27, 2017	-0.03	11.55	-8.49	0.81
Two-month spread	Jan 02, 1985 - Feb 27, 2017	-0.02	12.05	-12.71	1.38
Three-month spread	Jan 02, 1985 - Feb 27, 2017	0.01	12.14	-15.21	1.83

Table III-3. Price of 1 month WTI Calendar Spread Options 'at the money'

At the money ($X = Spread$)								
Trading days to expiration	Gibson & Schwartz		Poitras		New model		Actual payoff	
	Call	Put	Call	Put	Call	Put	Call	Put
10	0.52	0.45	0.06	0.06	0.55	0.40	0.53	0.21
20	0.55	0.50	0.09	0.09	0.61	0.45	0.53	0.22
40	0.60	0.53	0.13	0.13	0.61	0.51	0.54	0.21
80	0.63	0.59	0.18	0.18	0.61	0.52	0.54	0.20
120	0.65	0.66	0.22	0.22	0.62	0.54	0.55	0.19
RMSE								
10	0.33	0.44	0.45	0.45	0.21	0.21		
20	0.34	0.35	0.46	0.44	0.21	0.20		
40	0.32	0.32	0.44	0.42	0.20	0.24		
80	0.45	0.35	0.39	0.41	0.23	0.20		
120	0.48	0.45	0.46	0.45	0.32	0.31		

Table III-4. Price of 1 month WTI Calendar Spread Options 'in the money'

In the money ($X = Spread - 1$)

Trading days to expiration	Gibson & Schwartz		Poitras		New model		Actual payoff	
	Call	Put	Call	Put	Call	Put	Call	Put
10	1.03	0.32	1.00	0.00	1.12	0.13	1.34	0.03
20	1.16	0.37	1.00	0.00	1.13	0.19	1.34	0.02
40	1.19	0.41	1.00	0.00	1.15	0.24	1.35	0.02
80	1.25	0.45	1.01	0.01	1.16	0.27	1.37	0.02
120	1.32	0.54	1.01	0.02	1.17	0.29	1.38	0.01
RMSE								
10	0.73	0.50	0.48	0.25	0.48	0.10		
20	0.69	0.54	0.64	0.16	0.61	0.11		
40	0.61	1.10	0.63	0.20	0.57	0.22		
80	0.74	1.15	0.50	0.12	0.49	0.16		
120	0.50	1.32	0.64	0.09	0.61	0.13		

Table III-5. Price of 1 month WTI Calendar Spread Options 'out of money'

Out of the money ($X = Spread + 1$)								
Trading days to expiration	Gibson & Schwartz		Poitras		New model		Actual payoff	
	Call	Put	Call	Put	Call	Put	Call	Put
10	0.33	1.01	0.00	1.00	0.01	0.64	0.14	0.82
20	0.37	1.12	0.00	1.00	0.04	0.68	0.14	0.82
40	0.41	1.15	0.00	1.00	0.07	0.71	0.14	0.81
80	0.47	1.21	0.01	1.01	0.10	0.71	0.14	0.79
120	0.49	1.31	0.02	1.01	0.10	0.73	0.15	0.78
RMSE								
10	0.83	1.12	0.16	0.31	0.13	0.32		
20	0.82	1.13	0.15	0.29	0.09	0.25		
40	0.76	1.07	0.16	0.30	0.10	0.34		
80	0.92	0.91	0.15	0.32	0.08	0.31		
120	0.90	1.48	0.13	0.33	0.06	0.30		

REFERENCES

Black, F., and M, Scholes. 1973. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy* 81: 637–54.

Brennan, M. J. 1958. The Supply of Storage. *The American Economic Review* 48: 50–72.

Carr, P., and L. Wu. 2007. Stochastic Skew in Currency Options. *Journal of Financial Economics*. 86: 213–47.

Christoffersen, P., S. Heston, K. Jacobs. 2009. The Shape and Term Structure of the Index Option Smirk: Why Multifactor Stochastic Volatility Models Work So Well. *Management Science* 55: 1914–32.

Commandeur, J.J., and S.J. Koopman. 2007. An Introduction to State Space Time Series Analysis. New York: Oxford University Press.

Eraker, B. 2004. Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices. *Journal of Finance* 59: 1367–403.

Fama, E.F., and K.R. French. 1987. Commodity Futures Prices: Some Evidence on Forecast Power, Premiums, and the Theory of Storage. *The Journal of Business* 60: 55–73.

Franken, J.R.V., P. Garcia, and S.H. Irwin. 2009. Is Storage at a Loss Merely an Illusion of Spatial Aggregation? Journal of Agribusiness 27:65-84.

Gelman, A., J.H.S. Carlin, and D. Rubin. 2014. *Bayesian Data Analysis*. Vol. 2. Boca Raton, FL, USA: Chapman & Hall/CRC.

Gibson, R., and Schwartz, E. 1990. Stochastic Convenience Yield and the Pricing of Oil Contingent Claims. *Journal of Finance* 45: 959–76.

Heston, S.L. 1993. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *The Review of Financial Studies* 6: 327–343.

Heath, D., R. A. Jarrow and A.J. Morton. 1992. Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation. *Econometrica* 60: 77–105.

Hinz, J., and M. Fehr. 2010. Storage Costs in Commodity Option Pricing. SIAM *Journal on Financial Mathematics* 1: 729-51.

Irwin, S.H., P. Garcia, D.L. Good., and L.E. Kunda. 2011. Spreads and Non-convergence in Chicago Board of Trade Corn, Soybean, and Wheat Futures: Are Index Funds to Blame? *Applied Economic Perspectives and Policy* 33: 116–142.

Jones, C. 2003. The Dynamics of Stochastic Volatility: Evidence from Underlying and Options Markets. *Journal of Econometrics* 116: 181–224.

Nakajima K., and A. Maeda. 2007. Pricing Commodity Spread Options with Stochastic Term Structure of Convenience Yields and Interest Rates. *Asia-Pacific Financial Markets* 14: 157-84.

Murphy, J.A. 1990. A Modification and Re Examination of the Bachelier Option Pricing Model. *American Economist* 34: 34–41.

Pan, J. 2002. The Jump-risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study. *Journal of Financial Economics*. 63: 3–50.

Poitras, G. 1990. The Distribution of Gold Futures Spreads. *The Journal of Futures Markets* 10: 643–59.

Schachermayer, W., and J. Teichmann. 2008. How Close Are the Option Pricing Formulas of Bachelier and Black-Merton-Scholes? *Mathematical Finance* 18:155-70.

Schwartz, E.S. 1997. The Stochastic Behavior of Commodity Prices: Implications for Valuation and Hedging. *Journal of Finance* 52: 923-73.

Seok, J., and B.W. Brorsen. 2015. Pricing Corn Calendar Spread Options. Paper presented at the NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management, St. Louis, Missouri, April 20 – 21.

Shimko, D. 1994. Options on Futures Spreads: Hedging, Speculation and Valuation. *The Journal of Futures Markets* 14: 183–213.

Thompson, S. 1986. Returns to Storage in Coffee and Cocoa Futures Markets. *The Journal of Futures Markets* 6: 541-64.

Working, H. 1949. The Theory of Price of Storage. *American Economic Review* 39:1254-62

Zulauf, C.R., H. Zhou, and M.C. Roberts. 2006. Updating the Estimation of the Supply of Storage. *Journal of Futures Markets* 26: 657-76.

VITA

Eunchun Park

Candidate for the Degree of

Doctor of Philosophy

Thesis: THREE ESSAYS ON INSURANCE AND OPTION IN AGRICULTURAL ECONOMICS

Major Field: Agricultural Economics

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Agricultural Economics at Oklahoma State University, Stillwater, Oklahoma in May, 2017.

Completed the requirements for the Master of Economics in Food and Resource Economics at Korea University, Seoul, South Korea in 2013.

Completed the requirements for the Bachelor of of Economics in Food and Resource Economics at Korea University, Seoul, South Korea in 2010.

Experience:

Graduate Research Assistant
Department of Agricultural Economics, Oklahoma State University, 2013-2017

Research Assistant
Green Growth Research Group, Korea Energy Economic Institute,
Seoul, South Korea, 2012-2013

Graduate Teaching/Research Assistant
Department of Food and Resource Economics, Korea University, 2010-2013