The current issue and full text archive of this journal is available on Emerald Insight at:
www.emeraldinsight.com/0002-1466.htm

Received 2 August 2017
Revised 15 April 2018
Accepted 23 April 2018

A credibility-based yield forecasting model for crop reinsurance pricing and weather risk management

Wenjun Zhu
School of Finance, Nankai University, Tianjin, China
Lysa Porth
Department of Agricultural Risk Management and Insurance, Warren Centre for Actuarial Studies and Research, University of Manitoba, Winnipeg, Canada and
Department of Agribusiness and Agricultural Economics, University of Manitoba, Winnipeg, Canada and
Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Canada, and
Ken Seng Tan
Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Canada

Abstract

Purpose – The purpose of this paper is to propose an improved reinsurance pricing framework, which includes a crop yield forecasting model that integrates weather variables and crop production information from different geographically correlated regions using a new credibility estimator, and closed form reinsurance pricing formulas. A yield restatement approach to account for changing crop mix through time is also demonstrated.

Design/methodology/approach – The new crop yield forecasting model is empirically analyzed based on detailed farm-level data from Manitoba, Canada, covering 216 crop varieties from 19,238 farms from 1996 to 2011. As well, corresponding weather data from 30 stations, including daily temperature and precipitation, are considered. Algorithms that combine screening regression, cross-validation and principal component analysis are evaluated for the purpose of achieving efficient dimension reduction and model selection.

Findings – The results show that the new yield forecasting model provides significant improvements over the classical regression model, both in terms of in-sample and out-of-sample forecasting abilities.

Research limitations/implications – The empirical analysis is limited to data from the province of Manitoba, Canada, and other regions may show different results.

Practical implications – This research is useful from a risk management perspective for insurers and reinsurers, and the framework may also be used to develop improved weather risk management strategies to help manage adverse weather events.

Originality/value – This is the first paper to integrate a credibility estimator for crop yield forecasting, and develop a closed form reinsurance pricing formula.

Keywords Credibility, Insurance pricing, Reinsurance pricing, Yield forecasting

Paper type Research paper

1. Introduction

Weather risk plays an important role in forecasting crop yield, which is critical for agricultural insurance ratemaking. An improved crop yield forecasting model will enhance the scientific ratemaking framework for crop (re)insurers, and will support the acceleration of the agricultural insurance market worldwide (Ozaki *et al.*, 2008). However, in the presence of systemic weather risk, there are many challenges to efficiently and accurately forecast

Agricultural Finance Review
Vol. 79 No. 1, 2019
pp. 2-26
© Emerald Publishing Limited
0002-1466
DOI 10.1108/AFR-08-2017-0064

crop yields, including effects of possible climate changes, selecting predicting variables, restating crop mix, and modeling geographical differences across regions.

The objective of this paper is to address these difficulties by developing a new crop yield forecasting model and reinsurance pricing framework. A main focus will be on enhancing the actuarial ratemaking for agricultural reinsurance by integrating weather risks and production information from different geographical regions. This research will also add to the literature by proposing a yield restatement method that considers changing crop mix over time, and this step is essential to maintain the consistency of the data. In addition, a comprehensive weather index system is composed to reflect the nonlinear relationship of crop yields and weather variables. Efficient algorithms for selecting an optimal prediction matrix are also devised, and to address the geographical heterogeneity, an expanded regression credibility model is proposed to improve the crop yield forecasting model. The model is validated through empirical forecasting results, providing the best in-sample and out-of-sample forecasting results among the various approaches that are investigated in this paper.

An overview of the modeling framework of this paper can be expressed in the flow chart shown in Figure 1. One of the essential contributions of this paper is that it provides a comprehensive framework for insurance and reinsurance companies to address weather risks by constructing an integrated management framework with analytical formulas that include model construction, premium calculation, and risk assessment. The yield forecasting and pricing models developed in this paper may also be used by agricultural insurers and reinsurers to develop improved weather risk management strategies.

1.1 Restatement of yield for crop mix

The evolution of farming practices creates difficulties for loss modeling given that historical data become less representative of the current experience. For example, mixed cropping, which means growing more than one crop simultaneously on the same piece of land, helps the producer diversify, and protect against losses from extreme adverse weather conditions. Crop rotation is also common practice, planting different crop types in different years to provide nutrients to the soil and help mitigate the build-up of pathogens and pests. Further, changes in biotechnology and increases in commodity prices have influenced the types of crops that farmers plant. Together the effect is a change in crop mix over time, which creates concerns over the representativeness of the data. Therefore, an important first step in yield or loss modeling is to perform a restatement approach for historical yield data to help ensure that the older observations are good indicators of future crop production (Coble *et al.*, 2008, 2011; Sherrick *et al.*, 2004; Woodard, 2014; Woodard and Sherrick, 2011a). Despite its importance, yield restatement has rarely been discussed in literature. In this paper, an algorithm is provided to restate yield for crop mix. Serving as a pre-process step, the restatement algorithm helps to prepare the data so that it is more representative of present farming conditions and the overall risk profile of the current program.

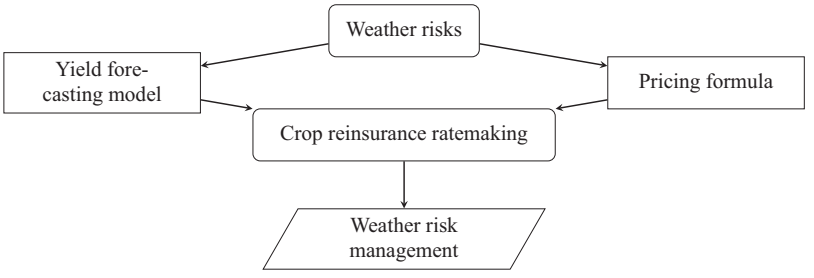


Figure 1.
Flow chat of the general modeling framework in this paper

1.2 Yield forecasting

Yield forecasting is challenging, largely due to extreme weather events that often lead to wide spread losses across many geographic regions (Cassman, 1999; Dai *et al.*, 2004; Lobell and Asner, 2003). A good yield forecasting model helps to predict crop yields before harvest actually takes place, and offers a scientific foundation for the ratemaking of traditional crop insurance contracts, such as multi-peril crop insurance. It is widely known that the traditional crop insurance design is subject to various challenges, including adverse selection and moral hazard (Chambers, 1989; Nelson and Loehman, 1987; Quiggen *et al.*, 1993). To help alleviate these issues, recent research has focused on index-based insurance (IBI) (Wang and Zhang, 2003). IBI is also faced with challenges, including basis risk, which refers to the mismatch between the actual loss on the farm and the payment to the farmer based on the index. A better understanding of the relationship between crop yield and weather risks becomes extremely important to help reduce basis risk.

The crop growth process is a natural system based on the interaction of soil, air, water, and crops. Therefore, an integrated yield model that combines meteorological and climate data may be beneficial (Campbell and Diebold, 2005; Kleibera *et al.*, 2011). However, a major challenge faced with integrating yield and weather information is how to efficiently select weather variables from a complex set of correlated weather variables and reduce the dimension of explanatory variables to an acceptable number. In this paper, a comprehensive weather index system is considered to describe the nonlinear relationship of weather variables and crop yields. Three model selection algorithms are proposed by combining screening regression (SR), principal component analysis (PCA) and cross-validation (CV), to efficiently help to achieve the goal of model selection and dimension reduction. Empirical results show that compared to the traditional multiple regression method, the three model selection algorithms presented in this paper have better in-sample and out-of-sample forecasting abilities.

1.3 Geographical heterogeneity and credibility approach

Crop yield is impacted by different weather variables across various geographical regions (i.e. geographical heterogeneity), and this effect has been studied on a relatively limited basis in literature (Cai *et al.*, 2013). In McCarl *et al.* (2008), the impact of changes in climate is shown to vary across different regions in the USA, and in addition the effect of spatial heterogeneity is discussed and found to be prevalent. In this paper, the effect of spatial heterogeneity is explored, and it is expected that for certain dominant weather variables there may be a positive effect on yield in some geographical regions, yet, a negative impact in others. To more closely examine this effect, this paper proposes a new credibility estimator to consider the geographical heterogeneity. Credibility theory has received relatively little attention in the area of agricultural insurance pricing, with the exception of research by Josephson *et al.* (2000), Pai *et al.* (2014), and Porth *et al.* (2014). The new credibility estimator proposed in this paper has advantages including unbiasedness and smaller mean quadratic loss, and the empirical analysis shows that the crop yield forecasting based on the new credibility estimator has smaller in-sample and out-of-sample forecasting errors.

The remainder of this paper proceeds as follows. Section 2 introduces the data, followed by a yield restatement approach for crop mix. In addition, a weather index system is developed, and three model selection algorithms are shown. Section 3 proposes a new credibility estimator to address geographical heterogeneity and improve crop yield forecasting. Closed form pricing formulas are also introduced to enhance the reinsurance pricing framework. Section 4 provides concluding remarks, and the appendices outline the proofs.

2. Yield forecasting

In general, the agriculture sector is faced with the challenge of data scarcity. In most countries, there is usually only one growing season each year, which leads to only one yield observation

for each location every year. Further, reinsurers do not often have access to farm-level data, and instead use highly aggregated yield level. In Canada, for example, crop insurers usually provide reinsurers with data by crop type aggregated to the provincial level. Therefore, a crop yield forecasting model with high out-of-sample prediction ability provides an important predictive modeling tool for agricultural reinsurance companies to compute rates, perform risk modeling, manage reserves, etc. In this section, the data are first discussed, followed by the development of a crop yield restatement algorithm to account for changing crop mix through time. Next, a weather index system is constructed, and three model selection algorithms are introduced.

2.1 Data introduction

A detailed farm-level crop yield data set is used to empirical evaluate the crop yield forecasting model developed in this paper. The farm-level data represents the province of Manitoba, Canada, and the data panel covers 216 crop types from 19,238 farms over the period 1996–2011. In addition, weather station data are analyzed, and this Adjusted and Homogenized Canadian Climate Data are obtained from Environment Canada. The data include daily temperature (maximum, minimum, and mean) from 24 weather stations, and daily precipitation from 30 weather stations in Manitoba. The weather data contain some missing points; therefore, in order to fill in the missing data in the weather data set, the Ordinary Kriging method is employed. "Kriging" is synonymous with "optimal prediction," which predicts unknown values from data observed at known locations. Through the use of geostatistic spatial models, such as Kriging, missing data are optimally predicted by minimizing the squared prediction errors (Matheron, 1963; Plant, 2012).

2.2 Restatement of yield for crop mix

In order to ensure that the crop yield data observed historically is a good indicator of future crop production, it is necessary to pre-process the yield data using a procedure commonly known as "restatement." Restatement refers to a procedure in which historical data are adjusted so that the resulting data are a better indication and more representative of the present situation. This pre-processing procedure is particularly important in the context of agriculture for reasons such as the evolution of technology, improved farming practice, changes in weather conditions, or any other factor that may have a significant impact on crop production over the years (Coble *et al.*, 2011; Woodard, 2014). In this paper, we focus on yield restatement due to crop mix. This is important when considering a reinsurance portfolio, which contains many different crop types. To the best of the author's knowledge, there is no literature that formally addresses crop yield restatement. Therefore, the approach outlined below is based on discussions with practitioners, and hence is consistent with market practice. Some important points about restatement procedures are as follows:

- Over the years, there have been many different types of crops produced; therefore, it is unrealistic to restate each and every single crop. Instead, we focus on restating a representative crop mix, which is determined from the most recent experience to provide a better reflection of the current risk profile.
- Due to the variations among crop types, our restatement procedure quantifies the crop exposure by using the actual land area, instead of using weight, monetary value, etc.
- Since we have yield data at the farm level, our restatement procedure first restates crop mix at the farm level and then restates at the municipality level, which is more likely the data aggregation level used by reinsurers. Note that the municipality level in Canada is comparable to the county-level yield data in the US system.
- There are T years of data and each year is indexed by $t = 1, \ldots, T$.

- There are d risk categories (i.e. municipalities) and each risk category is indexed by $i = 1, \ldots, d$. In our empirical studies, we have $T = 16$ and $d = 122$.
- For each time t and risk category i, there are J farms and each farm is indexed by $j = 1, \ldots, J$. Note that more precisely, J is a function of time t and risk category i. Here we suppress both subscripts for brevity.
- The total number of crops that have been produced over T years is denoted by K and each crop is indexed by $k = 1, \ldots, K$. Note that for any particular year and for any particular risk category, the number of crops that a farm produces will likely be substantially smaller than K.
- Let $y_{i,j,k,t}$ and $A_{i,j,k,t}$ denote the yield and acres for risk category i, farm j, crop k and in year t, respectively.

The first step in our proposed restatement procedure is to identify the "main crop mix" that the municipality has been producing in recent years. Here, the "main crop mix" is defined as the minimum number of crop types that covers at least 90 percent of the total farming acres over the most recent five years. The "optimal crop mix" consists of the "main crop mix" and "others," where "others" captures the remaining crops types that are not part of the "main crop mix," but are produced in the last five years. The "optimal crop mix" determined in this way is assumed to be representative of the next year's farming practice and hence it will be used as "benchmark" for restating the historical crop yields. By denoting K^* as the total number of crops in the "main crop mix" for a given municipality, the set $K = \{0, k_1, k_2, \ldots, k_{K^*}\}$ is then used to describe the "optimal crop mix" with indexes $k_1, k_2, \ldots, k_{K^*}$ for identifying crops that comprise the "main crop mix" and the index 0 for capturing crops that are in "others."

Using the "optimal crop mix" as the benchmark, for each municipality and for each year, we identify the subset of farms that has produced at least one of the crops in the "main crop mix." We use J^* to denote the number of such farms and $J = \{j_1, j_2, \ldots, j_{J^*}\}$ to denote the set of indexes for identifying these farms.

The restatement procedure is summarized in the following algorithm:

Algorithm 2.1. For each municipality, $i = 1, \ldots, d$.

Step 1. Determine the optimal crop mix. This is denoted by K^* *and* $K = \{0, k_1, k_2, \ldots, k_{K^*}\}$.

Step 2. For $t = 1, \ldots, T$.

Step 2a. Determine J^* and $J = \{j_1, j_2, \ldots, j_{J^*}\}$.

Step 2b. For $j \in J$ and $k \in K$, calculate the ratio of acres for each crop type in the "optimal crop mix" as:

$$a_{i,j,k,t} = \frac{A_{i,j,k,t}}{\sum_{l \in K} A_{i,j,l,t}} \tag{1}$$

Step 3. Restatement of the farm-level ratio of acre: assume the restated ratio $a^R_{i,j,k,t}$ satisfies a third-order polynomial, i.e.:

$$a^R_{i,j,k,t} = \alpha_{i,j,k} + \beta_{i,j,k} t + \gamma_{i,j,k} t^2 + \eta_{i,j,k} t^3,$$

where $\alpha_{i,j,k}, \beta_{i,j,k}, \gamma_{i,j,k}$ and $\eta_{i,j,k}$ are the coefficients of the polynomial. For $j \in J$ and $k \in K$, the corresponding coefficients are optimally determined by minimizing the sum of square errors as follows:

$$\min_{a_{i,j,k}, b_{i,j,k}, c_{i,j,k}} \sum_{t=1}^{T} \left(a^R_{i,j,k,t} - a_{i,j,k,t} \right)^2.$$

Step 4. Restatement of municipality level yield per unit acre. Finally from the restated farm-level ratio $a^R_{i,j,k,t}$, the yield per unit acre, $y^R_{i,t}$, *for* $t = 1, \ldots, T$, is restated as:

$$y^R_{i,t} = \frac{\sum_{j \in J}\sum_{k \in K} a^R_{i,j,k,t} y_{i,j,k,t}}{\sum_{j \in J}\sum_{k \in K} A_{i,j,k,t}} \tag{2}$$

The above algorithm produces $\{y^R_{i,t};\ i = 1, \ldots, d, t = 1, \ldots, T\}$ and these are the restated crop yields at the municipality level that will be used for subsequent analysis. While in Step 3, we postulate that the restated ratio $a_{i,j,k,t}$ is captured by a third-degree polynomial, other functions are plausible.

2.3 Weather index system

This section develops a weather index system to account for nonlinear impacts of weather variables on crop production in order to improve yield forecasting, and hence loss forecasting. A weather index is a nonlinear function of weather observations, which provides a direct and intuitive meteorological measure for certain weather risk. Weather derivatives, for example, may be a very efficient approach to transfer systemic weather risks in agriculture (Woodard and Garcia, 2008) and can be written on a cumulative weather index known as growing degree days (GDD). GDD is an indicator of the suitability for a crop to grow in terms of some benchmark temperature. It is assigned a zero value if the daily temperature falls below the base temperature ($\tilde{T}$); otherwise it is the difference between daily temperature and the base temperature. More explicitly, the value of GDD on day t is:

$$GDD_t = \max\left(0, T_t - \tilde{T}\right), \tag{3}$$

where T_t is the average of maximum and minimum daily temperature, defined as, $T_t = (T_{min} + T_{max})/(2)$, and $\tilde{T}$ is the base temperature. In addition to GDD, popular weather indexes include cooling degree days, heating degree days and cumulative average temperature[1]. Another index system, called crop heat units, is calculated from calibrated daily minimum and maximum temperatures (Brown, 1975), and was originally developed for field corn in Ontario.

In this paper, a weather index system is developed based on temperature and precipitation information, aiming at providing a simplified, yet, comprehensive measurement of weather. We define temperature thresholds θ_1, θ_2, and θ_3 and precipitation thresholds λ_1 and λ_2 as the following: θ_1 and θ_2 are the base temperatures, which represent the minimum temperatures that the crop can grow during the day and the night, respectively; θ_3 is the temperature during the day that the crop could grow at the highest rate. In the CHU system, $\theta_1 = 10$°C, $\theta_2 = 4.4$°C, and $\theta_3 = 30$°C. As such, we use this as a benchmark and analyze a wide range of temperature thresholds in order to assess the detailed relationships between the crop yields and weather variables. To be more specific, we set θ_1 from 6 to 10°C, θ_2 from 0 to 4°C, and θ_3 from 26 to 30°C, all with 1°C increment. As with precipitation, there is scarce literature discussing the selection of the thresholds; therefore, we use the first and third quartiles of historical precipitation to define λ_1 and λ_2, respectively. In addition, from an agronomy point of view, the growing season (i.e. from May to October for Canada) is the period when the weather may play the most important role in crop growth. As a consequence, we construct the weather indices during the growing season. Also note that the weather index system is developed on both a monthly and an annual basis. The advantage of studying the monthly weather indices is that more detailed dynamics between the crop yields and weather variables are likely to be detected. First, the (daily) weather index system is defined, then using different aggregation functions the daily indices are integrated along different periods (i.e. either monthly or annually through the

entire growing season) to create a design matrix for crop yield forecasting. The detailed definitions and notation of the weather index system and the aggregating functions are listed below.

Weather index system.

- Night growing degree low (NGDL):

$$NGDL_i = \min(MinT - \theta_{1i}, 0), \quad i = 1, \ldots 5,$$

where $\theta_{11} = 0$, $\theta_{12} = 1$, $\theta_{13} = 2$, $\theta_{14} = 3$, and $\theta_{15} = 4$(°C), indicating the night temperature above which the crops can grow. $MinT$ is the minimum daily temperature.

- Day growing degree low (DGDL):

$$DGDL_i = \min(MaxT - \theta_{2i}, 0), \quad i = 1, \ldots, 5,$$

where $\theta_{21} = 6$, $\theta_{22} = 7$, $\theta_{23} = 8$, $\theta_{24} = 9$, $\theta_{25} = 10$(°C), indicating the daytime temperature under which the crops may stop development. $MaxT$ is the maximum daily temperature.

- Day growing degree high (DGDH):

$$DGDH_i = \max(MaxT - \theta_{3i}, 0), \quad i = 1, \ldots, 5,$$

where $\theta_{31} = 26$, $\theta_{32} = 27$, $\theta_{33} = 28$, $\theta_{34} = 29$, $\theta_{35} = 30$(°C), indicating the maximum daytime temperature for the crops to grow.

- Precipitation high (PREH):

$$PREH = \max(P - \lambda_1, 0),$$

where $\lambda_1 = 0.1$ mm, which is about the 0.25 percent quantile of historical precipitations.

- Precipitation low (PREL):

$$PREL = \min(P - \lambda_2, 0),$$

where $\lambda_2 = 2$ mm, which is about the 0.75 percent quantile of historical precipitations.

The aggregation functions considered in this paper include maximum value (denoted as "max"), minima value (denoted as "min"), average value (denoted as "avg"), and total number of nonzero values (denoted as "cot"). These aggregation functions essentially divide the weather indices into three types:

(1) Average index: using function "avg," the average indices provide aggregate measures of weather conditions during a defined period.

(2) Extreme events: using function "min" and "max," these indices describe extreme events during a defined period.

(3) Extreme days: using function "cot," these indices count the number of days during a defined period experiencing extreme weather conditions.

After excluding the indices that duplicate the values of the existing variables in the design matrix or those making the matrix singular, we construct a 140-dimensional design matrix, where each column is an explanatory variable and each row is one year observation for the corresponding variables. The structure of the notation of the weather variables (i.e. indices) is I + function + _period, where I is the weather index, "function" is the aggregation function used, and "period" is the period through which the index is calculated.

For example, NGDL1cot_May means the total number of days that NGDL is above 0 during May, which provides knowledge about the downside low temperature risk during May. The detailed notation and the interpretations of these explanatory variables are listed in Table I.

2.4 Model selection algorithms

Statistical models are widely used as an alternative to agronomic process-based models in predicting crop yields. Many studies have used multiple regression models to identify the contributions of weather variables to crop yields and to perform the forecasting (Lobell and Burke, 2010; Shi *et al.*, 2013). The general form of the commonly used regression model can be expressed as:

$$Y_t = \beta_0 + \beta_1 V_{1,t} + \beta_2 V_{2,t} + \ldots + \epsilon_t, \tag{4}$$

where Y_t represents the crop yields which could be either time series for a single crop yield or average yields across different regions. $V_{1,t}, V_{2,t}, \ldots$ are explanatory weather variables, such as time, t, growing season mean temperature (T_{avg}), and total precipitation (P_{tol}), or their functions, for instance, T^2_{avg}, GDD_{tol}, etc.

A first and foremost challenge in this multiple regression model is the model selection problem, and caution is necessary because any misspecification, misinterpretation, or existence of multicolinearity will strongly affect the predicting results (Ramsey and Schafer, 2012). Common modeling practices use either all possible subsets method or stepwise methods. To be

Index (I)	Threshold (°C/mm)	Function (fun)	Notation	Example
T	–	max min avg	T+fun+_period	Tmax_May: Maximum temperature in May
P	–	max min avg	P+fun+_period	Pavg: Average precipitation during growing season
lightgray NGDL	$\theta_{11}=0$, $\theta_{12}=1$, $\theta_{13}=2$, $\theta_{14}=3$, $\theta_{15}=4$	max min avg cot	NGDL+j+fun+_period, $j=1,\ldots,5$	NGDL1min_Aug: Minimum NGDL (with $\theta_{11}=0$°C) during August
DGDL	$\theta_{21}=6$, $\theta_{22}=7$, $\theta_{23}=8$, $\theta_{24}=9$, $\theta_{25}=10$	max min avg cot	DGDL+j+fun+_period, $j=1,\ldots,5$	DGDL3cot_Jul: Number of days DGDL (with $\theta_{21}=8°C$) is above zero in July
DGDH	$\theta_{31}=26$, $\theta_{32}=27$, $\theta_{33}=28$, $\theta_{34}=29$, $\theta_{35}=30$	max min avg cot	DGDH+j+fun+_period, $j=1,\ldots,5$	DGDH5avg_Oct: Average DGDH (with $\theta_{35}=30°C$) during October
PREH	0.1	max avg cot	PREH+fun+_period	PREHavg_Jun: Average PREH (with $\lambda=0.1$ mm) during June
PREL	2	max avg cot	PREL+fun+_period	PRELcot_Sep: Number of days that PREL (with $\lambda=2$ mm) is above zero during September

Notes: "Index" is the weather index variables, "function" is the aggregation function used and "period" is the period through which the index is calculated, where a blank "period" represents the index for the whole growing season, otherwise "period" can be "May", "June", "July", "August", "September", and "October"

Table I. Notation and definitions of weather indices in the full design matrix

more specific, the all possible subsets method compares possible combinations of explanatory variables and uses statistics such as Mallows' C_p (Mallows, 2000) to select the best model. The stepwise methods add or remove variables until achieving the best model according to statistics such as Akaike information criterion (AIC) (Akaike, 1974) or Bayesian information criterion (BIC) (Schwarz, 1978). However, with a high-dimensional design matrix, such as the 140-dimensional explanatory variables constructed in our study, the multiple regression methods can be long and tedious and may not lead to an optimal model. Additionally, the aforementioned model selection methods tend to be more of an experience-based method. In addition to referring to statistical tools, such as C_p, AIC, or BIC, knowledge of biophysics, agronomy, and ecology are also required in the variable selection procedure, which can often be complicated and expensive.

In our study, three model selection algorithms, based on SR, PCA, and CV, are proposed to help achieve the objectives of model selection and dimension reduction. SR reduces the dimensionality by allowing only those "important" explanatory variables in the regression model, while PCA transforms the original highly correlated variables into the uncorrelated principal components (PCs), and retains the variation of the data as much as possible. The optimal threshold of screening is identified through CV. CV has an advantage of limiting the overfitting problem, which is common in agricultural loss modeling with very limited historical observations (Woodard and Sherrick, 2011b). This helps to ensure the selected models will provide satisfying out-of-sample prediction abilities, which is of more importance compared to in-sample prediction from the crop yield forecasting prospective. For a detailed definition of PCA and CV, refer to Jolliffe (2002) and Kohavi (1995). The details of these approaches are presented in Algorithms 2.2, 2.3, and 2.4. To the best of our knowledge, this is the first time that algorithms based on these methods (SR, PCA, and CV) are proposed for dimension reduction and model validation for crop yield forecasting.

Algorithm 2.2. (SR Algorithm) For each municipality i, $i = 1, \ldots, d$, $d = 122$:

Step 1. Calculate sample covariance coefficients between the yields and jth explanatory variables in the full design matrix ($j = 1, \ldots, p$), denoted as $\hat{\rho}_{i,j}^{W}$. The design matrix starts from the full design matrix $W_i^{(0)}$ with dimension $d_i^{(0)} = p = 44$.

Step 2. Calculate r candidate thresholds $\rho_i = (\rho_i^{(1)}, \ldots, \rho_i^{(r)})'$,

Step 3. For each $\rho_i^{(l)}, l = 1, \ldots, r$:

Step 3a. Update design matrix according to threshold $\rho_i^{(l)}$: Exclude explanatory variables with $\hat{\rho}_{i,j}^{W}$ smaller than $\rho_i^{(l)}$. Update the design matrix to be $W_i^{(l)}$ with the dimension being $d_i^{(l)}$.

Step 3b. Calculate $\lambda_i^{(l)}$, the out-of-sample predicting error of $W_i^{(l)}$, using CV.

Step 4. Calculate optimal design matrix W_i^*: Record the optimal threshold with the smallest predicting error as ρ_i^*, and the corresponding design matrix as W_i^*.

Algorithm 2.3. (PCASR Algorithm) For each municipality i, $i = 1, \ldots, n$, $n = 122$:

Step 1. Do PCA transformation to full design matrix $W_i^{(0)}$ and get new design matrix $Z_i^{(0)}$, including $s_i^{(0)}$ PCs that retain 85 percent or more variance of the full design matrix.

Step 2. Calculate sample covariance coefficients between the yields and jth components in $Z_i^{(0)}$ ($j = 1, \ldots, s_i^{(0)}$), denoted as $\hat{\rho}_{i,j}^{Z}$.

Step 3. Calculate r candidate thresholds $\rho_i = (\rho_i^{(1)}, \ldots, \rho_i^{(r)})'$.

Step 4. For each $\rho_i^{(l)}$, $l = 1, \ldots, r$:

Step 4a. Update design matrix according to threshold $\rho_i^{(l)}$: Exclude components with covariance coefficients $\hat{\rho}_{i,j}^{Z}$ smaller than $\rho_i^{(l)}$. Update the design matrix to be $Z_i^{(l)}$ with the dimension $s_i^{(l)}$.

Step 4b. Calculate $\lambda_i^{(l)}$, the out-of-sample predicting error of $Z_i^{(l)}$, using CV.

Step 5. Calculate optimal design matrix Z_i^*: Record the optimal threshold with the smallest predicting error as ρ_i^*, and the corresponding design matrix as Z_i^*.

Algorithm 2.4. (Screening PCA regression (SPCAR) Algorithm) For each municipality i, $i = 1, \ldots, n$, $n = 122$:

Step 1. Calculate sample covariance coefficients between the yields and jth explanatory variables in the full design matrix ($j = 1, \ldots, p$), denoted as $\hat{\rho}_{i,j}^{W}$. The number of dimension starts from $d_i^{(0)} = p = 44$.

Step 2. Calculate r candidate thresholds $\rho_i = (\rho_i^{(1)}, \ldots, \rho_i^{(r)})'$.

Step 3. For each $\rho_i^{(l)}, l = 1, \ldots, p$:

Step 3a. Update design matrix according to threshold $\rho_i^{(l)}$:exclude components with covariance coefficients smaller than $\rho_i^{(l)}$. Update the design matrix to be $W_i^{(l)}$ with the dimension $d_i^{(l)}$.

Step 3b. Calculate $\lambda_i^{(l)}$, the out-of-sample predicting error of $W_i^{(l)}$, using CV.

Step 4. Calculate design matrix W_i^*: record the optimal threshold with the smallest predicting error as ρ_i^*, and the corresponding design matrix as W_i^*.

Step 5. Do PCA transformation to design matrix W_i^* and get Z_i^*, containing s_i^* PCs that retain 85 percent or more variance of W_i^*. Z_i^* is the optimal design matrix.

2.5 Yield forecasting results

The three algorithms in Section 2.4 are applied to each of the 122 municipalities in the Manitoba data set. To demonstrate that these algorithms are effective at dimension reduction, we also analyze the classical regression method for the 122 municipalities. We find that the proposed algorithms can reduce the 140-dimensional explanatory matrix to a manageable set for all 122 municipalities in the province of Manitoba. In addition to assessing dimension reduction, some statistical measures such as the AIC and adjusted R^2 (AR2) are also examined and computed. The resulting histograms and the fitted densities of the AIC over 122 municipalities for the three algorithms and classical regression method are depicted in Figure 2 and the corresponding AR2 are depicted in Figure 3.

Another important issue to examine is the forecasting power of the models. Mean square errors (MSE) are calculated to assess the in-sample prediction ability and leave-one-out cross-validation mean square errors (Loo-CVMSE) are calculated for the out-of-sample prediction. The forecasting results of the classical regression (CR) method are listed in the first column of Table II, while the forecasting results based on the three algorithms are reported in Columns 3–5 of Table II. In the table, "CR" represents the classical regression method, "SR" is the SR algorithm, "PCASR" is the PCA SR method, and "SPCAR" is the screening PCA regression algorithm. The following observations are based on the forecasting results:

- All three proposed algorithms have better fitting abilities compared to the CR method, reducing the AICs and improving the AR2s. From Figures 2 and 3, we can observe that the AIC density of the CR method lies on the right of the other densities while the AR2 density of the CR method lies on the left of the others, indicating that the CR performs worse than the other models.
- All three proposed algorithms improve the forecasting performance in terms of both in-sample and out-of-sample criteria. For example, from Table II we can see that the average in-sample error of the CR is 0.0208, which is 2.5 times the in-sample error of SR (0.0082).
- While the CR has acceptable in-sample forecasting errors, it does not perform well in the out-of-sample forecasting. For example, the average out-of-sample forecasting errors of the CR method is 20.8711, which is 12 times higher than the SR, 634 higher than the PCASR, and 608 higher than the SPCAR.

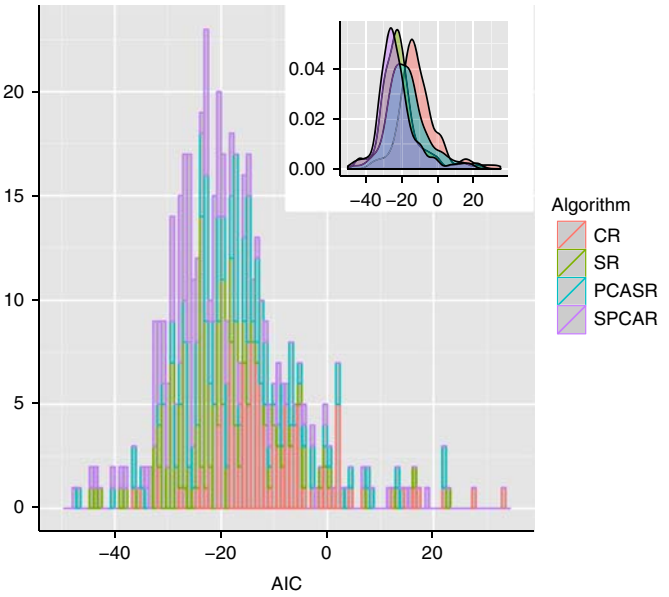


Figure 2.
Histogram and estimated density of AICs for three algorithms

Notes: CR, classical regression method; SR, the screening regression algorithm; PCASR, the PCA screening regression method; SPCAR, screening PCA regression algorithm

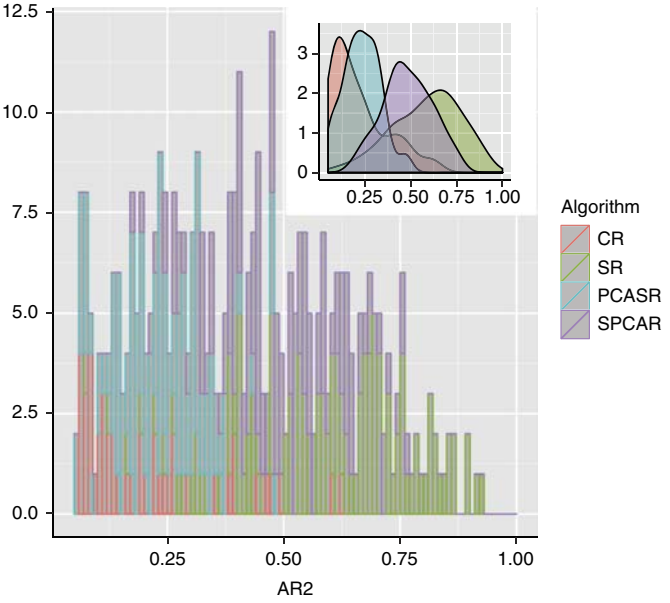


Figure 3.
Histogram and estimated density of AR2 for three algorithms

Notes: CR, classical regression method; SR, the screening regression algorithm; PCASR, PCA screening regression method; SPCAR, screening PCA regression algorithm

Statistic	CR	SR	PCASR	SPCAR
In-sample errors (MSE)				
Mean	0.0208	0.0082	0.0219	0.0142
Median	0.0109	0.0047	0.0119	0.0075
SD	0.0438	0.0159	0.0483	0.0368
Max	0.4247	0.1424	0.4947	0.3804
Out-of-sample errors (Loo-CVMSE)				
Mean	20.8711	1.7451	0.0329	0.0342
Median	13.8712	0.1291	0.0186	0.0158
SD	34.5929	3.5471	0.0677	0.0845
Max	255.6152	16.9406	0.6733	0.8467

Note: Important statistics are summarized in this table including the mean, median, standard deviation (SD), and maximum errors (Max)

Table II. Summary of forecasting results

It is difficult to conclude the best model among the three proposed algorithms, since there is a tradeoff between in-sample and out-of-sample forecasting abilities. For example, the SR algorithm has the best in-sample fitting ability based on the AR2 and in-sample MSE; however, a drawback is that it produces relatively large out-of-sample forecasting errors with a large standard deviation. This may be due to the fact that the SR performs well in some municipalities but not in others. The other two proposed algorithms, PCASR and SPCAR, although slightly worse than the SR in the in-sample forecasting, have better performance in the out-of-sample forecasting. In particular, the in-sample MSE of SPCAR is 35 percent better than PCASR (0.0142 vs 0.0219) while the out-of-sample of SPCAR is 4 percent worse than PCASR (0.0342 vs 0.0329). Considering that the out-of-sample forecasting capability is of greater importance for forecasting yields, we propose that the SPCAR is of more interest in yield forecasting from reducing the out-of-sample forecasting error point of view.

A main advantage of using a method such as the SPCAR is the ability to reduce the dimension to achieve an optimal design matrix that has only a few dominant explanatory variables or indices. Empirical analysis shows that there are great variations across municipalities, with the identified dominant weather indices varying from municipality to municipality. Therefore, it is of interest to compare the frequency of the weather indexes that are identified as dominant among all of the municipalities. The results are summarized in Figure 4. "Tmin_Oct" (i.e. minimum temperature during October) is the most dominant weather index that is identified by 106 out of 122 municipalities. "PRELavg_May" (i.e. average PREL during May) is the second most dominant weather index identified by 99 out of 122 municipalities. It is interesting to note that extreme low temperature and rainfall level have significant impacts on yield production.

3. Pricing framework

In this section, a reinsurance pricing framework is developed, which includes a new crop yield forecasting estimator to improve the out-of-sample forecasting ability. Closed form pricing formulas are also presented. Within this framework, the proposed forecasting estimator is constructed using an adjusted regression credibility estimator to take into consideration the geographical heterogeneity in the reinsurance pricing framework. This estimator has desirable statistical properties, such as unbiasedness and smaller mean quadratic losses, compared to the classic regression credibility estimator. The empirical analysis results further support the improvement of the proposed pricing framework in crop reinsurance ratemaking.

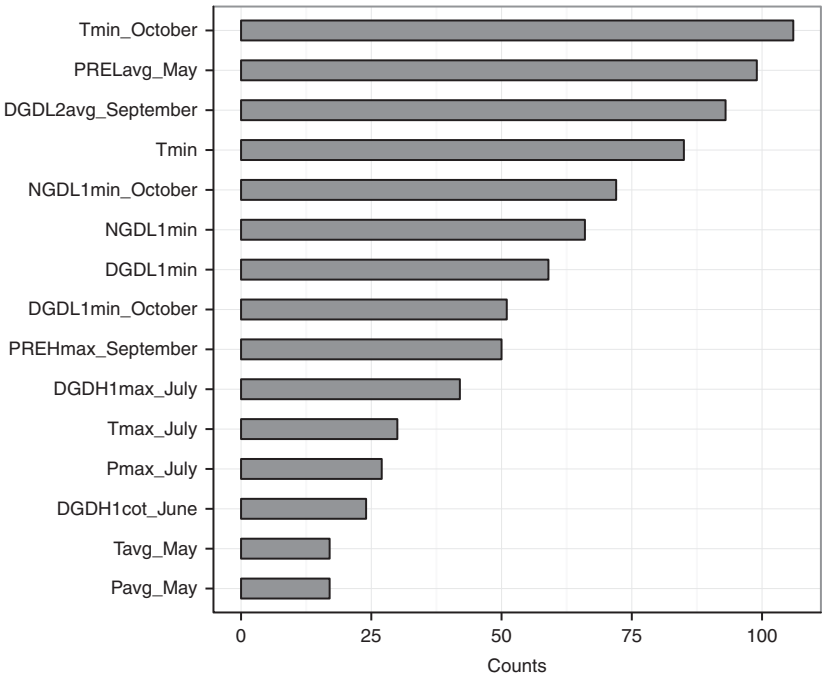


Figure 4.
Count of "dominant" weather indexes for each municipality

3.1 Geographical heterogeneity

In agricultural insurance/reinsurance ratemaking, geographical heterogeneity is typically ignored, yet, it plays an important role in pricing agricultural risk. Actuarial principles dictate that ratemaking approaches should be flexible, reproducible, and accurate. Ideally, this requires that reinsurers have stable and homogeneous risk portfolios, and pricing should be based on historical loss data. This helps to ensure that the resulting reinsurance premiums can be applied to all risks influenced by a variety of time factors, and can measure and represent the risk conditions reinsurance companies face.

Unfortunately, agricultural risk portfolios are usually time and spatially dependent and highly heterogeneous. Figure 5 shows the crop yields for 122 municipalities in Manitoba, Canada. To highlight the heterogeneity, we also list the summary of correlation coefficients of the most important weather indices selected in Section 2.4 in Table III, which summarizes the correlations of yields in 122 municipalities and the "key" weather indices. Taking "Tmax_Oct" as an example, we can see that the maximum correlation coefficient between the yield and Tmax_Oct is 0.46, while the minimum is −0.45. This means that this weather index may have positive impact on yield in some municipalities while it may have a negative impact on yield in other municipalities. This geographical heterogeneity indicates that the traditional method for pricing crop risks may fail to consider the spatial differences in the risk portfolio. In the next subsection, a new credibility estimator that integrates weather information and considers the geographical heterogeneity is proposed.

3.2 New credibility estimator

Let us asssume that there are d risk categories, and let iid random variable Θ_i describe the ith risk category, $i = 1, \ldots, d$. In practice, the risk category can be referred to as different geographical regions, different insurance or reinsurance companies, etc.

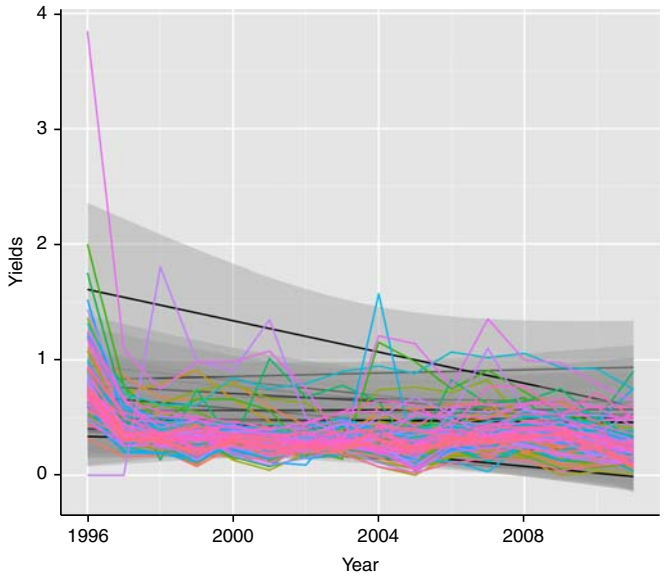


Figure 5.
Crop yields for 122 municipalities in Manitoba, Canada (1996–2011)

Index	$\rho_{\min}$	$\rho_{\max}$	$E(\rho)$	Median (ρ)	$\sqrt{\text{Var}(\rho)}$
Tmax_Oct	−0.45	0.46	−0.11	−0.12	0.14
Tmin	−0.64	0.62	−0.44	−0.49	0.16
PRELavg_May	−0.76	0.46	−0.54	−0.59	0.16
PREHmax_Sep	−0.37	0.66	0.39	0.42	0.15
NGDL1min_Oct	−0.67	0.57	−0.46	−0.49	0.17
NGDL1min	−0.64	0.62	−0.44	−0.49	0.16
DGDL2avg_Sep	−0.75	0.24	−0.50	−0.51	0.16
DGDL1min_Oct	−0.65	0.59	−0.47	−0.51	0.16
DGDL1min	−0.64	0.61	−0.45	−0.50	0.16
DGDH1max_Jul	−0.74	0.38	−0.38	−0.40	0.17

Note: Important statistics are summarized in this table, including the mean ($E(\rho)$), median, standard deviation ($\sqrt{\text{Var}(\rho)}$), maximum correlations ($\rho_{\max}$), and minimum correlations ($\rho_{\min}$)

Table III.
Summary of correlation coefficients of important weather Indices

Conditional on $\Theta_i = \theta_i$, let random vector $Y(\theta_i)$, $I(\theta_i)$, $L(\theta_i)$ denote the crop yield, indemnity, and liability in risk category i, respectively. We further assume $Y(\theta_i)$ follows a model with drift term and random term:

$$Y(\theta_i) = \mu_Y(\theta_i) + \sigma_Y(\theta_i)\varepsilon_i, \tag{5}$$

where ε_i is a standardized random variable with mean 0 and variance 1. This is a quite general model since both $\mu_Y(\theta_i)$ and $\sigma_Y(\theta_i)$ can be functions of time, risk categories, or other exogenous variable such as weather.

Hachemeister (1975) formulates a credibility regression model as a generalization of the traditional Bühlmann–Straub model (Bühlmann and Straub, 1970; Bühlmann, 1997; Bühlmann and Gisler, 2005), in which the loss data are assumed to follow a multiple regression model and the regression coefficients are credibility adjusted. To be more

specific, the classic regression credibility estimator of $\mu_Y(\theta_i)$, denoted as μ_Y^C (consider a single year hereafter), can be expressed as:

$$\hat{\mu}_Y^C = W_i^C\left[A_i^C\beta^C(\theta_i) + \left(\mathbb{I} - A_i^C\right)\beta_0^C\right], \tag{6}$$

where W_i^C is a $1 \times p$ design matrix for risk category i, and $\beta^Y(\theta_i)$ is a $p \times 1$ vector of regression coefficients. The other parameters are defined as:

$$\beta_0^Y = E\left(\beta^C(\theta_i)\right),$$

$$\sigma_Y^2 = E\left(\sigma_Y^2(\theta_i)\right),$$

$$\sum_i^Y = \sigma_Y^2\left(\left(W_i^Y\right)' W_i^Y\right)^{-1},$$

$$\Gamma_i^Y = Cov\left(\beta^C(\theta_i), \left(\beta^C(\theta_i)\right)'\right),$$

$$A_i^Y = \Gamma_i^Y\left(\Gamma_i^Y + \Sigma_i^Y\right)^{-1}.$$

An important concern about the credibility regression model is the model risk, where the basic assumption that the loss random variable follows the regression model needs to be carefully justified. With this in mind, we consider another arbitrary random variable $Z(\theta_i)$, assuming that it can be fully specified by the following regression model so that there is no model risk:

$$Z(\theta_i) = \mu_Z(\theta_i) = W_i\beta(\theta_i) + \sigma_Z(\theta_i)\epsilon, \tag{7}$$

where W_i is a $1 \times p$ design matrix, and $\beta(\theta_i)$ is a $p \times 1$ vector of regression coefficients. We further assume that $Z(\theta_i)$ is related to $Y(\theta_i)$ through the correlation coefficients defined as:

$$\rho_{YZ} = \frac{\mathrm{Cov}\left(\mu_Y(\theta_i)\mu_Z(\theta_i)\right)}{\sqrt{\mathrm{Var}\left(\mu_Y(\theta_i)\right)\mathrm{Var}\left(\mu_Z(\theta_i)\right)}}.$$

Let $Z = (Z_{ij})_{n\times d}$ be an observation of $Z(\theta_i)$ in year j, $i = 1, 2, \ldots, d$, $j = 1,2, \ldots, n$. We know that the credibility estimator of $Y(\theta_i)$ based on observations Z is defined as orthogonal projections (*Proj*) (see Bühlmann and Gisler, 2005; Definition 3.12):

$$\begin{aligned}\widehat{\mu_Y(\theta_i)} &= Proj\left(\mu_Y(\theta)\middle|L(1,Z)\right),\\ &= Proj\left(Proj\left(\mu_Y(\theta)\middle|L\left(1,Z,\mu_Z(\theta_i)\right)\right)\middle|L(1,Z)\right).\end{aligned} \tag{8}$$

Since $L(1, Z) \subset L(1, Z, \mu_Z(\theta_i))$, and $\rho(L(1, \mu_Z(\theta_i))) = \rho(L(1, Z, \mu_Z(\theta_i)))$ ($\rho(X)$ denotes the related σ-field generated by random variable X), therefore, by defining:

$$\begin{aligned}\tilde{\mu_Y}(\theta_i) &= \mathrm{Proj}\left(\mu_Y(\theta_i)\middle|L\left(1,Z,\mu_Z(\theta_i)\right)\right)\\ &= \mathrm{Proj}\left(\mu_Y(\theta_i)\middle|L\left(1,\mu_Z(\theta_i)\right)\right),\end{aligned}$$

we can express $\tilde{\mu_Y}(\theta_i)$ as a linear combination of $\mu_Z(\theta_i)$, namely:

$$\tilde{\mu_Y}(\theta_i) = a + b\mu_Z(\theta_i).$$

Thus, it satisfies the normal equations (see Bühlmann and Gisler, 2005; Corollary 3.17):

$$E\left(\mu_Y\tilde{(\theta_i)}\right) = a + bE\left(\mu_Z(\theta_i)\right) = E\left(\mu_Y(\theta_i)\right),$$

$$\mathrm{Cov}\left(\mu_Y\tilde{(\theta_i)}, \mu_Z(\theta_i)\right) = b\mathrm{Var}\left(\mu_Z(\theta_i)\right) = \mathrm{Cov}\left(\mu_Y(\theta_i), \mu_Z(\theta_i)\right).$$

Solving the above normal equations yields:

$$a = E\left(\mu_Y(\theta_i)\right) - bE\left(\mu_Z(\theta_i)\right),$$

$$b = \frac{\mathrm{Cov}\left(\mu_Y(\theta_i), \mu_Z(\theta_i)\right.}{\mathrm{Var}(\theta_Z(\theta_i))}.$$

To proceed, it is useful to introduce the following simpler notation:

$$\mu_Y = E\left(\mu_Y(\theta_i)\right), \quad \tau_Y = \sqrt{\mathrm{Var}\left(\mu_Y(\theta_i)\right)}, \quad \tau_Z = \sqrt{\mathrm{Var}\left(\mu_Z(\theta_i)\right)},$$

$$\sigma_Z^2 = E\left(\sigma_Z^2(\theta_i)\right), \quad M_i = \sigma_Z^2\left(W_i^T W_i\right)^{-1}, \quad \beta_0 = E(\beta(\theta_i)),$$

$$A_i = \Gamma_i(\Gamma_i + M_i)^{-1}, \quad \mu_Z = E\left(\mu_Z(\theta_i)\right) = E(W_i\beta(\theta_i)),$$

$$\Gamma_i = \mathrm{Cov}\left(\beta(\theta_i), \beta^T(\theta_i)\right), \tau_{YZ} = \mathrm{Cov}\left(\mu_Y(\theta_i), \mu_Z(\theta_i)\right).$$

Given the regression credibility estimator of $Z(\theta_i)$, we can derive the new forecasting estimator $\hat{\mu}_Y^N$ as:

$$\begin{aligned}
\hat{\mu}_Y^N &= b\mu_Z^C + a, \\
&= \mu_Y - \frac{\tau_{YZ}}{\tau_Z^2}\mu_Z + \frac{\tau_{YZ}}{\tau_Z^2}W_i\left(A_i\beta(\theta_i) + (I - A_i)\beta_0\right) \\
&= \mu_Y + \frac{\tau_{YZ}}{\tau_Z^2}\left(W_i A_i\left(\beta(\theta_i) - \beta_0\right)\right), \\
&= \mu_Y + \rho_{YZ}\frac{\tau_Y}{\tau_Z}\left(W_i A_i\left(\beta(\theta_i) - \beta_0\right)\right).
\end{aligned} \tag{9}$$

The above result is important in the introduction of an auxiliary variable $Z(\theta_i)$ with the true model known, which is related to $Y(\theta_i)$ using the correlation coefficients between $Y(\theta_i)$ and $Z(\theta_i)$. In this way, $\hat{\mu}_Y^N$ can be adjusted to reduce the risk of model misspecification. This idea was first proposed by Vylder (1976a, b) and has been applied in health insurance for adjusting large claims. A closer investigation of the expression of Equation (9) shows that when the correlation of $Y(\theta_i)$ and $Z(\theta_i)$ is high, namely, $|\rho_{YZ}|$ is close to 1, the estimator reduces to the classical regression credibility estimator μ_Y^C. On the other hand, if the true model for $Y(\theta_i)$ is far away from the specified regression model, $|\rho_{YZ}|$ will be small, and less weight will be allocated to the regression term.

Additionally, the newly proposed forecasting estimator $\hat{\mu}_Y^N$ has the desirable statistical property of unbiaseness. We also have the following proposition asserting that the new estimator is more efficient in the sense of a smaller mean quadratic loss:

P3.1. When $\lambda = \rho_{YZ}\frac{\tau_Y}{\tau_Z} \in [0, 1]$, the quadratic loss of the estimator $\hat{\mu}_Y^N$ defined in Equation (9) is no greater than that of $\hat{\mu}_Y^C$ defined in Equation (6). In other words:

$$E\left(\hat{\mu}_Y^N - \mu_Y(\theta_i)\right)^2 \leqslant E\left(\hat{\mu}_Y^C - \mu_Y(\theta_i)\right)^2. \tag{10}$$

Proof. Details are provided in Appendix 1.

Table IV summarizes the forecasting results of the new credibility estimator $\hat{\mu}_Y^N$. We compare the forecasting results with the SPCAR, which is shown to have the best out-of-sample forecasting ability, from the previous section, and the classical regression credibility estimator as a benchmark. The second column in Table IV shows the forecasting results for the SPCAR, the third column is for the classical regression credibility estimator, which is denoted as "RegCred," and the last column shows the results for the new credibility estimator, denoted as "NewCred." From both in-sample and out-of-sample errors, the new credibility estimator proposed in this paper has better forecasting.

3.3 Reinsurance pricing formulas and weather risk management

Agriculture insurance and reinsurance ratemaking procedures are commonly based on a random variable called the loss cost ratio (LCR). This ratio is also known as the Burning Ratio in some literature, and is defined as the ratio of indemnities over liabilities in order to normalize the loss exposures (Josephson *et al.*, 2000; Schnapp *et al.*, 2000).

The liability and indemnity for the ith risk category can be expressed as:

$$L(\theta_i) = c \cdot E\left(\mu_Y(\theta_i)\right); \tag{11}$$

$$I(\theta_i) = \max(0, L(\theta_i) - Y(\theta_i)), \tag{12}$$

where c is the coverage level, which in practice is usually 65, 75, 85 percent, etc. We emphasize that the indemnity and the liability are functions of crop yield, and hence a reliable crop yield forecasting model plays a fundamental role in crop insurance and

Statistic	SPCAR	RegCred	NewCred
In-sample errors (MSE)			
Mean	0.0142	0.0182	0.0075
Median	0.0075	0.0074	0.0058
SD	0.0368	0.0379	0.0158
Max	0.3804	0.4194	0.3754
Out-of-sample errors (Loo-CVMSE)			
Mean	0.0342	0.0456	0.0250
Median	0.0158	0.0186	0.0129
SD	0.0845	0.0582	0.0133
Max	0.8467	0.6218	0.4057

Table IV. Summary of credibility forecasting results

Notes: Important statistics are summarized in this table including the mean, median, standard deviation (SD), and maximum errors (Max). "RegCred" represents traditional regression credibility approach and "NewCred" represents the new proposed credibility estimator

reinsurance ratemaking. It follows from the definition of the indemnity and liability that the LCR in risk category i, $X(\theta_i)$, is defined as:

$$X(\theta_i) = \frac{I(\theta_i)}{L(\theta_i)} = \frac{\max(0, L(\theta_i) - Y(\theta_i))}{L(\theta_i)}. \tag{13}$$

While theoretically there are various designs for the reinsurance coverage, the excess of loss (XoL) reinsurance policy is the most common. In addition, the XoL reinsurance policy is found to be the optimal (Tan *et al.*, 2011) reinsurance design, including for agricultural reinsurance (Porth *et al.*, 2013). For these reasons, this paper will focus on the XoL reinsurance policy. An XoL with $A \times B$ structure implies that the insurer cedes the losses in the LCR layer above B, up to a limit of A. More explicitly, the loss random variable of the reinsurer for risk category i, $\pi(\theta_i)$, is given by:

$$\pi(\theta_i) = \min(A, \max(0, X(\theta_i) - B)). \tag{14}$$

The pure net premium of the reinsurer is then the expectation of the corresponding loss random variable; i.e.:

$$\widehat{\pi^N(\theta_i)} = E(\pi(\theta_i)) = E(\min(A, \max(0, X(\theta_i) - B))). \tag{15}$$

In practice, insurance policies are typically priced using some premium principle which takes into consideration the inherent risk, as well as additional expenses such as administration charges, etc. For an extensive list of premium principles, see Young (2004). In this paper, we consider two of the most popular premium principles known as the expectation premium principle (denoted as $\widehat{\pi^E(\theta_i)}$) and the standard deviation principle (denoted as $\widehat{\pi^{SD}(\theta_i)}$):

$$\widehat{\pi^E(\theta_i)} = E(\pi(\theta_i))(1 + \eta_1), \tag{16}$$

$$\widehat{\pi^{SD}(\theta_i)} = E(\pi(\theta_i)) + \eta_2 \sqrt{\mathrm{Var}(\pi(\theta_i))}. \tag{17}$$

where η_1 and η_2 are the respective loading coefficients. *P3.2* provides reinsurance pricing formulas for the two premium principles. The closed form formulas facilitate agricultural reinsurance pricing, and also assist the reinsurance companies in making sound and scientific weather management strategies:

P3.2. Let $\mu = E(\mu_Y(\theta_i))$, $\sigma^2 = Var(\mu_Y(\theta_i))$, and further assume that in Equation (5), ε, is normally distributed. Let $K_1 = c(1-B)$, $K_2 = c(1-A-B)$, $\Phi(\cdot)$ and $\phi(\cdot)$ be the cdf and pdf of the normal distribution, respectively. Then the reinsurance premium for the XoL reinsurance policy ($A \times B$), with coverage level c under the net premium principle, the expectation premium principle, and the standard deviation principle, respectively, are:

$$\widehat{\pi^N(\theta_i)} = \left(\frac{\sigma}{c\mu}\right)\left[\phi\left(\frac{(K_1-1)\mu}{\sigma}\right) + \left(\frac{(K_1-1)\mu}{\sigma}\right)\Phi\left(\frac{(K_1-1)\mu}{\sigma}\right) - \phi\left(\frac{(K_2-1)\mu}{\sigma}\right) - \left(\frac{(K_2-1)\mu}{\sigma}\right)\Phi\left(\frac{(K_2-1)\mu}{\sigma}\right)\right], \tag{18}$$

$$:= M_1, \tag{19}$$

$$\widehat{\pi^{E}(\theta_i)} = M_1(1+\eta_1); \tag{20}$$

$$\widehat{\pi^{SD}(\theta_i)} = M_1+\eta_2\sqrt{M_2-(M_1)^2}, \tag{21}$$

where:

$$M_2 = \left(\frac{\sigma}{c\mu}\right)^2\left\{\left[\left(\frac{(K_1-1)\mu}{\sigma}\right)^2+1\right]\Phi\left(\frac{(K_1-1)\mu}{\sigma}\right)+\left(\frac{(K_1-1)\mu}{\sigma}\right)\phi\left(\frac{(K_1-1)\mu}{\sigma}\right)\right.$$
$$+\left[\left(\frac{(K_2-1)\mu}{\sigma}\right)^2-2\left(\frac{(K_1-1)\mu}{\sigma}\right)\left(\frac{(K_2-1)\mu}{\sigma}\right)-1\right]\Phi\left(\frac{(K_2-1)\mu}{\sigma}\right)$$
$$\left.+\left(\frac{(K_2-1)\mu}{\sigma}-2\frac{(K_1-1)\mu}{\sigma}\right)\phi\left(\frac{(K_2-1)\mu}{\sigma}\right)\right\}, \tag{22}$$

Proof. Details are provided in Appendix 2.

Note that in the pricing formulas, the reinsurance premium can be expressed as a function of weather indices, $\pi(\mathbf{W}) = \pi(W_1, \ldots, W_p) : \mathbb{R}^p \mapsto \mathbb{R}$, where $\boldsymbol{W}$ is a p-dimensional weather index. The gradient vector of $\pi(\boldsymbol{W})$ is expressed as:

$$\nabla\pi(\boldsymbol{W}) = \left(\frac{\partial\pi}{\partial W_1}, \ldots, \frac{\partial\pi}{\partial W_p}\right)'.$$

It follows from the first-order Taylor expansion at an arbitrary point $\boldsymbol{W}_0$ that:

$$\pi(\boldsymbol{W}) = \pi(\boldsymbol{W}_0)+\nabla\pi(\boldsymbol{W})\big|_{W=W_0}(\boldsymbol{W}-\boldsymbol{W}_0).$$

Under the above linear approximation, the change in premium, as triggered by the weather indexes, is proportional to the gradient vector $\nabla\pi(\boldsymbol{W})$. Consequently, the gradient vector is a measure of the sensitivity of the premium with respect to the weather indices, and this further helps to measure the impact of weather risk on incomes, revenues, margins, and profits of the reinsurance companies. This pricing framework may provide reinsurers with the expertise to develop tailored programs for different weather-sensitive parts of their risk portfolio. Through appropriately managing the overall critical weather risk factors of the corporation, risk managers are able to make improved high-level strategic decisions that balance the risk vs return tradeoff.

4. Conclusion remarks

Weather risks are cited as one of the major difficulties having a substantial impact on the sustainability of agricultural insurance and reinsurance business. Faced with current challenges, such as potentially more frequent and severe extreme weather events, including drought and flood, insurers, and reinsurers with specialized agricultural business units must invest in improved predictive analytics in order to advance their weather risk modeling, pricing, reserving, and other risk management applications.

In this paper, a new crop yield forecasting model is developed to integrate weather variables and crop production information from different geographically correlated regions using a new credibility estimator, and closed form pricing formulas. A yield restatement approach to account for changing crop mix through time is demonstrated. The model is empirically analyzed based on detailed farm-level data from Manitoba, Canada, covering

216 crop varieties from 19,238 farms from 1996 to 2011, as well as corresponding weather data from 30 stations, including daily temperature and precipitation. Three dimension reduction and selection algorithms are analyzed, that combine SR, CV, and PCA. The results show that the forecasting model provides significant improvements over the classical regression model, both in terms of in-sample and out-of-sample forecasting abilities. This research is useful from a risk management perspective for insurers and reinsurers, and the framework may also be used to develop improved weather risk management strategies to help manage adverse weather events.

Note

1. For more detailed definitions, refer to Alatona *et al.* (2002); Campbell and Diebold (2005); Alexandridis and Zapranis (2013).

References

Akaike, H. (1974), "A new look at the statistical model identification", *IEEE Transaction on Automatic Control*, Vol. 19 No. 6, pp. 716-723.

Alatona, P., Djehicheb, B. and Stillbergerc, D. (2002), "On modelling and pricing weather derivatives", *Applied Mathematical Finance*, Vol. 9 No. 1, pp. 1-20.

Alexandridis, A.K. and Zapranis, A.D. (2013), *Weather Derivatives: Modeling and Pricing Weather-Related Risk*, Springer.

Brown, D.W. (1975), "Heat units for corn in Southern Ontario", Factsheet Ministry of Agriculture and Food Ontario No. 111/31, pp. 1–6.

Bühlmann, H. (1997), "Credibility in the regression case revisited", *ASTIN Bulletin*, Vol. 27 No. 1, pp. 83-98.

Bühlmann, H. and Gisler, A. (2005), *A Course in Credibility Theory and its Applications*, Springer.

Bühlmann, H. and Straub, E. (1970), "Glaubwürdigkeit für Schadensätze", *Bulletin of Swiss Arbitration Association*, Vol. 70, pp. 111-133.

Cai, R., Mullen, J.D., Bergstrom, J.C., Shurley, W.D. and Wetzstein, M.E. (2013), "Using a climate index to measure crop yield response", *Journal of Agricultural and Applied Economics*, Vol. 45 No. 4, pp. 719-737.

Campbell, S.D. and Diebold, F.X. (2005), "Weather forecasting for weather derivatives", *Journal of the American Statistical Association*, Vol. 100 No. 469, pp. 6-16.

Cassman, K.G. (1999), "Ecological intensification of cereal production systems: yield potential, soil quality, and precision agriculture", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 96 No. 11, pp. 5952-5959.

Chambers, R.G. (1989), "Insurability and moral hazard in agricultural insurance markets", *American Journal of Agriculture Economics*, Vol. 71 No. 3, pp. 604-616.

Coble, K.H., Harri, A., Anderson, J.D., Ker, A.P. and Goodwin, B. (2008), "Review of county yield trending procedures and related topics", technical report, USDA Risk Management Agency.

Coble, K.H., Miller, M.F., Rejesus, R.M., Boyles, R., Knight, T.O. and Goodwin, B.K. (2011), "Methodology analysis for weighting of historical experience", technical report, USDA Risk Management Agency.

Dai, A., Trenberth, K.E. and Qian, T. (2004), "A global data set of palmer drought severity index for 1870–2002: relationship with soil moisture and effects of surface warming", *Journal of Hydrometeorology*, Vol. 5, pp. 1117-1130.

Hachemeister, C.A. (1975), "Credibility for regression models with application to trend", in Kahn., P.M. (Ed.), *Credibility: Theory and Applications*, Academic Press, Inc., New York, NY, pp. 129-163.

Jolliffe, I.T. (2002), *Principal Component Analysis*, 2nd ed., John Wiley & Sons, Ltd.

Josephson, G.R., Lord, R.B.L. and Mitchell, C.W. (2000), "Actuarial documentation of multiple peril crop insurance ratemaking procedures", prepared for USDA/Risk Management Agency by Milliman & Robertson, Inc., available at: www.rma.usda.gov/pubs/2000/000801_mpci_rate.pdf

Kleibera, W., Rafterya, A.E. and Gneitinga, T. (2011), "Geostatistical model averaging for locally calibrated probabilistic quantitative precipitation forecasting", *Journal of the American Statistical Association*, Vol. 106 No. 496, pp. 1291-1303.

Kohavi, R. (1995), "A study of cross-validation and bootstrap for accuracy estimation and model selection", *International Joint Conference on Artificial Intelligence*, Vol. 14 No. 2, pp. 1137-1145.

Lobell, D.B. and Asner, G.P. (2003), "Climate and management contributions to recent trends in US agricultural yields", *Science*, Vol. 299 No. 5609, pp. 1032-1032.

Lobell, D.B. and Burke, M.B. (2010), "On the use of statistical models to predict crop yield responses to climate change", *Agriculture and Forest Meteorology*, Vol. 150 No. 11, pp. 1443-1452.

Mallows, C.L. (2000), "Some comments on c_p", *Technometrics*, Vol. 15 No. 1, pp. 661-676.

Matheron, G. (1963), "Principles of geostatistics", *Economic Geology*, Vol. 58 No. 8, pp. 1246-1266.

McCarl, B.A., Villavicencio, X. and Wu, X. (2008), "Climate change and future analysis: is stationary dying?", *American Journal of Agriculture Economics*, Vol. 90 No. 5, pp. 1241-1247.

Nelson, C.H. and Loehman, E.T. (1987), "Further toward a theory of agricultural insurance", *American Journal of Agriculture Economics*, Vol. 69 No. 3, pp. 523-531.

Ozaki, V.A., Goodwin, B.K. and Shirota, R. (2008), "Parametric and nonparametric statistical modelling of crop yield: implications for pricing crop insurance contracts", *Applied Economics*, Vol. 40 No. 9, pp. 1151-1164.

Pai, J., Boyd, M. and Porth, L. (2014), "Insurance premium calculation using credibility analysis: an example from livestock mortality insurance", *Journal of Risk and Insurance*, Vol. 82 No. 2, pp. 341-357.

Plant, R.E. (2012), *Spatial Data Analysis in Ecology and Agriculture Using R*, CRC Press.

Porth, L., Tan, K.S. and Weng, C. (2013), "Optimal reinsurance analysis from a crop insurer's perspective", *Agriculture Finance Review*, Vol. 73 No. 2, pp. 310-328.

Porth, L., Zhu, W. and Tan, K.S. (2014), "A credibility-based Erlang mixture model for pricing crop reinsurance", *Agricultural Finance Review*, Vol. 74 No. 2, pp. 162-187.

Quiggen, J., Karagiannis, G. and Stanton, J. (1993), "Crop insurance and crop production: an empirical study of moral hazard and adverse selection", *Economics of Agricultural Crop Insurance*, Vol. 37 No. 2, pp. 95-113.

Ramsey, F.L. and Schafer, D.W. (2012), *The Statistical Sleuth: A Course in Methods of Data Analysis*, 3rd ed., Cengage Learning.

Schnapp, F., Driscoll, J., Zacharias, T. and Josephson, R. (2000), "Ratemaking considerations for multiple peril crop insurance", report prepared for USDA/Risk Management Agency.

Schwarz, G.E. (1978), "Estimating the dimension of a model", *Annual of Statistics*, Vol. 6 No. 2, pp. 461-464.

Sherrick, B., Zanini, F., Schnitkey, G. and Irwin, S. (2004), "Crop insurance valuation under alternative yield distributions", *American Journal of Agricultural Economics*, Vol. 86 No. 2, pp. 406-419.

Shi, W., Tao, F. and Zhang, Z. (2013), "A review on statistical models for identifying climate contributions to crop yields", *Journal of Geographical Sciences*, Vol. 23 No. 3, pp. 567-576.

Tan, K.S., Weng, C. and Zhang, Y. (2011), "Optimality of general reinsurance contracts under CTE risk measure", *IME*, Vol. 49 No. 2, pp. 175-187.

Vylder, F.D. (1976a), "Geometrical credibility", *Scandinavian Actuarial Journal*, pp. 121-149.

Vylder, F.D. (1976b), "Optimal semilinear credibility", *Bulletin of Swiss Arbitration Association*, pp. 27-40.

Wang, H.H. and Zhang, H. (2003), "On the possibility of a private crop insurance market: a spatial statistics approach", *Journal of Risk and Insurance*, Vol. 70 No. 1, pp. 111-124.

Woodard, J.D. (2014), "A conditional distribution approach to assessing the impact of weather sample heterogeneity on yield risk estimation and crop insurance ratemaking", *North American Actuarial Journal*, Vol. 18 No. 2, pp. 279-293.

Woodard, J.D. and Garcia, P. (2008), "Weather derivatives, spatial aggregation, and systemic risk: implications for reinsurance hedging", *Journal of Agricultural and Resource Economics*, Vol. 33 No. 1, pp. 34-51.

Woodard, J.D. and Sherrick, B.J. (2011a), "Actuarial impacts of loss cost ratio ratemaking in US crop insurance programs", *Journal of Agriculture and Resource Economics*, Vol. 36 No. 1, pp. 211-228.

Woodard, J.D. and Sherrick, B.J. (2011b), "Estimation of mixture models using cross-validation optimization: implications for crop yield distribution modeling", *American Journal of Agricultural Economics*, Vol. 93 No. 4, pp. 968-982.

Young, V.R. (2004), "Premium principles", in Teugels, J. and Sundt, B. (Eds), *Encyclopedia of Actuarial Science*, John Wiley & Sons, West Sussex, Vol. 3, pp. 1323-1331.

Further reading

Carriquiry, M.A. and Osgood, D.E. (2012), "Index insurance, probabilistic climate forecasts, and production", *Journal of Risk and Insurance*, Vol. 79 No. 1, pp. 287-300.

Conway, G. and Toenniessen, G. (1999), "Feeding the world in the twenty-first century", *Nature*, Vol. 402 No. 6761, pp. C55-C58.

Harri, A., Erdem, C., Keith, H.C. and Thomas, O.K. (2009), "Crop yield distributions: a reconciliation of previous research and statistical tests for normality", *Applied Economic Perspectives and Policy*, Vol. 31 No. 1, pp. 163-182.

Insurance Bureau of Canada (IBC) (2014), "2014 facts of the property & casualty insurance industry", Fact Book, Insurance Bureau of Canada.

Ito, H., Ai, J. and Ozawa, A. (2015), "Managing weather risks: the case of J. League soccer teams in Japan", *Journal of Risk and Insurance*, Vol. 83 No. 4, pp. 877-912.

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

Motha, R.P. and Baier, W. (2005), "Impacts of present and future climate change and climate variability on agriculture in the temperature regions: North America", *Climate Change*, pp. 137-164.

Okhrin, O., Odening, M. and Xu, W. (2013), "Systemic weather risk and crop insurance: the case of China", *Journal of Risk and Insurance*, Vol. 80 No. 2, pp. 351-372.

Pall, P., Aina, T., Stone, D.A., Stott, P.A., Nozawa, T., Hilberts, A.G.J., Lohmann, D. and Allen, M.R. (2011), "Anthropogenic greenhouse gas contribution to flood risk in England and Wales in Autumn 2000", *Nature*, Vol. 470 No. 7334, pp. 382-385.

Salinger, M. (2005), "Climate variability and change: past, present and future: an overview", *Climate Change*, Vol. 70 Nos 1/2, pp. 9-29.

Appendix 1. Proof of *P3.1*

Let us start the proof by defining:

$$QL^N = E\left(\hat{\mu}_Y^N - \mu_Y(\theta_i)\right)^2$$

$$QL^C = E\left(\hat{\mu}_Y^C - \mu_Y(\theta_i)\right)^2,$$

then:

$$QL^N = E\left(\mu_Y - \mu_Y(\theta_i) + \lambda\left(\hat{\mu}_Y^C - \mu_Z\right)\right)^2$$

$$= E\left(\hat{\mu}_Y^C - \mu_Y(\theta_i) + \lambda\left(\hat{\mu}_Y^C - \mu_Z\right) + \left(\mu_Y - \hat{\mu}_Y^C\right)\right)^2$$

$$= E\left(\hat{\mu}_Y^C - \mu_Y(\theta_i)\right)^2 + \lambda^2 E\left(\hat{\mu}_Y^C - \mu_Z\right)^2$$
$$+ E\left(\mu_Y - \hat{\mu}_Y^C\right)^2 + 2\lambda E\left(\hat{\mu}_Y^C - \mu_Y(\theta_i)\right)\left(\hat{\mu}_Y^C - \mu_Z\right)$$
$$+ 2\lambda E\left(\hat{\mu}_Y^C - \mu_Z\right)\left(\mu_Y - \hat{\mu}_Y^C\right) + 2E\left(\hat{\mu}_Y^C - \mu_Y(\theta_i)\right)\left(\mu_Y - \hat{\mu}_Y^C\right)$$

Note that $E(\hat{\mu}_Y^C) = E[E(\hat{\mu}_Y^C|\theta_i)] = E(\mu_Z(\theta_i)) = \mu_Z$, hence:

$$E\left(\hat{\mu}_Y^C - \mu_Z\right)^2 = Var\left(\hat{\mu}_Y^C\right)$$

$$E\left(\mu_Y - \hat{\mu}_Y^C\right)^2 = Var\left(\hat{\mu}_Y^C\right) + (\mu_Y - \mu_Z)^2$$

also:

$$2\lambda\left(E\left(\hat{\mu}_Y^C - \mu_Y(\theta_i)\right)\left(\hat{\mu}_Y^C - \mu_Z\right) + E\left(\hat{\mu}_Y^C - \mu_Z\right)\left(\mu_Y - \hat{\mu}_Y^C\right)\right)$$

$$= 2\lambda E\left(\hat{\mu}_C^Y - \mu_Z\right)\left(\mu_Y - \mu_Y(\theta_i)\right)$$

$$= 2\lambda E\left(\left(\mu_Y - \mu_Y(\theta_i)\right)E\left(\hat{\mu}_Y^C - \mu_Z|\theta_i\right)\right)$$

$$= 0,$$

and:

$$E\left(\hat{\mu}_Y^C - \mu_Y(\theta_i)\right)\left(\mu_Y - \hat{\mu}_Y^C\right)$$

$$= E\left(\mu_Y(\theta_i) - \mu_Y\right)\left(\hat{\mu}_Y^C - \mu_Y\right) + E\left(\mu_Y - \hat{\mu}_Y^C\right)\left(\hat{\mu}_Y^C - \mu_Y\right)$$

$$= -E\left(\hat{\mu}_Y^C - \mu_Y\right)^2$$

$$= -Var\left(\hat{\mu}_Y^C\right) - (\mu_Y - \mu_Z)^2.$$

Therefore, the quadratic loss of estimator $\hat{\mu}_Y^N$ can be expressed as:

$$QL^N = QL^C - (1-\lambda^2)\mathrm{Var}\left(\hat{\mu}_Y^C\right) - (\mu_Y - \mu_Z)^2.$$

Namely, if $\lambda = \rho_{YZ}(\sigma_Y/\sigma_Z) \in [0,1]$, we have $QL^N < QL^C$, and Relation (10) holds.

Appendix 2. Proof of *P3.2*

Since:

$$\pi(\theta_i) = \min(A, \max(0, X(\theta_1) - B));$$
$$= \max(X(\theta_i) - B, 0) - \max(X(\theta_i) - (A+B), 0),$$

where $X(\theta_i)$ is defined as Equation 13. Therefore:

$$\pi(\theta_i) = \frac{1}{L(\theta_i)}[\max(L(\theta_i)(1-B)-Y(\theta_i),0) - \max(L(\theta_i)(1-(A+B))-Y(\theta_i),0)]$$

$$= \max\left(1-B-\frac{Y(\theta_i)}{L(\theta_i)},0\right) - \max\left(1-(A+B)-\frac{Y(\theta_i)}{L(\theta_i)},0\right)$$

In general, for a random variable $Y \sim N(\mu, \sigma)$, we have the following values:

$$E\left(1_{[Y<H]}\right) = \Phi\left(\frac{H-\mu}{\sigma}\right),$$

$$E\left(Y1_{[Y<H]}\right) = \mu\Phi\left(\frac{H-\mu}{\sigma}\right) - \sigma\phi\left(\frac{H-\mu}{\sigma}\right),$$

$$E\left(Y^2 1_{[Y<H]}\right) = (\sigma^2+\mu^2)\Phi\left(\frac{H-\mu}{\sigma}\right) - \sigma(H+\mu)\phi\left(\frac{H-\mu}{\sigma}\right),$$

$$E\left((H-Y)^2 1_{[Y<H]}\right) = \left[(H-\mu)^2+\sigma^2\right]\Phi\left(\frac{H-\mu}{\sigma}\right) + \sigma(H-\sigma)\phi\left(\frac{H-\mu}{\sigma}\right),$$

$$E\left((H_1-Y)(H_2-Y)1_{[Y<H_2]}\right) = \left[(H_1-\mu)(H_2-\mu)+\sigma^2\right]\Phi\left(\frac{H_2-\mu}{\sigma}\right) + \sigma(H_1-\mu)\phi\left(\frac{H_2-\mu}{\sigma}\right),$$

where 1_A is indicator function for event A, H,H_1,H_2 are constants and $H_1 > H_2$.

Therefore:

$$E(\pi(\theta_i)) = (1-B)E\left(1_{\left[\frac{Y(\theta_i)}{L(\theta_i)}<1-B\right]}\right) - E\left(\frac{Y(\theta_i)}{L(\theta_i)}1_{\left[\frac{Y(\theta_i)}{L(\theta_i)}<1-B\right]}\right)$$

$$-(1-A-B)E\left(1_{\left[\frac{Y(\theta_i)}{L(\theta_i)}<1-A-B\right]}\right) - E\left(\frac{Y(\theta_i)}{L(\theta_i)}1_{\left[\frac{Y(\theta_i)}{L(\theta_i)}<1-A-B\right]}\right);$$

$$= \left\{\frac{\sigma}{c\mu}\phi\left(\left(\frac{(K_1-1)\mu}{\sigma}\right)\right) - \frac{1-K_1}{c}\Phi\left(\frac{(K_1-1)\mu}{\sigma}\right)\right\}$$

$$-\left\{\frac{\sigma}{c\mu}\phi\left(\frac{(K_2-1)\mu}{\sigma}\right) - \frac{1-K_2}{c}\Phi\left(\frac{(K_2-1)\mu}{\sigma}\right)\right\},$$

which is exactly Equation (18). Meanwhile:

$$E\left[(\pi(\theta_i))^2\right] = E\left(\left(1-B-\frac{Y(\theta_i)}{L(\theta_i)}\right)^2 1_{\left[\frac{Y(\theta_i)}{L(\theta_i)}<1-B\right]}\right) + E\left(\left(1-A-B-\frac{Y(\theta_i)}{L(\theta_i)}\right)^2 1_{\left[\frac{Y(\theta_i)}{L(\theta_i)}<1-A-B\right]}\right)$$

$$-2E\left(\left(1-B-\frac{Y(\theta_i)}{L(\theta_i)}\right)\left(1-A-B-\frac{Y(\theta_i)}{L(\theta_i)}\right)1_{\left[\frac{Y(\theta_i)}{L(\theta_i)}<1-A-B\right]}\right),$$

$$= \left(\frac{\sigma}{c\mu}\right)^2 \left\{ \left[\left(\frac{(K_1-1)\mu}{\sigma}\right)^2 + 1 \right] \Phi\left(\frac{(K_1-1)\mu}{\sigma}\right) + \left(\frac{(K_1-1)\mu}{\sigma}\right) \phi\left(\frac{(K_1-1)\mu}{\sigma}\right) \right.$$
$$+ \left[\left(\frac{(K_2-1)\mu}{\sigma}\right)^2 - 2\left(\frac{(K_1-1)\mu}{\sigma}\right)\left(\frac{(K_2-1)\mu}{\sigma}\right) - 1 \right] \Phi\left(\frac{(K_2-1)\mu}{\sigma}\right)$$
$$\left. + \left(\frac{(K_2-1)\mu}{\sigma} - 2\frac{(K_1-1)\mu}{\sigma}\right) \phi\left(\frac{(K_2-1)\mu}{\sigma}\right) \right\},$$

denote:

$$M_1 = \left(\frac{\sigma}{c\mu}\right) \left[\phi\left(\frac{(K_1-1)\mu}{\sigma}\right) + \left(\frac{(K_1-1)\mu}{\sigma}\right) \Phi\left(\frac{(K_1-1)\mu}{\sigma}\right) \right.$$
$$\left. - \phi\left(\frac{(K_2-1)\mu}{\sigma}\right) - \left(\frac{(K_2-1)\mu}{\sigma}\right) \Phi\left(\frac{(K_2-1)\mu}{\sigma}\right) \right],$$

$$M_2 = \left(\frac{\sigma}{c\mu}\right)^2 \left\{ \left[\left(\frac{(K_1-1)\mu}{\sigma}\right)^2 + 1 \right] \Phi\left(\frac{(K_1-1)\mu}{\sigma}\right) + \left(\frac{(K_1-1)\mu}{\sigma}\right) \phi\left(\frac{(K_1-1)\mu}{\sigma}\right) \right.$$
$$+ \left[\left(\frac{(K_2-1)\mu}{\sigma}\right)^2 - 2\left(\frac{(K_1-1)\mu}{\sigma}\right)\left(\frac{(K_2-1)\mu}{\sigma}\right) - 1 \right] \Phi\left(\frac{(K_2-1)\mu}{\sigma}\right)$$
$$\left. + \left(\frac{(K_2-1)\mu}{\sigma} - 2\frac{(K_1-1)\mu}{\sigma}\right) \phi\left(\frac{(K_2-1)\mu}{\sigma}\right) \right\},$$

then, $\mathrm{Var}(\pi(\theta_i)) = M_2 - (M_1)^2$, and Equation (21) holds obviously.

Corresponding author
Lysa Porth can be contacted at: lysa.porth@umanitoba.ca

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**