Actuarial Ratemaking in Agricultural Insurance

by

Wenjun Zhu

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor in Philosophy
in
Actuarial Science

Waterloo, Ontario, Canada, 2015

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

A scientific agricultural (re)insurance pricing approach is essential for maintaining sustainable and viable risk management solutions for different stakeholders including farmers, governments, insurers, and reinsurers. The major objective of this thesis is to investigate high dimensional solutions to refine the agricultural insurance and reinsurance pricing. In doing so, this thesis develops and evaluates several high dimensional approaches for constructing actuarial ratemaking framework for agricultural insurance and reinsurance, including two credibility approaches, a high dimensional copula approach, and a multivariate weighted distribution approach.

This thesis comprehensively examines the ratemaking process, including reviews of different detrending methods and the generating process of the historical loss cost ratio's (LCR's, which is defined as the ratio of indemnities to liabilities). A modified credibility approach is developed based on the Erlang mixture distribution and the liability weighted LCR. In the empirical analysis, a comprehensive data set representing the entire crop insurance sector in Canada is used to show that the Erlang mixture distribution captures the tails of the data more accurately compared to conventional distributions. Further, the heterogeneous credibility premium based on the liability weighted LCR's is more conservative, and provides a more scientific approach to enhance the reinsurance pricing.

The agriculture sector relies substantially on insurance and reinsurance as a mechanism to spread loss. Climate change may lead to an increase in the frequency and severity of spatially correlated weather events, which could lead to an increase in insurance costs, or even the unavailability of crop insurance in some situations. This could have a profound impact on crop output, prices, and ultimately the ability to feed the world rowing population into the future. This thesis proposes a new reinsurance pricing framework, including a new crop yield forecasting model that integrates weather and crop production information from different risk geographically related regions, and closed form reinsurance pricing formulas. The framework is empirically analyzed, with an original weather index system we set up, and algorithms that combine screening regression (SR), cross validation (CV) and principle component analysis (PCA) to achieve efficient dimension reduction and model selection. Empirical results show that the new forecasting model has improved both in-sample and out-of-sample forecasting abilities. Based on this framework, weather risk management strategies are provided for agricultural reinsurers.

Adverse weather related risk is a main source of crop production loss, and in addition to farmers, this exposure is a major concern to insurers and reinsurers who act as weather risk underwriters. To date, weather hedging has had limited success, largely due to challenges

regarding basis risk. Therefore, this thesis develops and compares different weather risk hedging strategies for agricultural insurers and reinsurers, through investigating the spatial dependence and aggregation level of systemic weather risks across a country. In order to reduce basis risk and improve the efficiency of weather hedging strategies, this thesis refines the weather variable modeling by proposing a flexible time series model that assumes a general hyperbolic (GH) family for the margins to capture the heavy-tail property of the data, together with the Lévy subordinated hierarchical Archimedean copula (LSHAC) model to overcome the challenge of high-dimensionality in modeling the dependence of weather risk. Wavelet analysis is employed to study the detailed characteristics within the data from both time and frequency scales. Results show that it is of great importance of capturing the appropriate dependence structure of weather risk. Further, the results reveal significant geographical aggregation benefits in weather risk hedging, which means that more effective hedging may be achieved as the spatial aggregation level increases.

It has been discussed that it is necessary to integrate auxiliary variables such as weather, soil, and other information into the ratemaking system to refine the pricing framework. In order to investigate a possible scientific way to reweight historical loss data with auxiliary variables, this thesis proposes a new premium principle based on multivariate weighted distribution. Some designable properties such as linearity and stochastic order preserving are derived for the new proposed multivariate weighted premium principle. Empirical analysis using a unique data set of the reinsurance experience in Manitoba from 2001 to 2011 compares different premium principles and shows that integrating auxiliary variables such as liability and economic factors into the pricing framework will redistribute premium rates by assigning higher loadings to more risky reinsurance contracts, and hence help reinsurers achieve more sustainable profits in the long term.

Acknowledgements

I would like to sincerely express my deepest appreciation to my supervisors: Dr. Ken Seng Tan and Dr. Lysa Porth. Dr. Tan you have been inspiring me, with your spirit of adventure and attribute of a genius, to be a researcher with curiosities about the unknown and thank you for providing me such a precious opportunity to be exposed to scientific research. Dr. Porth, you have enthusiastically encouraged me, and provided me with unreserved support and insightful guidance. Without both of your supervision this dissertation would not have been possible.

I would also like to pass my gratitude to my committee members, Dr. Mary Hardy and Dr. Adam Kolkiewicz, for the kind willingness to offer your expert opinions and helpful suggestions. I am always grateful for your time and availability to discuss my questions. In addition, a sincere thank you to Dr. Calum Turvey for your kind willingness to be my external committee member and for the time you have dedicated to review my thesis. I also thank Dr. Jason Thistlethwaite for acting as my internal-external committee member.

A special thank you as well to Dr. Joshua Woodard for your inspiration with your enthusiasm for research and your good work, and for providing me with an opportunity to collaborate with you.

I would also like to thank Dr. Phelim Boyle, Dr. Chengguo Weng and Dr. Tony Wirjanto from our department for sparking ideas and providing suggestions regarding my research work.

I would also acknowledge funding support from the Society of Actuaries Hickman Scholar Program and the China Scholarship Council. I am also grateful to Manitoba Agriculture Services Corporation (MASC) for their assistance in obtaining the data.

My heart and soul are with my beloved mum and dad. Thank you for all the miracles, joy, peace, and happiness you have given me in my life. Without your unconditional support and understanding, I would not keep my persevering and hard work. I would also like to thank all of my friends who supported me in writing this thesis. At the end I would like to express my thanks to my boyfriend, who I met at the right time in the correct place, who has spent many sleepless nights with me working on the thesis and who has always been encouraging me to strive towards my dreams.

Dedication

In memory of my great-grandmother. You left fingerprint of grace and blessings on our lives. Your love and sacrifice for the family will never be forgotten.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

It is estimated that global agricultural production must increase by 60% to feed the world's population, which will reach 9 billion by 2050. The recent "Sigma" report by Swiss Re (2013) emphasizes that agriculture insurance is an indispensable part of agricultural risk management and helps to smooth farm income as well as to promote/encourage food investment. Furthermore, the study also suggests that the presentation of agriculture insurance in emerging markets is currently very low but potential premiums by 2025 may reach an estimated USD 14 -19 billion, representing 3-4 fold increases from the 2011 figure.

The most common crop insurance program, Multiple Peril Crop Insurance (MPCI), serves to provide financial protection to farmers from yield risks as a result of natural disasters such as droughts, insects, hurricane, etc. Private reinsurance is an essential part of a sound agricultural insurance system, largely due to challenges of managing losses that often spatially correlated. The reinsurance arrangement, like the Standard Reinsurance Agreement (SRA) in the U.S. and the Federal-Provincial Reinsurance Fund in Canada, help to encourage the participation of private reinsurance and protect the insurers from catastrophic losses. A sound and scientific rating approach for agricultural insurance and reinsurance ratemaking is essential in maintaining a sustainable program in the long run. From an actuarial point of view, the rating problem is equivalent to finding the proper rate based on available historical observations. To be more specific, given a loss random variable $\mathbf{X} \in \mathbb{R}^{d\times n}$

$$\mathbf{X} = \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2n} \\ \vdots & \vdots & \ddots & \\ X_{d1} & X_{n2} & \ddots & X_{dn} \end{pmatrix}$$

with observations of n years from d risk sectors, $x \in \mathbb{R}^{n \times d}$,

$$\mathbf{x} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \\ x_{d1} & x_{d2} & \ddots & x_{dn} \end{pmatrix}$$

actuaries are involved in estimating the parameter of the loss distribution of X in order to assign appropriate premium rates to each risk sector.

In actuarial literature, an "independent and identical distribution assumption" is often assumed, which considerably simplifies the underlying problem. This assumption can be interpreted in at least two ways:

- The loss is called *homogeneous with no contagion in the mass of risks* if $X_{1i}, X_{2i}, \dots, X_{di}$ are independently and identically distributed for all fixed i.
- The loss is called *homogeneous with no contagion in time* if $X_{j1}, X_{j2}, \dots, X_{jn}$ are independently and identically distributed for all fixed j.

Relaxing the assumptions of independence and homogeneity, either (or both) of the above statements introduces more challenges in estimation, but potentially makes the resulting actuarially rates more appropriate. Given the special risk characteristics of the agricultural insurance and reinsurance, for instance, large exposure to natural catastrophes and at times spatial correlations, special considerations must be given to the ratemaking process compared to more typical property & casualty procedures. To be more specific, serial correlations and trends are introduced if we drop the independent and homogeneity assumption in time, and will add challenges in the estimation of loss process (such as crop yield process) distributions. Meanwhile, dropping the homogeneous assumption will introduce geographical dependence or inter-business correlations to the loss data.

The main objective of this thesis is to address some of the outstanding, yet essential, issues in agricultural ratemaking, particularly from an actuarial point of view and with special attention on investigating high dimensional solutions for the construction of a scientific and validated ratemaking framework for agricultural insurance and reinsurance.

1.1 Yield Distribution

Crop yield distribution estimation and forecasting lies at the heart of agricultural insurance program development. A scientific and accurate yield distribution and forecasting model helps ensure the resulting premium rates are actuarially fair, since it gives better predictions of the expected loss as well as the yield shortfalls, which directly relates to the loss distributions. Technological developments and agronomic advancements are thought to increase the average and possibly reduce variance of the current yields compared to the past. On the other hand, severe weather may be increasing in both frequency and severity, which would increase the overall risk portfolio of the insurers and reinsurers. This thesis proposes the Erlang mixture to model the loss process of crop insurance program in Canada from the prospective of the reinsurer. Our analysis suggests that compared to parametric distributions commonly used in agriculture economics, our proposed framework provides better fitting results and leads to more conservative pricing method for reinsurance companies.

1.2 Credibility Rating

In agriculture insurance, portfolio risks are diverse due to widespread geographical losses, climate influence, moral hazards, etc. Despite its complexity, the ratemaking approach adopted is rather naive. For example, simple premium discounts and surcharges will be applied to differentiate customers with different historical claims. A relative index, called the Management Experience Transfer (MET) index in Saskatchewan and the Individual Productivity Index (IPI) in Manitoba, is used to measure the success of the individual farmer. The index is greater than 1 if the yields of the insured are better than the area average, smaller than 1 if the yields are lower than the area average, and 1 if no historical yield data is available for a certain farmer. In general, the MPCI premium is calculated as

$$\text{APH yields} \times \text{Acres planted} \times \text{Coverage level} \times \text{Crop Base Price} \times \text{Price election percentage},$$

where the APH yields stands for Actual Production Historical yields, and is based on 4 to 10 years of historical yield data; the coverage level represents the deductible level, and usually ranges from 50% to 85%; the crop base price is usually set by the government agency (e.g., Risk Management Agency (RMA) of United States Department of Agriculture (USDA) in the States) at the beginning of the growing season based on the current market price information (Josephson et al., 2000).

Credibility theory has proven to be a useful experience-based ratemaking tool in pricing Property & Casualty (P&C) insurance policies. Credibility rating takes into consideration the heterogeneity of loss data in time and different insured (or risk) groups. Classical credibility theory starts from the publication of Bühlmann (1967). A series of extensions based on this model have emerged since then, the most widely applied models include the Bühlmann-Straub model (Bühlmann and Straub, 1970) and the regression credibility model (Hachemeister and Kahn, 1975). Recently, Wen and Wu (2011) propose a credibility model with a general dependence structure over risks. Credibility approach is applied to mortality risk modeling by Hardy and Panjer (1998) to derive a methodology to calculate the adverse deviation margin added to mortality rate. Pai et al. (2014) refer to a Bayesian credibility model for livestock insurance pricing. With the exception of the aforementioned applications and improvements, credibility theory is rarely applied in agriculture ratemaking procedure. In order to improve the ratemaking process for the agricultural insurance and reinsurance sectors, this thesis proposes some augmented models based on the credibility approach to enhance the crop insurance ratemaking framework and improve crop yield forecasting model.

1.3 Systemic Weather Risks

Weather variability is the primary cause of loss in agriculture by either a single identifiable event such as hail, fire, flood, etc., or adverse events during certain extended period, such as continued rainfall, long droughts, etc. Systemic weather risk is cited as one of the main reasons for the failure of private crop insurance (Miranda and Glauber, 1997). Further, potential effects related to climate change may lead to an increase in the frequency and severity of spatially correlated weather events, which could lead to an increase in insurance and reinsurance costs. This could have a profound impact on crop output, prices, and ultimately the ability to feed the worlds growing population into the future. Adverse weather events, especially extreme weather events, lead to spatially correlated catastrophes and involve large geographical regions. This thesis will address these issues by developing a new crop yield forecasting model that incorporates a comprehensive weather index system.

1.4 High Dimensional Copula Approach

The agriculture sector is subject to a variety of risks, including severe extreme natural hazards, that are usually spatially correlated and affecting many people. Copula models have

been popular in quantifying systemic and spatially-dependent risks. However, standard models of systemic risk nearly always assume linear correlations, which have been inadequate in capturing nonlinear dependence across different geographic regions (Goodwin and Hungerford, 2014). In this thesis, a new Lévy subordinated Hierarchical Archimedean copula (LSHAC) model is proposed to model systemic weather risk. Empirical results show that the LSHAC model has better estimation performance compared to the classical Gaussian copula and the traditional hierarchical Archimedean copulas (HAC). Constructed from Lévy subordinators, the LSHAC model has more flexibility in modeling the tail dependence of the weather variables across different regions with fewer parameters compared to Gaussian copulas.

1.5 Multivariate Weighting Approach

The crop insurance sector has some unique features that are not commonly shared with most business lines of P & C insurance. First, agricultural losses tend to be highly spatially correlated and are at times encountered in extreme amounts. Second, over the years, the crop insurance program has experienced many program structure changes, leading to significant changes coverage levels and premium rates. Finally, agricultural insurance industry is a weather sensitive sector that is largely exposed to climate change effects and systemic weather risk. Therefore, some researchers have suggested to incorporate additional variables to weight the historical losses experiences in order to reflect the weather conditions and programs changes. For example, a study by RMA suggests a binning procedure integrating weather variables to weight historical loss data. In this thesis, a multivariate weighting distribution is proposed to integrate additional auxiliary variables into the ratemaking framework to enhance the crop reinsurance pricing.

1.6 Organization of Thesis

This thesis focuses on developing sound and improved actuarial and statistic tools in the context of an agricultural ratemaking framework. Chapter 2 addresses the pricing challenges in crop insurance and reinsurance, including shortness of data, and geographical correlated losses with high variations, and develops a scientific pricing framework that combines a new distribution family, Erlang mixture, and a modified credibility approach. A comprehensive data set representing the entire Canadian crop insurance sector is used as an empirical example, and the ten provincial crop insurance regions are used as a framework for the

credibility model. The new developed credibility-based Erlang model is shown to be superior and provides enhanced reinsurance pricing.

In Chapter 3, a new reinsurance pricing framework is proposed by developing a new crop yield forecasting model that integrates weather and crop production information from different geographically correlated regions. Furthermore, closed-form reinsurance pricing formulas are also derived. The model is empirically validated by analyzing the available weather data. Model selection algorithms, combining Cross-Validation (CV) and Principle Component Analysis (PCA), are proposed. The results show that the new forecasting model has improved in sample and out-of-sample forecasting capabilities. Based on these results, weather risk management suggestions are provided for agricultural reinsurance companies.

Chapter 4 discusses a new copula family known as the Lévy subordinated Hierarchical Archimedean copula (LSHAC). Motivated by the idea that new Archimedean copulas can be developed from Lévy subordinators, this chapter presents a general framework and notation system for the LSHAC model. This chapter also proposes a three-stage estimating procedure for the LSHAC model, with special attention on the estimation of the hierarchical structure of the copula functions. An empirical estimation example using daily temperatures from eight Canadian provinces demonstrates the advantage of the modeling capability of the proposed LSHAC model, relative to traditional elliptical copulas and the classical HAC models. Different weather risk hedging strategies for agricultural insurers are developed and compared. Empirical results support the importance of dependence structure assumptions.

Chapter 5 proposes a premium principle based on the multivariate weighted distribution to incorporate auxiliary variables to improve crop reinsurance ratemaking. The premium principle based on multivariate weighted distribution has some designable properties including linearity, stochastic ordering preserving, etc. It is also advantageous over univariate weighting distribution premium principle because it satisfies strict increasing relative risk loading if the weighting variables satisfy some stochastic ordering conditions. The empirical study compares the pricing results based on some popular premium principles and shows that by incorporating auxiliary information, the multivariate weighted distribution premium principle is able to assign higher loading to more risky contracts and achieve more sustainable long-run profits.

The contribution of this thesis lies on refining the actuarial pricing framework of agricultural insurance and reinsurance with high dimensional modeling solutions. By applying and extending credibility approach, high dimensional copula approach and multivariate weighting distribution approach, this thesis provides scientific methodologies to integrate a variety of auxiliary information, including the weather impacts, economic conditions and loss expe-

riences from neighbouring regions, to enhance crop insurance and reinsurance pricing and decision making.

Chapter 2

A Credibility-based Erlang Mixture Model for Pricing Crop Reinsurance

2.1 Introduction

A sound and scientific agricultural (re)insurance pricing approach is essential for maintaining sustainable and viable risk management solutions for farmers, governments, insurers, and reinsurers alike. In the ratemaking process, the goal is to determine the fair risk charge, often through the use of available historical observations. However, agricultural insurance can be quite difficult to price due to unique challenges, including shortness of data, and highly variable losses from year to year, which are often geographically correlated across regions (Porth et al., 2014). Pricing models that lack a scientific framework and possible consistency over time present a significant concern, including reinsurers who are currently faced with a competitive landscape, with lower premium rates for some business lines, and as such, tightening margins. Therefore, a robust pricing framework may help lead to steady underwriting performance and improved profitability over the long term.

The objective of this chapter is to address some of the fundamental issues surrounding crop insurance ratemaking, from the perspective of the reinsurer, through the development of a scientific pricing framework that can be consistently and widely adopted by the agricultural sector. In doing so, this chapter comprehensively examines the ratemaking process, including a review of the generating process of the historical loss cost ratio's (LCR's, which is defined as the ratio of indemnities to liabilities) in order to gain a better understanding

of the underlying yield distributions. Another parametric distribution known as the Erlang mixture is investigated to improve the goodness of fit. Further, the liability weighted LCR is introduced as a more conservative definition for the pricing model. Based on these results, two modified versions of the Bühlmann-Straub credibility model are developed to enhance the pricing framework by combining information from the observed data of certain risk categories (i.e. historical LCR's for the individual region/province), also taking into consideration the experience from the collective risk pool (i.e. the entire crop insurance program in Canada across the nine other geographic regions/provinces).

Trending Process

A fundamental issue to be considered before ratemaking is a solid understanding of the generating process of the historical data. In time series analysis, scholars generally prefer to work with stationary processes, with constant mean and finite variance (Hamilton, 1994). However, in crop insurance, the yield and other related data series, such as the LCR, do not satisfy this assumption of stationarity. Instead, the crop time series data usually contains trends defined as either deterministic or stochastic, and the corresponding testing and detrending methods remain controversial. Many statistical tests have been proposed by econometricians over the past several decades, including the Augmented Dickey-Fuller test (ADF test; Dickey, 1976; Dickey and Fuller, 1979), the Phillips-Perron test (PP test; Perron, 1986; Phillips and Perron, 1988), the Dickey-Fuller with Generalized Least Squares Detrending test (DF-GLS test; Elliott et al., 1996), and the Kwiatkowski-Phillips-Schmidt-Shin test (KPSS test; Kwiatkowski et al., 1992). Unit root tests are also developed for panel data (Andrew Levin, 2002; Maddala and Wu, 1999; Pesaran, 2007). (add more recent reference)

In empirical studies, researchers typically suggest applying linear trends (or higher order polynomial trends) to crop yield data with deterministic trends (Gallagher, 1986; Harri et al., 2009; Luttrell and Gilbert, 1976; Sherrick et al., 2004; Turvey and Zhao, 1999). Some other research studies have focused on the presence of stochastic trends (Goodwin and Ker, 1998; Moss and Shonkwiler, 1993). A major limitation of these studies is that the current tools available for testing trends in time series are based on an asymptotic assumption, i.e., the sample size is assumed to be very large,(explain asymptotic assumptions) which is not satisfied in the case of crop insurance due to shortness of data. In view of this limitation, the aforementioned four methods for testing trends in time series are reviewed in this study, particularly focusing on the implications of working with small sample sizes, as is the case in crop insurance. Through Monte Carlo Simulation experiments, we show that in the small sample cases the size distortion is quite large and the power performance is poor for all four

of the tests. This finding demonstrates the potential serious limitation in relying on these tests to obtain accurate and reliable information about trends when data samples are small. As such, a focus of this chapter is on establishing a scientific approach to pricing that is robust under different trending processes.

Distribution Frameworks

Deciding which distribution to use for modelling crop yield leads to another challenge. Past studies have tried to model yields with some known single parametric distributional attributes, including the normal distribution (Just and Weninger, 1999), log-normal distribution (Jung and Ramezani, 1999; Stocks, 2000; Tirupattur et al., 1996), beta distribution (Nelson and Preckel, 1989; Sherrick et al., 2004; Tirupattur et al., 1996), gamma distribution (Gallagher, 1986), and weibull distribution (Sherrick et al., 2004). Nonparametric estimation methods have also been applied (Goodwin and Ker, 1998; Ker and Goodwin, 2000). As consensus has yet to be reached regarding yield modelling, this area of research remains of central importance given the fundamental importance it serves in the ratemaking process (Sherrick et al., 2004).

One caution in distribution fitting is working with highly parameterized models, which can lead to overfitting, and hence poor forecasting and inadequate pricing. Mixture models studied by researchers such as Lanoue et al. (2010), Woodard and Sherrick (2011b) and Yang (2011) have indicated that, in comparison to the single distribution models, these models offer greater flexibility and have better out-of-sample performances. In this chapter we explore the applicability of another mixture model, known as the Erlang mixture, for agricultural risk modelling. We also investigate its desirable properties.

To test the proposed Erlang Mixture distribution, different models are compared in terms of likelihoods and BIC values using a comprehensive data set that represents the entire crop insurance sector of Canada. The data includes historical indemnities and liabilities (from which the LCR is calculated), over 1974-2009, across 276 crop types, and 10 geographic regions (provinces).

Modified Credibility Approach

Credibility theory has received little attention in the area of agricultural insurance pricing, with the exception of two papers, Josephson et al. (2000), and Pai et al. (2014). In this chapter, a modified Bühlmann-Straub Credibility Model is developed based on the Erlang Mixture distribution, in an effort to enhance the reinsurance pricing framework. The motivation of utilizing a credibility approach is to help address the challenge of shortness of data, and improve the statistical estimates of the expected losses. The credibility approach

combines information from the historical LCR's for the individual region/province, as well as the collective experience for the entire crop reinsurance program in Canada across nine other geographic regions/provinces. An empirical pricing example is presented to support the proposed pricing framework.

Main Contributions of This Chapter

This chapter for the first time investigates the difficulties in deciding trending process with small data sample in agricultural insurance and introduces the Erlang mixture model in the context of agricultural risk modeling. By extending the traditional Bühlmann-Straub credibility model (Bühlmann, 1967; Bühlmann and Gisler, 2005; Bühlmann and Straub, 1970), two extended versions of the Bühlmann-Straub credibility model are also presented (Section 2.5). This chapter also recommends the liability weighted LCR to reweigh historical losses and provides a more conservative reinsurance pricing framework (Proposition 2A.1).

The remainder of this chapter is as follows. In the next section, four statistical tests to examine the trending process of crop loss data are considered, and the corresponding simulation results are provided. In Section 2.3, a statistical description of the data set is reviewed, and the proposed Erlang Mixture distribution family is presented in Section 2.4. Based on these results, the modified Bühlman-Straub Credibility approach is developed in Section 2.5, and the liability weighted LCR is proposed as a more conservative way to aggregate historical losses for the pricing model. Section 2.6 provides some empirical evidence regarding the appropriateness of the credibility models for pricing crop insurance/reinsurance policies. Finally, the study is concluded in Section 2.7 with some empirical recommendations and future research directions.

2.2 Trend Testing

2.2.1 Deterministic Trends and Stochastic Trends

For a time series y_t and $t \geq 0$, a model with deterministic trend is defined as:

$$y_t = d_t + u_t \tag{2.1}$$

where d_t is some deterministic function of t, and u_t is some stationary process. This model can also be described as a trend-stationary process or integrated of order zero process ($I(0)$ process). On the other hand, if u_t contains an autoregressive unit root, a model with a

stochastic trend is defined as:

$$u_t = \sum_{i=0}^{p} \phi_i u_{t-i} + v_t \tag{2.2}$$

where v_t is some stationary process, and $\text{Max}(\phi_i, i = 1, 2, \ldots, p) = 1$. This model can also be defined as a difference-stationary process or integrated of order one process ($I(1)$ process).

Numerous tests have been proposed by econometricians for testing unit roots. They are known as Unit Root Tests, with the null hypothesis that the time series is $I(1)$ against the alternative that the time series is $I(0)$. Or conversely, if the null hypothesis is the time series is $I(0)$ against the alternative hypothesis that the series is $I(1)$, they are known as Stationary Tests. Under a Gaussian process assumption, Elliott et al. (1996) prove asymptotically that many tests manage to achieve power functions that are extremely close to the power envelope given by Neyman-Pearson Lemma, which is the upper bound for the power function of any tests based on the same likelihood. The asymptotic assumption is highly questionable for most real world systems. Monte Carlo simulation is used in many studies to examine the finite sample performance of these tests (DeJong et al., 1991; Phillips and Perron, 1988; Schwert, 1989). However, these studies are limited in the sense that they all consider sample sizes larger than 100, which is still relatively large compared to the particular situation in crop insurance where data of only several decades (with annual observations) are available at best. Hence, a better understanding of how these tests perform in sample sizes far smaller than the asymptotic cases, prior to executing these tests for the purpose of testing trends, becomes critically important.

2.2.2 Simulation Results for Tests

This subsection reviews the four most frequently used tests for unit root followed by a simulation study.

Unit Root Tests

ADF test: Dickey (1976); Dickey and Fuller (1979) developed the ADF test based on the $AR(p)$ model defined as:

$$y_t = \beta' D_t + \phi y_{t-1} + \sum_{j=1}^{p} \psi_j \Delta y_{t-j} + \varepsilon_t \tag{2.3}$$

with $H_0 : \phi = 1$ against $H_1 : |\phi| < 1$, and where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. D_t is the trend term of the model, in the following simulation analysis, a "Constant Model" refers to $D_t = 1$, and the "Trend Model" refers to $D_t = (1, t)'$.(define D_t)

PP test: Perron (1986); Phillips and Perron (1988) proposed a unit root test based on a nonparametric regression model defined as:

$$\Delta y_t = \beta' D_t + \phi y_{t-1} + u_t \tag{2.4}$$

where u_t is I(0), which allows for heterosksdasticity and correlation.

DF-GLS test: Elliott et al. (1996) proposed a family of tests whose asymptotic power functions are tangent to the Gaussian power envelop by considering the asymptotic approximation based on the local-to-unity alternative $c = T(\phi - 1)$. Among this family of tests, the DF-GLS test, which is a modified and efficient version of the ADF t-test, is shown to improve power given the same sample size, and when the trending process of the time series is unknown.

Stationary test

KPSS test: Kwiatkowski et al. (1992) proposed a test with null hypothesis that the time series is trend stationary, against an alternative that it has a unit root. The model is specified as

$$\begin{aligned} y_t &= \beta' D_t + \mu_t + u_t, && (2.5)\\ \mu_t &= \mu_{t-1} + \varepsilon_t, && (2.6)\\ \varepsilon_t &\sim WN(0, \sigma_{\varepsilon_t}^2). && (2.7) \end{aligned}$$

with $H_0 : \sigma_\varepsilon^2 = 0$ against $H_1 : \sigma_\varepsilon^2 > 0$, and u_t is a stationary process, which is allowed to be heteroskedastic.

To study the size and power performance of these four tests, we conduct simulation experiments based on the following model with 1000 replicates:

$$y_t = \phi y_{t-1} + u_t; \qquad u_t \sim N(0, \sigma^2) \tag{2.8}$$

The size distortions and power performances of these tests are listed in Table 2.1 to Table 2.4, where the Constant Model refers to $D_t = 1$, and the Trend Model refers to $D_t = (1, t)'$. Each row corresponds to various sample sizes (e.g., 25, 50, up to 1000), and the columns are the number of lags chosen in the testing model. The results are compared to a nominal size of

0.05. These tables show that when the sample size is larger than 500, all four tests perform well as indicated by a small size distortion and a high power close to one, even if the D_t is not correctly specified in the test. However, when the sample size is smaller than 50, the test results have larger size distortions and unsatisfying power performances, namely, they provide inaccurate information regarding the trends of the process. Moreover, the size distortions and power performances are much more sensitive to the misspecification of the trend when the sample sizes are smaller than 250, which could result in misleading testing results. For example, in Table 2.1, when the sample size is 50, the average power of the ADF test is 158.24% higher when the trend model is correctly specified. Additionally, for the KPSS stationary test, Monte Carlo simulation results are found to be consistent with the findings reported by Kwiatkowski et al. (1992) for sample sizes equal to 30. Therefore, we conclude that the current statistical tests for trending processes are not revealing when sample sizes are prohibitively small, as is the challenge faced in crop insurance. In the following analysis, we will use the DF-GLS test and KPSS test for agricultural data since they are providing the best testing results.

2.3 Data Description and Properties

By using a unique and a comprehensive data set that covers the entire crop insurance sector of Canada (provided by Agriculture and Agri-Food Canada (AAFC)), this section provides an in-depth analysis on the statistical characteristics of the historical LCR's (Subsection 2.3.1) and the trend testing results (Subsection 2.3.2). The data set includes actual indemnities and liabilities, from 1974 through 2009, across 276 crop types, and 10 geographic regions (i.e. provinces) in Canada. The ten provinces considered include Alberta (AB), Manitoba (MB), Ontario (ON), British Columbia (BC), New Brunswick (NB), Nova Scotia (NS), Prince Edward Island (PEI), Québec (QC), Saskatchewan (SK), and Newfoundland and Labrador (NFLD).

In this chapter, the indemnities and liabilities are aggregated to a provincial level. From this, the loss cost ratio's (LCR's), are calculated as the ratio of indemnities to liabilities. Reinsurers commonly utilize the LCR to normalize the loss exposure and examine the underlying risk profile and compute premiums. This normalization is important because historically there have been significant increases in liabilities and yields, increasing program participation, as well as improvements in biotechnology, farming practices, etc. (Coble et al., 2008; Harri et al., 2009; Miranda and Glauber, 1997; Sherrick et al., 2004; Woodard et al., 2012; Woodard and Sherrick, 2011a). Given that reinsurers often face constraints regarding the

Size Distortion(ADF): Constant Mode					
Sample/Lags	2	4	6	8	10
T=25	0.0568	0.0715	0.0985	0.1497	0.2642
T=50	0.0519	0.0533	0.0556	0.0629	0.0761
T=100	0.0505	0.0513	0.0515	0.0512	0.0547
T=250	0.0482	0.0461	0.0500	0.0495	0.0458
T=500	0.0487	0.0529	0.0458	0.0542	0.0508
T=1000	0.0439	0.0493	0.0454	0.0426	0.0485
Size Distortion(ADF): Trend Model					
T=25	0.0806	0.1149	0.1704	0.3036	0.5553
T=50	0.0625	0.0685	0.0802	0.0910	0.1148
T=100	0.0554	0.0580	0.0609	0.0653	0.0640
T=250	0.0472	0.0553	0.0544	0.0506	0.054
T=500	0.0493	0.0524	0.0509	0.0523	0.0484
T=1000	0.0490	0.0528	0.0531	0.0542	0.0492
Power Performance(ADF): Constant Model					
Sample/Lags	2	4	6	8	10
T=25	0.2164	0.2484	0.3118	0.4319	0.6041
T=50	0.2583	0.2475	0.2585	0.2674	0.2865
T=100	0.4440	0.386	0.3594	0.3321	0.3128
T=250	0.9518	0.9095	0.8684	0.8195	0.7663
T=500	1	0.9999	0.9998	0.9992	0.9978
T=1000	1	1	1	1	1
Power Performance(ADF): Trend Model					
T=25	0.0836	0.1138	0.1793	0.3096	0.3088
T=50	0.0908	0.0916	0.0882	0.1024	0.1069
T=100	0.1674	0.1478	0.1358	0.1274	0.1255
T=250	0.7499	0.6568	0.5750	0.4941	0.4911
T=500	0.9998	0.9971	0.9930	0.9805	0.9813
T=1000	1	1	1	1	1

Table 2.1: Size distortion and power performance of ADF test. The norminal size is 0.05, and the Alternative Hypothesis is $\phi = 0.9$. Lags refers to the order of integration in the ADF regression model (value of p in equation (2.3)).

Size Distortion(PP): Constant Mode				
	Z_α		Z_τ	
Sample/Lags	l_4	l_{12}	l_4	l_{12}
T=25	0.0474	0.0447	0.0608	0.0587
T=50	0.0574	0.0533	0.0593	0.0588
T=100	0.0510	0.0568	0.0526	0.0589
T=250	0.0513	0.0576	0.0535	0.0528
T=500	0.0529	0.0562	0.0517	0.0577
T=1000	0.0508	0.0507	0.052	0.0517
Size Distortion(PP): Trend Model				
T=25	0.0365	0.0356	0.0607	0.0578
T=50	0.0587	0.0540	0.0644	0.0626
T=100	0.0659	0.0618	0.0640	0.0619
T=250	0.0618	0.0627	0.0625	0.0585
T=500	0.0539	0.0584	0.0570	0.0589
T=1000	0.056	0.0534	0.0576	0.0545
Power Performance(PP): Constant Model				
	Z_α		Z_τ	
Sample/Lags	l_4	l_{12}	l_4	l_{12}
T=25	0.0883	0.0945	0.0839	0.0866
T=50	0.1925	0.1892	0.1479	0.1441
T=100	0.4759	0.4845	0.3591	0.3692
T=250	0.9924	0.9912	0.9711	0.9701
T=500	1.0000	1.0000	1.0000	1.0000
T=1000	1.0000	1.0000	1.0000	1.0000
Power Performance(PP): Trend Model				
T=25	0.0467	0.0479	0.0722	0.0713
T=50	0.1003	0.0959	0.0999	0.0961
T=100	0.2769	0.2693	0.2350	0.2329
T=250	0.9098	0.9053	0.8644	0.8588
T=500	1.0000	1.0000	1.0000	1.0000
T=1000	1.0000	1.0000	1.0000	1.0000

Table 2.2: Size distortion and power performance of PP test. The norminal size is 0.05, and the Alternative Hypothesis is $\phi = 0.9$. The test statistics Z_α and Z_τ are calculated according to the expressions in the original paper. $l_4 = \text{floor}[4(T/100)^{1/4}]$ and $l_{12} = \text{floor}[12(T/100)^{1/4}]$.

Size Distortion(DF-GLS): Constant Mode					
Sample/Lags	2	4	6	8	10
T=25	0.1276	0.0983	0.0881	0.0841	0.1143
T=50	0.0914	0.0872	0.0782	0.0693	0.0633
T=100	0.0670	0.0705	0.0633	0.0596	0.0597
T=250	0.0612	0.0600	0.0583	0.0554	0.0560
T=500	0.0539	0.0554	0.0545	0.0485	0.0557
T=1000	0.0535	0.0534	0.0540	0.0489	0.0515
Size Distortion(DF-GLS): Trend Model					
T=25	0.0594	0.0353	0.0243	0.0232	0.0523
T=50	0.0395	0.0324	0.0252	0.0202	0.0170
T=100	0.0479	0.0415	0.0376	0.0450	0.0420
T=250	0.0508	0.0494	0.0496	0.0503	0.0445
T=500	0.0491	0.0486	0.0490	0.0452	0.0438
T=1000	0.0410	0.0460	0.0475	0.0446	0.0511
Power Performance(DF-GLS): Constant Model					
Sample/Lags	2	4	6	8	10
T=25	0.2180	0.1613	0.1156	0.1057	0.1257
T=50	0.3010	0.2352	0.1881	0.1516	0.1216
T=100	0.5326	0.4442	0.3805	0.314	0.2717
T=250	0.8852	0.8427	0.7800	0.7301	0.6768
T=500	0.9806	0.9642	0.9451	0.9132	0.8957
T=1000	0.9991	0.9975	0.9934	0.9879	0.9790
Power Performance(DF-GLS): Trend Model					
T=25	0.0676	0.0384	0.026	0.0224	0.0492
T=50	0.0724	0.0511	0.0355	0.0244	0.0174
T=100	0.1907	0.1571	0.1267	0.1378	0.1041
T=250	0.8207	0.7345	0.6478	0.5604	0.4837
T=500	0.9910	0.9845	0.9677	0.9479	0.9197
T=1000	1.0000	0.9999	0.9995	0.9990	0.9970

Table 2.3: Size distortion and power performance of DF-GLS test. The norminal size is 0.05, and the Alternative Hypothesis is $\phi = 0.9$. Lags refers to the order of integration in the ADF regression model (value of p in equation (2.3)).

Size Distortion(KPSS): Constant Mode			
Sample/Lags	l_0	l_4	l_1
T=25	0.0518	0.0311	0.0369
T=50	0.0497	0.0402	0.0412
T=100	0.0502	0.0441	0.0457
T=250	0.0494	0.0501	0.0503
T=500	0.0482	0.0460	0.0495
T=1000	0.0511	0.0461	0.0473
Size Distortion(KPSS): Trend Model			
T=25	0.0574	0.0399	0.0387
T=50	0.0535	0.0431	0.0406
T=100	0.0531	0.0453	0.0464
T=250	0.0571	0.0467	0.0500
T=500	0.0522	0.049	0.0505
T=1000	0.0514	0.0515	0.0506
Power Performance(KPSS): Constant Model			
Sample/Lags	l_0	l_4	l_{12}
T=25	0.1191	0.0877	0.0842
T=50	0.3080	0.2560	0.2576
T=100	0.6048	0.5281	0.5250
T=250	0.9117	0.8498	0.8403
T=500	0.9888	0.9657	0.9630
T=1000	0.9998	0.9944	0.9949
Power Performance(KPSS): Trend Model			
T=25	0.0742	0.0512	0.0498
T=50	0.1406	0.0993	0.1016
T=100	0.3859	0.2988	0.2897
T=250	0.8451	0.7558	0.7568
T=500	0.9872	0.9628	0.9634
T=1000	0.9999	0.9968	0.9965

Table 2.4: Size distortion and power performance of KPSS test. The norminal size is 0.05, and the Alternative Hypothesis is $\phi = 0.9$. $l_4 = \text{floor}[4(T/100)^{1/4}]$ and $l_{12} = \text{floor}[12(T/100)^{1/4}]$.

detail of the crop data provided to them, loss modelling becomes difficult. Therefore, a pricing model that is robust under different trending processes is imperative.

2.3.1 Statistical Characteristics of the LCR Data

This subsection describes the statistical characteristics of the Canadian crop LCR data set used in this chapter. For all ten provinces, the data is positively skewed, and most provinces have high kurtosis, reflecting a heavy tail property. Empirical Value at Risk (VaR's) and Conditional Tail Expectations (CTE's) at different levels also report heavy tails. In crop insurance, a LCR greater than 0.25 is usually suggestive of a substantial and widespread loss. As shown in Table 2.5, the tail quantiles for most provinces are higher than 0.25, which suggests the catastrophic nature of the loss experience. This is particularly true for the three largest crop insurance provinces, Alberta, Manitoba and Saskatchewan, which together comprise more than 50 % of the crop insurance program in Canada. These statistical characteristics are important to consider in developing a reinsurance pricing model, where it is important to ensure that peak loss experiences are accounted for in loading and reserving considerations.

LCR	AB	MB	ON	BC	NB	NS	PEI	QC	SK	NFLD
μ	0.13	0.09	0.07	0.08	0.14	0.05	0.08	0.06	0.11	0.14
σ	0.09	0.07	0.05	0.05	0.11	0.03	0.05	0.04	0.09	0.09
γ_1	2.31	1.77	1.39	0.58	1.79	0.65	1.18	1.25	1.97	0.66
γ_2	6.33	2.16	1.02	-2.97	4.39	0.11	1.96	1.84	3.93	-0.24
$VaR_{0.85}$	0.20	0.20	0.11	0.14	0.21	0.08	0.14	0.10	0.19	0.25
$VaR_{0.9}$	0.23	0.23	0.16	0.16	0.30	0.09	0.15	0.12	0.22	0.28
$VaR_{0.95}$	0.33	0.29	0.19	0.18	0.37	0.12	0.18	0.15	0.40	0.34
$CTE_{0.85}$	0.29	0.26	0.17	0.17	0.34	0.11	0.17	0.13	0.29	0.31
$CTE_{0.9}$	0.35	0.28	0.19	0.18	0.40	0.12	0.19	0.15	0.36	0.34
$CTE_{0.95}$	0.48	0.32	0.19	0.18	0.57	0.13	0.25	0.19	0.43	0.35

Table 2.5: Statistical Description of the LCR Data.μ-Mean, σ-Standard Deviation, γ_1-Skewness, γ_2-Excess Kurtosis.

2.3.2 Trend Testing Results

As discussed in the previous sections, the first step of loss modeling is to understand the data trending process, and then eliminate the trends accordingly. Various detrending methods can result in substantial differences in the underlying data series, and at times lead to misspecified

estimators of the distributions (Zapata and Rambaldi, 1989). Therefore, different trends require correspondingly appropriate detrending methods. When the trend is deterministic, the data should be detrended by eliminating the time trend, yet, when the trend is stochastic, first differences should be used.

In this chapter, we begin the pricing process using the DF-GLS unit root test. The DF-GLS test is shown to have the best performance for small sample cases among the three unit root tests according to Table 2.1 to Table 2.3, followed by the KPSS stationary test. Table 2.6 shows the test results of the LCR data for each province according to the DF-GLS test and KPSS test. Five of the ten provinces are found to have inconsistent results for both tests. This makes it difficult to draw conclusions from the tests in order to determine the most suitable detrending method. As such, a model that is robust under various detrending methods can be very helpful to ensure a sound pricing framework, and avoid misclassification of the trending process.

LCR	AB	MB	ON	BC	NB	NS	PEI	QC	SK	NFLD
time trend	1	1	1	1	1	1	1	1	1	1
DF-GLS	I(1)	I(0)	I(1)	I(0)	I(0)	I(1)	I(0)	I(1)	I(0)	I(1)
KPSS	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)

Table 2.6: Test Results of the LCR Data. (Constant Model, Significant level=0.05). 1 in the first line refers to an insignificant time trend, I(0) refers to a deterministic trend, and I(1) is a stochastic trending process.

2.4 Erlang Mixture Distribution

In this section, we propose using an Erlang mixture distribution for modelling agricultural crop data. We will also provide evidence that this distribution is able to capture the tails of the data more accurately. To the best of our knowledge this distribution has not been studied in agricultural risk modelling and (re)insurance pricing. The Erlang Mixture family is a very important class of distribution because theoretically it is dense in the space of positive distributions. In other words, there always exists a series of mixture of Erlangs that converges in distribution to an arbitrary positive distribution (Tijms, 1994).

The probability distribution function (p.d.f.) of the mixed Erlang model is defined as

$$f(x|\theta,\alpha) = \sum_{i=1}^{M} \alpha_i \frac{x^{r_i-1}e^{-x/\theta}}{\theta^{r_i}(r_i-1)!} \quad x > 0,$$

where the scale parameter θ is assumed to be the same for each mixing component, and the shape parameters r_i's, $(i = 1, \ldots, M)$ are increasing integers, which are assumed to be known in each round of estimation. The number of mixing Erlang distributions M, the mixing coefficients α_i's, $(i = 1, \ldots, M)$, and θ, are found by the Expectation-Maximization (EM) Algorithm proposed by Lee and Lin (2010) to maximize the log-likelihood functions.

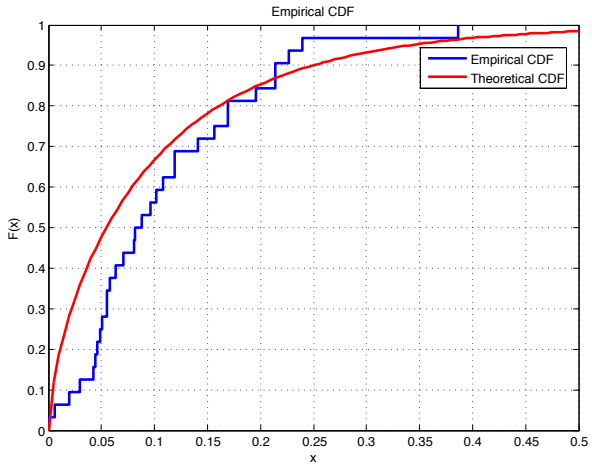
To test the plausibility of this distribution for modelling crop data, we use the Bayesian information criterion (BIC). The BIC was first introduced by Schwarz in 1978, and is defined as

$$\text{BIC} = -2\log(\hat{\mathcal{L}}) + K \cdot \log(N),$$

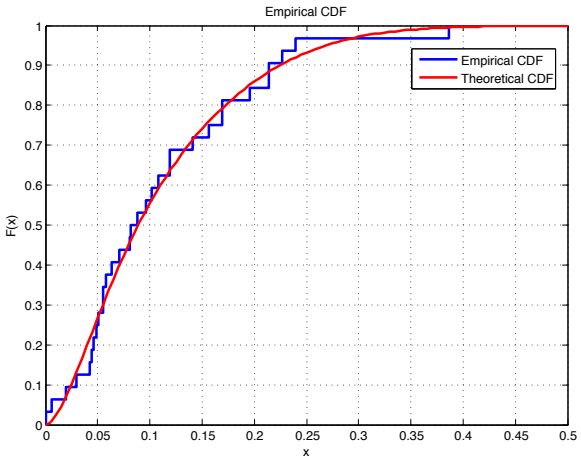
where $\hat{\mathcal{L}}$ is estimated maximum value of the likelihood function, K is the number of parameters in the model, and N is the sample size. The advantage of using the BIC as our model selection criterion is that it is valid not only for nested models but also for non-nested models (Burnham and Anderson, 2002).

In order to compare the goodness of fit of the Erlang Mixture distribution model, some commonly used distributions, such as Gamma, Weibull, Beta, Normal and Loglogistic are also considered as benchmarks. Further, the Erlang Mixture model is considered with respect to the two different detrending processes considered previously. The first method involves regressing the LCR's with respect to time trend and hereafter is referred to as *regressed data*, and the second method involves differencing the LCR's and hereafter is referred to as *differenced data*. To be more specific, the regressed data are detrend by reducing the first order polynomial function of time from the original LCR data, and the difference data are created by first differencing. The fitting results for the regressed data as well as differenced data are found in Table 2.7, respectively. For both detrending methods, the Erlang Mixture distributions have the highest likelihoods and the lowest BICs for all the ten provinces. While the Erlang Mixture distributions are superior, the Gamma and Weibull distributions also perform well. The empirical cumulative distribution functions (eCDFs) of Erlang distribution and Gamma distribution for the province SK are presented in Figure 2.1, and the QQ-plot for Erlang distribution is displayed in Figure 2.2 (add QQ-plots etc to the fitted distributions).

As mentioned earlier, reinsurers are faced with the challenge of asymmetric information.



(a) eCDF for Gamma distribution (SK).



(b) eCDF for Erlang mixture distribution (SK).

Figure 2.1: eCDFs for LCR from SK for Gamma distribution and Erlang mixture distribution.

Reinsurers often only have access to limited historical time series and often this data is highly aggregated with mixed coverage levels and rates through time, which may mask potential trends. From this perspective, within the framework of a parametric approach, the Erlang distribution has a promising advantage in that it is robust regardless of the detrending methods, which is not the case with the other distributions. Therefore, the Erlang mixture distribution may help to improve the ratemaking and loss reserving process for insurers and reinsurers, particularly when faced with small data samples as in crop insurance.

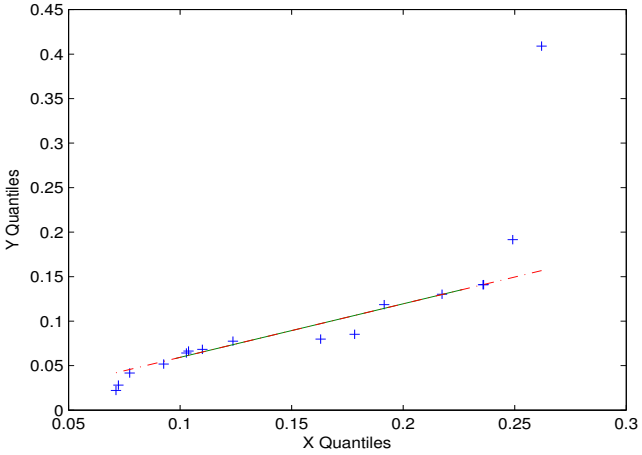


Figure 2.2: QQ-plot for LCR from SK fitted with Erlang mixture distribution.

Maximum Likelihood Estimations: Regressed Data							
Province	Criteria	Erlang	Gamma	Weibull	Beta	Loglogistic	Normal
AB	Log-Likelihood	135.57	50.16	49.76	3.06	43.95	37.92
	BIC	-242.48	-93.16	-92.36	1.05	-80.74	-68.67
MB	Log-Likelihood	200.84	57.61	57.12	4.26	52.17	42.85
	BIC	-383.77	-108.06	-107.07	-1.35	-97.16	-78.53
ON	Log-Likelihood	216.22	76.93	75.99	10.62	70.57	59.01
	BIC	-414.53	-146.70	-144.80	-14.08	-133.97	-110.85
BC	Log-Likelihood	334.73	50.88	52.74	-4.11	43.68	59.95
	BIC	-651.53	-94.59	-98.32	15.38	-80.18	-112.74
NB	Log-Likelihood	205.84	42.58	41.36	3.00	34.58	29.19
	BIC	-397.34	-77.99	-75.56	1.16	-62.00	-51.21
NS	Log-Likelihood	276.44	75.68	75.72	1.69	67.53	75.76
	BIC	-524.22	-144.18	-144.28	3.78	-127.90	-144.36
PEI	Log-Likelihood	175.53	60.69	60.75	1.52	52.81	58.94
	BIC	-333.14	-114.22	-114.34	4.12	-98.45	-110.71
QC	Log-Likelihood	255.88	70.68	71.08	1.85	63.71	69.00
	BIC	-493.84	-134.19	-134.99	3.48	-120.25	-130.84
SK	Log-Likelihood	151.23	50.47	49.73	4.19	43.95	35.20
	BIC	-284.54	-93.77	-92.29	-1.22	-80.74	-63.23
NFLD	Log-Likelihood	265.15	35.67	34.53	-2.60	25.24	35.21
	BIC	-512.38	-64.18	-61.90	12.37	-43.31	-63.25
Maximum Likelihood Estimations: Differenced Data							
AB	Log-Likelihood	428.14	-2.65	2.30	-15.99	-9.75	29.96
	BIC	-842.07	12.42	2.50	39.08	26.60	-52.81
MB	Log-Likelihood	251.84	14.07	17.10	-12.15	6.35	33.47
	BIC	-485.90	-21.04	-27.09	31.41	-5.60	-59.84
ON	Log-Likelihood	318.79	35.95	36.48	-2.71	26.24	46.59
	BIC	-626.90	-64.80	-65.84	12.53	-45.37	-86.07
BC	Log-Likelihood	332.68	44.39	45.38	-4.06	34.88	57.41
	BIC	-651.13	-81.66	-83.64	15.23	-62.66	-107.71
NB	Log-Likelihood	384.53	-10.61	-5.81	-13.55	-17.92	22.17
	BIC	-754.83	28.33	18.74	34.22	42.96	-37.23
NS	Log-Likelihood	347.04	40.07	43.08	-3.18	32.65	57.95
	BIC	-679.86	-73.02	-79.05	13.47	-58.19	-108.80
PEI	Log-Likelihood	372.08	17.06	21.23	-10.42	10.02	41.98
	BIC	-729.93	-27.01	-35.35	27.96	-12.94	-76.85
QC	Log-Likelihood	398.93	32.67	35.54	-7.79	25.10	50.08
	BIC	-776.53	-58.23	-63.97	22.69	-43.08	-93.05
SK	Log-Likelihood	349.67	7.06	10.24	-15.26	-0.41	26.59
	BIC	-685.11	-7.01	-13.36	37.63	7.92	-46.07
NFLD	Log-Likelihood	353.31	13.05	15.30	-12.66	5.03	28.42
	BIC	-685.28	-18.99	-23.49	32.43	-2.95	-49.72

Table 2.7: Maximum Likelihood Estimating Results for LCR Data Under Two different Detrending Methods.

2.5 Modified Bühlmann-Straub Credibility Models

The objective of this section is to present an enhanced pricing methodology for crop insurance through the integration of a modified credibility model, and the Erlang-mixture model. The credibility model is a widely accepted ratemaking tool used by actuaries in the property and casualty sectors to price insurance policies for risk exposures such as automobile and health. More recently, credibility model has also been used for modeling operational risk (See Bühlmann et al., 2007).

The motivation for using a credibility approach is to enhance pricing through the addition of supplemental information. For example, suppose we are interested in predicting future claims of a particular risk class, the traditional pure risk premium approach involves using past claim experience data pertinent to the risk class. However, in addition to using the history of claims for the given risk class, credibility theory (see Bühlmann (1967); Bühlmann and Straub (1970)) argues that the prediction power can be improved by also exploiting the claim experience of the collective pool of risk classes with similar characteristics. By restricting the class of estimators to be a linear combination of past observations, an optimal credibility-based prediction estimator that minimizes the squared difference between the predicted values and the claim experience can be obtained. This results in an estimator that is simple and intuitive. The credibility estimator becomes a linear combination between the individual risk class claim experience and the collective claim experience with relative weights dependent on the credibility of the claim experience of the risk class to the collective risk pool. In the special case in which the experience of a risk class is fully credible, then the experience of the risk class is solely used for future loss prediction.

The classical Bühlmann-Straub credibility model assumes that claims are independent conditional on a given risk category, and the random variables characterizing the risk profiles are identically and independently distributed (see Bühlmann and Gisler, 2005). However, in the application of agriculture insurance, the identical conditional distribution assumptions may not be appropriate because there are some priori differentiations for each province. This priori differentiation intuitively makes sense since each province is faced with different weather, technical, and economic situations, therefore, their risk characteristics should contain some priori differences. In addition to giving a priori differentiation, we obtain the credibility premium based on two different risk assumptions for our loss data, including a homogeneous risk assumption and a heterogeneous risk assumption. The two modified Bühlmann-Straub credibility models are discussed next.

For $i = 1, 2, \ldots, d$ and $j = 1, 2, \ldots, T$, where d and T denote, respectively, the number of

risk categories (i.e. number of provinces in our context), and the number of years of data, we define the following notation:

- X_{ij}: loss cost ratio (LCR) of risk i in year j,
- I_{ij}: indemnity of risk i in year j,
- L_{ij}: liability (risk exposure) of risk i in year j.
- Θ_i: the parameter of the i-th risk.

In our context, the LCR $X_{ij} = \frac{I_{ij}}{L_{ij}}$ is the basic random variable for the credibility rating. The parameter Θ_i is defined as a random variable that describes the risk characteristics for the i-th risk. The assumptions underlying the homogeneous and heterogeneous risk models, and the resulting credibility premiums are stated below.

M1: **Homogeneous Risk Model:**

A1: $\mathbf{X} = \{X_{ij} : i = 1, 2, \ldots, d; j = 1, 2, \ldots, T\}$ are independent conditional on Θ_i, with its (conditional) mean and (conditional) variance given, respectively, by

$$E[X_{ij}|\Theta_i] = a_i \mu(\Theta_i) \tag{2.9}$$

$$Var[X_{ij}|\Theta_i] = a_i \frac{\sigma_1^2(\Theta_i)}{b_i}. \tag{2.10}$$

We use the parameters a_i and b_i to describe the priori differentiation, due to different weather, technical, and economic situations, between each geographical regions. $\mu(\Theta_i)$ and $\sigma_1^2(\Theta_i)$ are functions of Θ_i.

A2: $(\Theta_1, \mathbf{X_1}), (\Theta_2, \mathbf{X_2}) \ldots, (\Theta_d, \mathbf{X_d})$ are independent, and $\Theta_1, \Theta_2, \ldots, \Theta_d$ are independent and identically distributed.

As shown in Appendix 2A.2, for the two assumptions stated above, the credibility premium for the i-th risk class (province) is given by

$$\text{CreP}_i^{\text{Hom}} = Z_i^{\text{Hom}} \bar{X}_i^{\text{Hom}} + (1 - Z_i^{\text{Hom}}).\mu_i \tag{2.11}$$

In the above formula,

- $\bar{X}_i^{\text{Hom}}$, which denotes the historical average of the LCR of risk category i, is given by

$$\bar{X}_i^{\text{Hom}} = \frac{1}{T} \sum_{j=1}^{T} X_{ij}. \tag{2.12}$$

- $\mu_0 = E\big(\mu(\Theta_i)\big)$ can be interpreted as the risk premium for the entire collective risk pool (without taking into consideration the priori difference between the provinces).
- Z_i^{Hom}, which is the credibility factor, captures the weight that is assigned to the historical data of the i-th province in the credibility premium. Formally this is defined as

$$Z_i^{\text{Hom}} = \frac{a_i b_i \cdot T}{a_i b_i \cdot T + \kappa^{\text{Hom}}} \tag{2.13}$$

with the credibility coefficient κ^{Hom} given by

$$\kappa^{\text{Hom}} = \frac{v^{\text{Hom}}}{a^{\text{Hom}}}, \qquad \text{where } v^{\text{Hom}} = E[\sigma_1^2(\Theta_i)], \quad a^{\text{Hom}} = Var[\mu(\Theta_i)]. \tag{2.14}$$

A more intuitive expression for the credibility factor is to use the following equivalent representation (Bühlmann and Gisler, 2005):

$$Z_i^{\text{Hom}} = \frac{T}{T + \tilde{\kappa}^{\text{Hom}}} \qquad \text{where } \tilde{\kappa}^{\text{Hom}} = \frac{E[Var[X_{ij}|\Theta_i]]}{Var(E[X_{ij}|\Theta_i])}.$$

This demonstrates that the credibility factor depends explicitly on three factors. More specifically, the credibility factor increases as

  - the number of observations T increases,
  - the variability within the risk classes (as measured by $E[Var[X_{ij}|\Theta_i]]$) decreases,
  - the heterogeneity of the collective risk pool (as measured by $Var(E[X_{ij}|\Theta_i])$) increases.

The above interpretations are intuitive. The credibility premium (2.11) is a weighted average between the observed LCR average of the i-th risk category and the risk premium for the entire collective risk pool. The past observations become more credible and lead to higher credibility factor with more observations, or the smaller within risk category variability, or with larger between risk category variability.

M2: **Heterogeneous Risk Model:**

A1: $\mathbf{X} = \{X_{ij} : i = 1, 2, \ldots, d; j = 1, 2, \ldots, T\}$ are independent conditional on Θ_i, , with its (conditional) mean and (conditional) variance given, respectively, by

$$E[X_{ij}|\Theta_i] = a_i \mu(\Theta_i) \tag{2.15}$$

$$Var[X_{ij}|\Theta_i] = a_i \frac{\sigma_2^2(\Theta_i)}{b_i \cdot L_{ij}} \tag{2.16}$$

where a_i and b_i are interpreted similarly as in the previous model and $\sigma_2^2(\Theta_i)$ is another function in term of Θ_i.

A2: $(\Theta_1, \mathbf{X_1}), (\Theta_2, \mathbf{X_2}) \ldots, (\Theta_d, \mathbf{X_d})$ are independent, and $\Theta_1, \Theta_2, \ldots, \Theta_d$ are independent and identically distributed.

To capture the heterogeneity of the data, the conditional variance of X_{ij} depends not only on parameters a_i, b_i, and a function in term of Θ_i as in the homogeneous risk model, but, also on an extra factor L_{ij}, which measures the year j risk exposure of the i-th province. Under the above assumptions, Appendix 2A.2 similarly establishes that the credibility premium for province i becomes

$$\text{CreP}_i^{\text{Het}} = Z_i^{\text{Het}} \bar{X}_i^{\text{Het}} + (1 - Z_i^{\text{Het}})\mu_i \tag{2.17}$$

where

$$\bar{X}_i^{\text{Het}} = \sum_{j=1}^{T} \frac{L_{ij}}{L_{i\cdot}} X_{ij} \tag{2.18}$$

$$Z_i^{\text{Het}} = \frac{a_i b_i L_{i\cdot}}{a_i b_i L_{i\cdot} + \kappa^{\text{Het}}} \tag{2.19}$$

$$L_{i\cdot} = \sum_{j=1}^{T} L_{ij}, \quad \mu_i = a_i \mu_0, \tag{2.20}$$

$$\kappa^{\text{Het}} = \frac{v^{\text{Het}}}{a^{\text{Het}}}, \quad v^{\text{Het}} = E[\sigma_2^2(\Theta_i)], \quad a^{\text{Het}} = Var[\mu(\Theta_i)]. \tag{2.21}$$

The interpretations of these parameters are similar to the previous model. The main difference between the homogeneous and heterogeneous assumptions is that consideration is given to the improvement of the risk exposure in each year in the heterogeneous model. This is important because trends in risk exposure, as a result of increasing yields and increasing commodity prices, leads to considerable variability. This makes the heterogeneous credibility model more reasonable by factoring both geographic and time variations of liability into the premium calculation.

Another interesting result arising from these models is in the use of the historical data in the credibility estimator. In the former model, $\bar{X}_i^{\text{Hom}}$ in (2.12) is simply the arithmetic average of the LCR's over the past T years, while in the latter model, $\bar{X}_i^{\text{Het}}$ in (2.18) is defined as the liability weighted average of historical LCR's. This gives rise to two different ways of averaging past LCR's, depending on the model assumptions.

This difference could have some important implications for ratemaking. In particular, as formally established in Appendix 2A.1 (see Proposition 2A.1), under some additional tech-

nical assumptions, we have $\bar{X}_i^{\text{Het}} \geq \bar{X}_i^{\text{Hom}}$; i.e., the liability weighted LCR's is at least as large as the simple averaging LCR's. More importantly, this also implies that ratemaking based on the liability weighted LCR's should produce a risk premium that is higher than the corresponding rate based on the simple averaging LCR's.

Additionally, we assume that the following relationship holds

$$\sigma_1^2(\Theta_i) = \frac{1}{\bar{L}_i}\sigma_2^2(\Theta_i) \tag{2.22}$$

where $\bar{L}_i = \frac{1}{T}\sum_{j=1}^{T} L_{ij}$. This relation is reasonable since $\sigma_1^2(\Theta_i)$ and $\sigma_2^2(\Theta_i)$ can be interpreted as the conditional variance of LCR (X_{ij}) and the indemnity (I_{ij}), respectively (recall (2.10) and (2.16)). Then we have

$$\text{CreP}_i^{\text{Hom}} = \frac{a_i b_i}{a_i b_i + \frac{v^{\text{Hom}}}{T a^{\text{Hom}}}} \bar{X}_i^{\text{Hom}} + \frac{a_i}{\frac{a_i b_i}{\frac{v^{\text{Hom}}}{T a^{\text{Hom}}}} + 1} \mu_0 \tag{2.23}$$

$$\text{CreP}_i^{\text{Het}} = \frac{a_i b_i}{a_i b_i + \frac{v^{\text{Hom}}}{T a^{\text{Hom}}}} \bar{X}_i^{\text{Het}} + \frac{a_i}{\frac{a_i b_i}{\frac{v^{\text{Hom}}}{T a^{\text{Hom}}}} + 1} \mu_0 \tag{2.24}$$

where the $\bar{X}_i^{\text{Hom}}$ and $\bar{X}_i^{\text{Het}}$ are defined in (2.12) and (2.18), respectively. The above results suggest that if $\bar{X}_i^{\text{Het}} \geq \bar{X}_i^{\text{Hom}}$ then $\text{CreP}_i^{\text{Het}} \geq \text{CreP}_i^{\text{Hom}}$. Thus the credibility premium based on the liability weighted LCR's is more conservative than the corresponding premium based on the simple averaging LCR's.

2.6 Applications of Credibility Models

In Sections 2.3 and 2.4, an extensive analysis was conducted on a unique set of crop data, covering 10 provinces in Canada over the years 1974 to 2009. Among all the plausible distributions that were investigated, the Erlang-mixture model was found to provide the best goodness of fit to the historical LCRs. Section 2.5 then described two modified versions of the Bühlmann-Straub credibility model, which may provide a better ratemaking framework for crop (re)insurance policies. Continuing to explore the same set of data, this section provides additional empirical evidence regarding the appropriateness of the credibility models for pricing crop (re)insurance policies. In particular, Subsection 2.6.1 discusses how the credibility models can be used to integrate each province's historical data with the pooled historical data of all ten provinces to optimally determine the credibility premium (or equivalently the forecasted LCR) for year 2010. Then by using the province of Manitoba as an example, an

	Homogeneous Model		Heterogeneous Model	
Province	Z_i^H	PX_i^H	Z_i^{Het}	PX_i^{Het}
AB	0.9561	0.1333	0.9526	0.1405
MB	0.9574	0.0884	0.9618	0.1119
ON	0.9548	0.0553	0.9699	0.0639
BC	0.9777	0.0436	0.9774	0.0691
NB	0.9654	0.0930	0.9727	0.1121
NS	0.9729	0.0534	0.9597	0.0914
PEI	0.9737	0.0745	0.9517	0.1030
QC	0.9729	0.0748	0.9792	0.0903
SK	0.9553	0.1135	0.9743	0.1202
NFLD	0.9765	0.1156	0.9638	0.1538

Table 2.8: Predicted Credibility LCR for the Year 2010. Z_i^H and Z_i^{Het}, PX_i^H and PX_i^{Het} are credibility factors and predicted credibility LCRs for the i-th province in the year 2010 under homogeneous and heterogeneous assumptions, respectively.

attempt is made in Section 2.6.2 to reproduce the empirically observed reinsurance data for the year 2010.

2.6.1 Credibility Premium with Erlang Mixture distribution

The crop data for ten Canadian provinces (see Subsection 2.3) is used to conduct an empirical analysis for the credibility models described in the last section to predict the year 2010 LCR for each province. In our analysis, the prior distribution is assumed to be normal and the parameters are estimated using the maximum likelihood estimation (MLE) method. For both homogeneous and heterogeneous risk models, the Erlang Mixture Model is calibrated to the LCR's. Furthermore, parameters such as a_i and b_i $(i = 1, 2, \ldots, d)$ are estimated based on the calibrated Erlang distribution, providing the prior information for each risk category (province). The forecasted credibility LCRs for year 2010 for both risk models (i.e. (2.11) and (2.17), respectively), together with the credibility factors (i.e. (2.13) and (2.19), respectively), are listed in Table 2.8.

The reported results clearly exemplifies the importance of the risk assumptions. The assumption of heterogeneity leads to higher predicted LCR's (relative to homogeneous risk assumption) and thus is more conservative if the model is used for pricing insurance contracts. Provinces such as Manitoba (MB), Ontario (ON), Saskatchewan (SK), New Brunswick (NB) and Quebec (QC) are the major crop producers in Canada, and it is of interest to note that switching from the homogeneous risk model to the heterogeneous risk model results in

increased credibility factors for these provinces, thus giving more credits to their historical data. On the other hand, the credibility factors for provinces such as Nova Scotia (NS), Prince Edward Island (PEI), and Newfoundland and Labrador (NFLD), which comprise a much smaller portion of the crop sector in canada, result in decreased credibility factors. When switching from the homogeneous risk model to the heterogeneous risk model, the credibility factors decrease by 1.366%, 2.285%, and 1.309%, respectively. This means that the predicted LCR's for these provinces are giving more credit to information from the entire collective risk pool.

2.6.2 Manitoba Reinsurance Example

The crop reinsurance treaty that is written based on the LCR is typically a layer reinsurance contract structure(Porth et al., 2013). For example, the reinsurance treaty for the province of Manitoba for the year 2010 has lower attachment and upper level of LCR = 15% and LCR = 27.5%, respectively. This implies that if the observed LCR for the insured year is less than 15%, there is no reinsurance payout. If the observed LCR is greater than 15%, then the reinsurers are liable for the loss in excess of 15% up to a maximum of 12.5% (which is the spread of the attachment points). The actual payout is then adjusted by the liability exposure forecasted at the inception of the contract. Manitoba Agriculture Service Corporation (MASC) retains 10% of the liability so that the remaining 90% of the liability is ceded to private reinsurers. For year 2010, the liability exposure for the private reinsurers was 1,856,000,000 with reinsurance premium $28,700,000.

Using year 2010 reinsurance data from the province of Manitoba, the objective of this subsection is to consider various methods of pricing reinsurance contacts, and assess their effectiveness by relating the price to that observed in the market. To proceed, two critical assumptions are imposed. One is that the reinsurance premium P of the contract is assumed to be determined by the expectation premium principle, defined as

$$P = (1 + \theta)E(L_R), \tag{2.25}$$

where θ is the loading factor and L_R is the reinsurer loss exposure random variable. For the year 2010 Manitoba reinsurance program, we have

$$L_R = \max(X_{MB,2010} - 0.15, 0.125) \cdot L_{i,2010} \tag{2.26}$$

where $X_{i,2010}$ denotes Manitoba's LCR random variable in year 2010. Second we assume

$\theta = 0.35$, which seems to be consistent with crop reinsurance market practice (Porth et al., 2014).

Based on the above assumptions, the remaining task is to explore ways of determining the expectation in (2.25) in order to fully specify the reinsurance premium. Here we examine the following three methods:

> **Method I:** The random variable $X_{MB,2010}$ in (2.26) is assumed to be modeled by the Erlang-mixture model with its parameter values calibrated to the historical LCR data of Manitoba (see Section 2.4). $E(L_R)$ is then evaluated accordingly from the calibrated Erlang-mixture model.
>
> **Method II:** The credibility premium approach as described in Subsection 2.6.
>
> **Method III:** This approach is similar to the preceding credibility approach, but, with the important difference that instead of risk pooling from all ten provinces, the present approach exploits information only from a smaller subset of provinces. Here we use the K-means clustering technique to partition the 10 provinces into groups with similar risk (as measured by the average risk). In our case, the 10 provinces are classified into three groups as shown in Table 2.9. Note that Manitoba is grouped together with the provinces of Ontario, British Columbia, Nova Scotia, Prince Edward Island and Quebec, and this group has the lowest risk exposure (average LCRs is 0.072). The credibility approach is then applied to this group for determining the reinsurance premium.

Note that the Method I corresponds to the classical net premium approach but with risk loading. While Methods II and III are based on the credibility approach, the latter method has the advantage that it relies on a smaller subset of data (i.e. Group I in Table 2.9). This reduces the amount of data needed to be collected and hence it is more manageable, especially where there is a large number of risk categories to begin with. Also, borrowing information from other risk categories with some similar traits appears to be more reasonable and credible.

The results are summarized in Table 2.10. For the credibility based methods II and III, both homogeneous and heterogeneous models are implemented. In addition to the reinsurance premium, the predicted credibility LCRs are also reported. The last column of the table gives the pricing error relative to the actual 2010 reinsurance premium, where positive (negative) value indicates that it is larger (smaller) than the actual price. Using the 2010 price as the actual price, one immediate conclusion that can be drawn from the results is that net premium based only on the Erlang mixture model and the credibility premiums under the

Group	Provinces	Mean LCR
I	MB ON BC NS PEI QC	0.072
II	AB SK	0.120
III	NB NFLD	0.140

Table 2.9: Classification for Ten Regions/Provinces in Canada.

homogeneous assumptions are not satisfactory, as they underestimate the observed premium by more than 10%. These observations are quite surprising since the best goodness of fit, the Erlang mixture model, has a heavier tail and hence should result in a more conservative estimate of the reinsurance premium. However, the reinsurance premium calculated from these methods is still less than the observed market premium.

The impact of changing the assumption from homogeneous to heterogeneous is also highlighted. If we use all 10 provinces in our heterogenous credibility model, then the resulting reinsurance premium is more conservative in that it is higher than the observed premium by almost 10%. On the other hand, if we adopt Method III which uses a smaller subset of provinces that are more alike, the underestimation error reduces from 13.9% to 4.0% by using the heterogenous credibility model, as opposed to the homogeneous model.

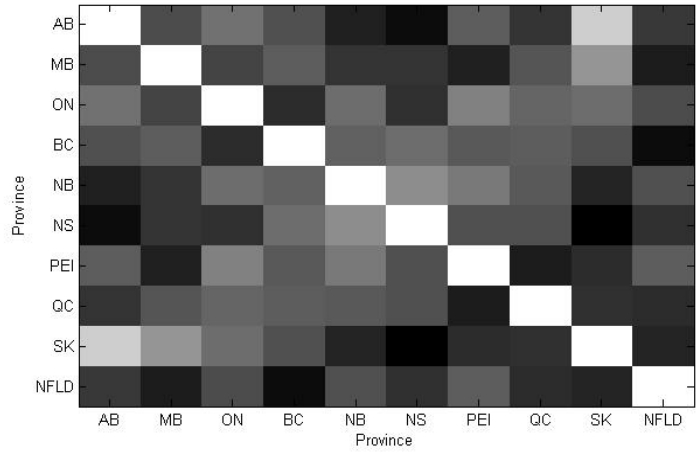


Figure 2.3: Covariance Coefficient Matrix of Ten Regions/Provinces in Canada

2.7 Conclusions

Crop insurance is faced with a major challenge of limited data which leads to concerns in pricing insurance and reinsurance contracts. The objective of this chapter was to address

Method		Z_{MB}	PX_{MB}	Reinsurance Premium	Discrepancy
I		-	-	\$25,138,608	−12.4%
II	Homogeneous	0.9548	0.0884	\$24,906,567	−13.2%
	Heterogeneous	0.9618	0.1119	\$31,538,179	9.9%
III	Homogeneous	0.9573	0.0877	\$24,710,466	−13.9%
	Heterogeneous	0.9614	0.0977	\$27,548,036	−4.0%

Table 2.10: Comparison of credibility factors, forecasted LCRs and reinsurance premiums under various pricing methods. The value under "Discrepancy" column gives the pricing error relative to the year 2010 reinsurance premium of \$28,700,000.

some of the fundamental issues surrounding crop insurance ratemaking, from the perspective of the reinsurer, through the development of a scientific framework that can be consistently and widely adopted by the agricultural sector. We show by simulation that the current tools available for testing trends in time series cannot provide accurate information for small sample sizes as is in the case in crop insurance. A unique data set comprised of the entire crop insurance sector in Canada was analyzed, and from this we show that the Erlang Mixture distribution family, which is newly introduced in this chapter for agricultural insurance ratemaking, captures the heavy tail property of the loss data better than single distribution models. In addition, the Erlang Mixture model is shown to be robust under different trending methods, which is important in dealing with the difficult problem of deciding which trending process to use for small samples.

To enhance the crop (re)insurance pricing framework, a modified credibility model is introduced. Credibility premiums based on two different risk assumptions of the LCR data are calculated, including a homogeneous risk assumption and a heterogeneous risk assumption. Further, the "liability weighted" LCR is proposed to aggregate historical loss data. The results show that the heterogeneous credibility premium based on the liability weighted LCR's is more conservative, and provides a more scientific and robust approach to enhance reinsurance pricing.

2A Appendix: Proofs

2A.1 A proposition of liability weighted LCR's

For T years of observations, define two random variables P, Q as

$$P = \frac{1}{T}\sum_{i=1}^{T} X_i, \tag{2A.27}$$

$$Q = \frac{\sum_{i=1}^{T} X_i L_i}{\sum_{i=1}^{T} L_i}, \tag{2A.28}$$

where X_i, L_i $(i = 1, \ldots, T)$, are the LCR and liability in the i-th year. Therefore, P and Q define two ways to aggregate historical losses, namely, P is the simple average LCR and Q is the liability weighted LCR. Then we have the following proposition.

Proposition 2A.1. *In probability space $(\Omega, \mathcal{F}, P)$, we define the following random variables:*

1. $X : \Omega \mapsto [0, 1]$ *is LCR,*
2. $Y : \Omega \mapsto \mathbb{R}_+$ *is crop yield,*
3. $L : \Omega \mapsto \mathbb{R}_+$ *is liability.*

The relationship holds:

$$X = \frac{(L - Y)_+}{L} = \begin{cases} 0, & Y(\omega) \geq L(\omega), \omega \in \Omega, \\ \frac{L-Y}{L}, & Y(\omega) < L(\omega), \omega \in \Omega. \end{cases} \tag{2A.29}$$

Assume liability is a function of yield, $f : \mathbb{R}_+ \mapsto \mathbb{R}_+$, namely, $L(\omega) = f(Y(\omega))$ for all $\omega \in \Omega_0$, where $\Omega_0 = \{\omega : Y(\omega) < L(\omega)\}$. Then if for all $x \in I = X(\Omega_0)$, function $f(x)$ satisfies the following:

1. $f(x) \geq 0$ *with* $f(0) = 0$,
2. $f'(x) \geq 0$,
3. $f'(x) \geq \frac{f(x)}{x}$.

Then $Q \geq P$ almost surely.

Proof. Let us start the proof by defining $L_{(i)}$ as the ith order statistic of random variables

$L, i = 1, \ldots, T$, i.e.,

$$L_{(1)}(\omega) \leq L_{(2)}(\omega), \ldots, L_{(T)}(\omega), \quad \omega \in \Omega$$

- If $\omega \in \Omega/\Omega_0$, then,

$$X_{(1)}(\omega) \leq X_{(2)}(\omega), \ldots, X_{(T)}(\omega), \quad \omega \in \Omega/\Omega_0$$

  since $X_{(i)}(\omega) \in \Omega/\Omega_0 = 0$.

- If $\omega \in \Omega_0$, let function $l : I \mapsto [0, 1]$ be $l(y) = 1 - \dfrac{y}{f(y)}$. From the assumptions, we have $l'(y) \geq 0$, since

$$l'(x) = \frac{yf'(y) - f(y)}{f^2(y)} \geq 0 \quad \forall y \in I.$$

  So $l(y)$ is an increasing function on I, which means that

$$X_{(1)}(\omega) \leq X_{(2)}(\omega), \ldots, X_{(T)}(\omega), \quad \forall \omega \in \Omega_0$$

Therefore,

$$X_{(1)}(\omega) \leq X_{(2)}(\omega), \ldots, X_{(T)}(\omega), \quad \forall \omega \in \Omega$$

namely,

$$\frac{X_{(1)}}{\sum_{i=1}^{T} L_{(i)}}(\omega) \leq \frac{X_{(2)}}{\sum_{i=1}^{T} L_{(i)}}(\omega), \ldots, \frac{X_{(T)}}{\sum_{i=1}^{T} L_{(i)}}(\omega), \quad \forall \omega \in \Omega,$$

Then,

$$\sum_{i=1}^{T} \frac{1}{T} X_{(i)} \leq \sum_{i=1}^{T} L_{(i)} \frac{X_{(i)}}{\sum_{i=1}^{T} L_{(i)}} \quad \text{a.s.} \tag{2A.30}$$

which is equivalent to that $Q \geq P$ almost surely.
The inequality 2A.30 follows by directly applying the Chebyshev Sum Inequality. If

$$a_1 \leq a_2, \ldots, \leq a_T \tag{2A.31}$$

$$b_1 \leq b_2, \ldots, \leq b_T \tag{2A.32}$$

$\frac{1}{T} \sum_{i=1}^{T} a_i \sum_{i=1}^{T} b_i \leq \sum_{i}^{T} a_i b_i$

□

2A.2 Deriving the Cedibility Premiums

Credibility premiums search for the best linear combination of the past observations in the sense of minimizing the quadratic loss (for simplification of the notation, in the following derivation we suppress the subscript i, which indicates the risk category.):

$$QL = E[(\mu(\Theta) - \alpha_0 - \sum_{j=1}^{T} \alpha_j X_j)^2]$$

Hence, the solution $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_N)'$ needs to satisfy the following for every $k = 1, 2, \ldots, T$:

$$\frac{\partial QL}{\partial \alpha_0} = (-2)\dot{E}(\mu(\Theta) - \alpha_0 - \sum_{j=1}^{T} \alpha_j X_j) = 0$$
$$\frac{\partial QL}{\partial \alpha_k} = (-2)\dot{E}[X_k(\mu(\Theta) - \alpha_0 - \sum_{j=1}^{T} \alpha_j X_j)] = 0$$

Namely,

$$E(\mu(\Theta)) = \alpha_0 + \sum_{j=1}^{T} \alpha_j E(X_j)$$
$$E(\mu(\Theta)X_k) = \alpha_0 + \sum_{j=1}^{T} \alpha_j E(X_j X_k)$$

Also note that:

$$\begin{aligned} E(\mu(\Theta)) &= E(E(X_{T+1}|\Theta)) = E(X_{T+1}) \\ E(\mu(\Theta)X_k) &= E(E(\mu(\Theta)X_k|\Theta)) \\ &= E(\mu(\Theta)E(X_k|\Theta)) \\ &= E(E(X_{T+1}|\mu(\Theta))E(X_k|\Theta)). \\ &= E(X_{T+1}X_k) \end{aligned}$$

Thus

$$E(X_{T+1}) = \alpha_0 + \sum_{j=1}^{T} \alpha_j E(X_j) \tag{2A.33}$$

$$Cov(X_{T+1}, X_k) = \sum_{j=1}^{T} \alpha_j Cov(X_j, X_k). \tag{2A.34}$$

The above two equations are called "Normal Equations".
For M1:

$$\begin{aligned} Cov(X_{T+1}, X_k) &= Cov(E(X_{T+1}|\Theta), E(X_k|\Theta)) + E(Cov(X_{T+1}|\Theta, X_k|\Theta) \\ &= a^2 a^H \\ Cov(X_j, X_k) &= Cov(E(X_j|\Theta), E(X_k|\Theta)) + E(Cov(X_j|\Theta, X_k|\Theta) \\ &= a^2 a^H + \frac{a}{b} v^H \delta_{kj} \end{aligned}$$

where $\delta_{kj} = \begin{cases} 0, & k = j, \\ 1, & \text{else} \end{cases}$
From (2A.33),

$$a\mu_0 = \alpha_0 + a\mu_0 \sum_{j=1}^{T} \alpha_j. \tag{2A.35}$$

From (2A.34),

$$\begin{aligned} a^2 a^H &= \sum_{j=1, j\neq k}^{T} \alpha_j a^2 a^H + a_k (a^2 a^H + \frac{a}{b} v^H) \\ &= \sum_{j=1}^{N} a^2 a^H \alpha_j + \frac{a}{b} v^H \alpha_k. \end{aligned}$$

Hence,

$$\alpha_0 = \frac{\frac{v^H}{a^H}}{Nab + \frac{v^H}{a^H}} a\mu_0, \quad \alpha_j = \frac{ab}{Nab + \frac{v^H}{a^H}}$$

Thus, we finally get:

$$CreP^H = Z^H \bar{X}^H + (1 - Z^H)\mu$$

where $Z^H = \frac{Nab}{Nab+\frac{v^H}{a^H}}$, $\mu = a\mu_0$, and $\bar{X}^H = \frac{1}{T}\sum_{j=1}^{T} X_k$.
For M2, similarly:

$$\begin{aligned} Cov(X_{T+1}, X_k) &= Cov(E(X_{T+1}|\Theta), E(X_k|\Theta)) + E(Cov(X_{T+1}|\Theta, X_k|\Theta) \\ &= Var(E(X_k|\Theta)) + 0 \\ &= a^2 a^{\text{Het}} \\ Cov(X_j, X_k) &= Cov(E(X_j|\Theta), E(X_k|\Theta)) + E(Cov(X_j|\Theta, X_k|\Theta) \\ &= a^2 a^{\text{Het}} + \frac{a}{bL_k} v^{\text{Het}} \delta_{kj}. \end{aligned}$$

From (2A.33),

$$a\mu_0 = \alpha_0 + a\mu_0 \sum_{j=1}^{T} \alpha_j \tag{2A.36}$$

From (2A.34),

$$\begin{aligned} a^2 a^{\text{Het}} &= \sum_{j=1, j\neq k}^{T} \alpha_j a^2 a^{\text{Het}} + a_k (a^2 a^{\text{Het}} + \frac{a}{bL_j} v^{\text{Het}}) \\ &= \sum_{j=1}^{N} a^2 a^{\text{Het}} \alpha_j + \frac{a}{bL_j} v^{\text{Het}} \alpha_k \end{aligned}$$

Hence, by defining $L = \sum_{j=1}^{T} L_j$, we have

$$\alpha_0 = \frac{\frac{v^{\text{Het}}}{a^{\text{Het}}}}{Lab + \frac{v^{\text{Het}}}{a^{\text{Het}}}} a\mu_0, \quad \alpha_k = \frac{L_k ab}{Lab + \frac{v^{\text{Het}}}{a^{\text{Het}}}}$$

Thus, we finally get:

$$CreP^{\text{Het}} = Z^{\text{Het}} \bar{X}^{\text{Het}} + (1 - Z^{\text{Het}})\mu$$

where $Z^{\text{Het}} = \frac{Lab}{Lab + \frac{v^{\text{Het}}}{a^{\text{Het}}}}$, $\mu = a\mu_0$, and $\bar{X}^{\text{Het}} = \sum_{j=1}^{T} \frac{L_j}{L} X_j$.

Chapter 3

A Credibility-based Yield Forecasting Model for Crop Reinsurance Ratemaking and Weather Risk Management

3.1 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 agricultural insurance market worldwide (Ozaki et al., 2008). However, in the presence of systemic weather risks, there are many challenges in efficiently and accurately forecasting crop yields, including effects of possible climate changes, selecting predicting variables, restating crop mix, and modeling geographical differences across regions.

The objective of this chapter 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 add to the literature by proposing scientific approach to restate yields through consideration of changing crop mix over the historical years in order to maintain the consistency of data. In addition, a comprehensive weather index system is composed to reflect the nonlinearity relationship of crop yields and weather variables. Efficient algorithms for selecting an optimal predic-

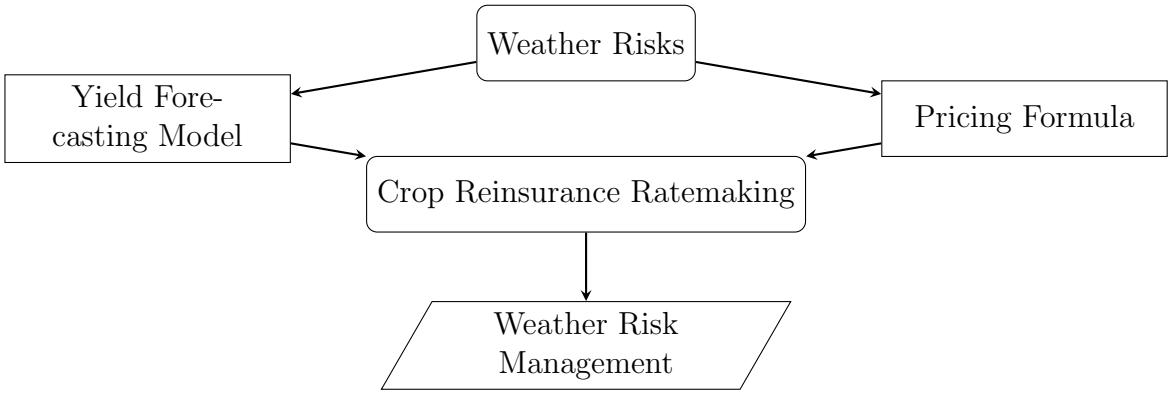


Figure 3.1: Flow chat of the general modeling framework in Chapter 3.

tion 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 chapter. Finally, weather risk management for reinsurers is discussed by proposing a *Division to Integration* risk management procedure that incorporates the forecasting and pricing model developed.

An overview of the modeling framework of this chapter can be expressed in the flow chart shown in Figure 3.1. One of the essential contributions of this chapter is that it provides a comprehensive framework for reinsurance companies to address weather risks by constructing an integrated management framework with analytical formulas that include model construction, premium calculation, and risk assessment.

Climate Change & Weather Risks

Agricultural risk management is faced with a number of challenges, largely due to exposure of natural catastrophes, creating losses that are at times spatially correlated. Some findings suggest that possible effects of climate change may contribute to more prevalent extreme weather, namely, extreme heat, drought, wildfires, and heavy precipitation, which are occuring more frequently and with more intensity (Cassman, 1999; Conway and Toenniessen, 1999; IPCC, 2013; Motha and Baier, 2005; Pall et al., 2011; Salinger, 2005). From both a short and long term perspective, the economic impacts brought by extreme weather events can have significant implications. Lacking information and knowledge of the covariant weather risks can create a number of challenges for different stakeholders including, farmers, insurers and reinsurers, and governments.

In recent years, the Property & Casualty (P&C) industry has experienced tremendous losses.

In 2012 for example, there were approximately $27.8 billion of industry losses in Canada, among which a large amount of losses were from severe weather events (Insurance Bureau of Canada (IBC), 2014). In the wake of these losses, however, reinsurers were able to maintain capacity and absorb severe losses. This outcome has helped lend support to the reinsurance market in terms of an efficient and cost effective risk management approach to help reduce risk exposure and vulnerability. Therefore, agricultural insurers often seek to improve weather risks diversification through the use of such risk management tools as reinsurance. Further, in some regions the availability of reinsurance capacity drives the success and sustainability of the agricultural insurance program.

Restatement of Crop Mix

The restatement of crop mix is an important issue in agricultural (re)insurance design and pricing. The evolution of farming practices creates difficulties for loss modeling given that historical dates 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 to diversify and to protect against losses from extreme adverse weather conditions. Another common farming practice to reduce risk is crop rotation, in which farmers plant different type of crops in different years to help give nutrients to the soil and help to mitigate the build-up of pathogens and pests. Further, changes in biotechnology and increases in commodity prices have influenced the crop that farmers plant, and these changes in crop mix over time also create concerns over the representativeness of the data. Therefore, a scientific restatement of historical yield data helps to ensure that the historical observations are good indicators of future crop production (Coble et al., 2011; Woodard, 2014).

Despite its importance, yield restatement has rarely been discussed in the literature. In this chapter, we provide an algorithm to restate crop mix for yield forecasting. Serving as a pre-process step, the restatement algorithm helps prepare the data so that it is more representative of present farming conditions and the overall risk profile of the current program.

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 (MPCI). 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., 1994). In alleviate these issues, recent research has focused on index-based insurance (IBI). Other issues associate with IBI, on the other hand, surface. One critical issue is the 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 how the crop yield in response to the weather becomes extremely important in reducing basis risk.

The crop growth process is a nature system based on the interaction of soil, air, water, and crops. An integrated yield model combining meteorological and climate data is beneficial for farmers to understand the influences of weather variables and adjust their cultivating strategies accordingly (Campbell and Diebold, 2005; Kleibera et al., 2011). A major challenge faced with integrating yield and weather information is how to efficiently select scientific variables from a complex set of correlated weather variables and reduce the dimension of explanatory variables to an acceptable number. In this chapter, a comprehensive weather index system is considered to describe the nonlinear relationship of weather variables and crop yields. Three efficient model selection algorithms are proposed by combining Screening Regression (SR), Principal component Analysis (PCA) and Cross Validation (CV), which efficiently help to achieve the goal of model selection and dimension reduction. Empirical results show that compared to the traditional multiple regression method, the model selection algorithms have better in-sample and out-of-sample forecasting abilities.

Geographical Heterogeneity & Credibility Approach

Weather variables impact crop yield differently across various regions (i.e., geographical heterogeneity), and this has been studied on a very limited basis (Cai et al., 2013). McCarl et al. (2008) found that the impact of climate change varies among different regions in the U.S. In our study, the effect of spatial heterogeneity is explored, and for certain dominant weather variables there may be a positive effect on yield in some municipalities, yet a negative impact in other geographical regions. To examine this issue, a new credibility estimator is proposed to consider the geographical heterogeneity. In addition, the new proposed credibility estimator also shows several other advantages, including unbiasedness and smaller mean quadratic loss compared to classic regression credibility model.

Main Contributions in This Chapter

This chapter contributes the literature in both methodology development and practical applications. In therm of methodology, we propose a new credibility estimator that incorporates weather variables and production information in principal componentgeographically correlated regions (See Equation (3.9)). We prove that this new credibility estimator is an unbiased estimator with smaller mean quadratic loss compared to the classic regression

credibility estimator (Proposition 3.3.1). Closed form reinsurance pricing formulas are also derived (Proposition 3.3.2).

This chapter also contributes the practical reinsurance pricing and risk management. In particular, for the first time, we address the issue of crop mix restatement in reinsurance pricing, and propose a restatement algorithm to maintain the consistency of historical data (Algorithm 3.2.1). Efficient algorithms that combine screening regression (SR), cross validation (CV) and principal component analysis (PCA) are developed to achieve efficient dimension reduction and model selection (Algorithms 3.2.2 to 3.2.4). Empirical results show that the crop yield forecasting model proposed in this paper has improved both in-sample and out-of-sample forecasting abilities. In addition, a comprehensive weather index system is developed to reflect the nonlinearity between crop yields and weather variables (Table 3.1). Finally, based on the framework discussed in this paper, we propose a Division to Integration weather risk management procedure, providing practical risk management suggestions to agricultural reinsurers.

The remainder of this chapter proceeds as follows. Section 3.2 constructs a comprehensive weather index system and introduces three model selection algorithms to facilitate crop yield forecasting. Section 3.3 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 3.4 provides conclusion remarks with weather risk management suggestions for agricultural reinsurers. The appendix outlines the proofs.

3.2 Yield Forecasting

Agriculture reinsurance is faced with the challenge of limited loss experience as there is usually only one growing season for many crops per year in Canada. Further, reinsurers often have access to highly aggregated loss data, typically at the county level rather than the farm level. As a result, an accurate crop yield forecasting model is essential to improve agricultural reinsurance ratemaking. Accurate crop yield predictions, particularly those with high out-of-sample prediction ability, provide important information for agriculture reinsurance companies to measure loss conditions ahead of time, and hence assist in computing fair premium rates that are sustainable in the long term. In this section, three model selection algorithms are proposed to facilitate crop yield forecasting by achieving better in-sample and out-of-sample forecasting abilities.

3.2.1 Data Introduction

In this chapter, a detailed farm-level crop yield data set from Manitoba, Canada, is studied. The data panel covers 216 types of crops from 19238 farms from 1996 to 2011. The weather data analyzed is the Adjusted and Homogenized Canadian Climate Data (AHCC) from Environment Canada. It includes daily temperature (maximum, minimum, and mean) from 24 weather stations, and daily precipitation from 30 weather stations in Manitoba. The weather data contains some missing points, therefore, in order to address the missing data problem in the weather dataset, the Ordinary Kriging method is employed. "Kriging" is synonymous with "optimal prediction", which predicts unknown values from data observed at known locations. This term was first used by Matheron in 1963 in honour of Krige, who did preceding work on this method. Through the use of geostatistic spatial model, such as Kriging, missing data are optimally predicted by minimizing the squared prediction errors (Matheron, 1963; Plant, 2012).

3.2.2 Restatement of 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 condition, or any other factor that has significant impact on crop production (Coble et al., 2011; Woodard, 2014).

While the restatement of crop mix is important, it should be emphasized that there is scarce literature which formally addresses this issue. The approach that we describe below is based on conversations with some practitioners, and is consistent with market practice. Some important points about the restatement procedure 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 land area that is actually used, 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 is comparable to the county-level yield data in the U.S. system.

Before describing our restatement procedure, it is useful to recall some previous notations as well as to define some new terms:

- 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 at 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, respectively, the yield and acres for risk category i, farm j, crop k and in year t.

The first step in our proposed restatement procedure is to identify the "main crop mix" that the municipality has been producing in recent years. The "main crop mix" is defined as the minimum number of crop mix that covers at least 90% 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 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 for the next year's farming practice and hence it will be used as a "benchmark" for restating all the historical crop yields. It is noteworthy that this "optimal crop mix" is not unique, and depends on the definition of "main crop mix". By denoting K^* as the total number of crops in the "main crop mix" for a given municipality, the set $\mathcal{K} = \{0, k_1, k_2, \ldots, k_{K^*}\}$ 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, we identify, for each municipality and for

each year, 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 $\mathcal{J} = \{j_1, j_2, \ldots, j_{J^*}\}$ to denote the set of indexes for identifying these farms.

The restatement procedure is summarized in the following algorithm:
Algorithm 3.2.1. *For each municipality $i = 1, \ldots, d$,*

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

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

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

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

$$a_{i,j,k,t} = \frac{A_{i,j,k,t}}{\sum_{l \in \mathcal{K}} A_{i,j,l,t}} \tag{3.1}$$

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 $j \in \mathcal{J}$ and $k \in \mathcal{K}$, the corresponding coefficients are optimally determined by minimizing the sum of square errors as follows:

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

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*

$$y^R_{i,t} = \frac{\sum_{j \in \mathcal{J}} \sum_{k \in \mathcal{K}} a^R_{i,j,k,t} y_{i,j,k,t}}{\sum_{j \in \mathcal{J}} \sum_{k \in \mathcal{K}} A_{i,j,k,t}} \tag{3.2}$$

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.

3.2.3 Weather Index System

Given the high sensitivity and low frequency nature of agricultural risks, insurers often cede a portion of the risk in their portfolio to private reinsurers in order to improve diversification. Given the data limitations reinsurers are faced with, reinsurers seek to improve loss forecasting approaches. This section develops a comprehensive weather index system to account for nonlinear impacts of weather variables on crop production in order to improve yield forecasting, and hence loss forecasting. The development of a comprehensive weather index system may also be useful for the weather derivatives and weather-linked insurance market, as reinsurers require for more sophisticated methods to help quantify weather risk conditions.

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, 2008b) 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 ($\widetilde{T}$); otherwise it is the difference between daily temperature and the base temperature. More explicitly, the value of GDD on day t is

$$GDD_t \quad = \quad \max(0, T_t - \widetilde{T}), \tag{3.3}$$

where T_t is the average of maximum and minimum daily temperature, defined as, $T_t = \frac{T_{min} + T_{max}}{2}$, and $\widetilde{T}$ is the base temperature. Besides the GDD, popular weather indexes include the Cooling Degree Days (CDD), Heating Degree Days (HDD) and Cumulative Average Temperature (CAT) [1]. Another index system, called Crop Heat Units (CHU), is calculated from calibrated daily minimum and maximum temperatures (Brown, 1969). It was originally developed for field corn and has been in used in Ontario for the last 30 years.

In this chapter, a detailed 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,

[1] For more detailed definitions, refer to Alaton et al. (2002), Campbell and Diebold (2005) and Alexandridis and Zapranis (2013).

$\theta_1 = 10°C, \theta_2 = 4.4°C$, and $\theta_3 = 30°C$. As such, we use this as a benchmark and analyze a wide range of the temperature thresholds in order to assess the detailed relationships between the crop yields and weather variables, which will help in the forecasting of yields. 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 for both a monthly and an annual basis. The advantage of studying the monthly weather indices is that more detailed dynamic between the crop yields and weather variables is 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 notations of the weather index system and aggregating functions are listed below. Note that all the weather indices are indexed by t, we suppress the index for brevity.

Weather Index System

- Night Growing Degree Low (NGDL)

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

  where $\theta_{11} = 0, \theta_{12} = 1, \theta_{13} = 2, \theta_{14} = 3, \theta_{15} = 4(°C)$, 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 $\theta_{31} = 26, \theta_{32} = 27, \theta_{33} = 28, \theta_{34} = 29, \theta_{35} = 30(^{\circ}C)$, 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 0.25% quantile of historical precipitations.

- Precipitation Low (PREL)

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

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

The aggregation functions considered in this chapter 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:

- **Average Index:** Using function "avg", the average indices provide aggregate measures of weather conditions during a defined period.
- **Extreme Events:** Using function "min" and "max", these indices describe extreme events during a defined period.
- **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 notations 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 zero during May, which directly provides knowledge about the downside low temperature risk during May. The detailed notations and the interpretations of these explanatory variables are listed in Table 3.1.

Table 3.1: Notations and Definitions of Weather Indices in the Full Design Matrix. "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", "Jun", "Jul", "Aug", "Sep", and "Oct".

Index (I)	Threshold ($^\circ C$ / 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
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 $\theta_{11} = 0^\circ C$) 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 $\theta_{21} = 8^\circ C$) 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 $\theta_{35} = 30^\circ C$) 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

3.2.4 Model Selection Algorithms

In geographic and agronomic science, 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

$$Y_t = \beta_0 + \beta_1 V_{1,t} + \beta_2 V_{2,t} + \ldots + \epsilon_t, \tag{3.4}$$

where Y_t represents the crop yields which could be either time series for 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 effect the predicting results (Ramsey and Schafer, 2013). Common modeling practices use either the *all possible subsets method* or *stepwise methods*. To be more specific, the *all possible subsets method* compares possible combinations of explanatory variables and uses statistics such as Mallows' C_p (Mallows, 1973) 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 experience-based method. Since in addition to referring to statistical tools such as C_p, AIC or BIC, knowledge of the biophysics, agronomy, and ecology are required in the variable selection procedure, which can often be complicated and expensive.

In our study, three model selection algorithms are proposed to help achieve the objectives of model selection and dimension reduction, combining Screening Regression (SR) with principal component Analysis (PCA). 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 (PC's), and retains the variation of the data as much as possible. The optimal threshold of screening is identified through cross validation (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), so that 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 more detailed definition of PCA and CV, refer to Jolliffe (2002) and Kohavi (1995). The details of these algorithms will be presented in Algorithms 3.2.2, 3.2.3, and 3.2.4. 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 3.2.2. ***(Screening Regression (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 $\boldsymbol{\rho}_i = (\rho_i^{(1)}, \ldots, \rho_i^{(r)})'$,*

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

Step 3a. *Update the design matrix according to the 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 the 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 3.2.3. ***(PCA Screening Regression (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 $s_i^{(0)}$ PC's that retain 85% or more variance of the full design matrix.*

Step 2. *Calculate the 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 $\boldsymbol{\rho}_i = (\rho_i^{(1)}, \ldots, \rho_i^{(r)})'$*

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

Step 4a. *Update the 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 the 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 3.2.4. ***(Screening PCA Regression (SPCAR) Algorithm)*** *For each municipality i, $i = 1, \ldots, n, n = 122$:*

Step 1. *Calculate the 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 $\boldsymbol{\rho}_i = (\rho_i^{(1)}, \ldots, \rho_i^{(r)})'$.*

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

Step 3a. *Update the 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 3. *Calculate the 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 4. *Do PCA transformation to design matrix W_i^* and get Z_i^*, containing s_i^* PC's that retain 85% or more variance of W_i^*. Z_i^* is the optimal design matrix.*

3.2.5 Yield Forecasting Results

The three algorithms in Section 3.2.4 are applied to each of the 122 municipalities in the Manitoba dataset. To demonstrate that these algorithms are effective at dimension reduction, we also execute the classical regression method expressed in Equation 3.4 to the 122 municipalities. We find that the proposed algorithms can reduce the 140-dimensional explanatory matrix to a manageable set for all 122 municipalities that we have investigated. In addition to assessing dimension reduction, some statistical measures such as the Akaike Information Criterion (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 three algorithms and classic regression method are depicted in Figure 3.2 and the corresponding AR2 are depicted in Figure 3.3.

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 method (CR) are listed in the first column of Table 3.2, while the forecasting results based on the three algorithms are reported in columns 3 to 5 of Table 3.2. In the table, "CR" represents the classical regression method, "SR" is the screening regression algorithm, "PCASR" is the PCA screening regression 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. Form Figure 3.2 and Figure 3.3, 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. According to the AR2 results in Figure 3.3, SR and SPCAR algorithms create Adjusted R^2 larger than 0.5, while CR and PCASR perform worse with most AR2 smaller than 0.5.

- All three proposed algorithms improve the forecasting performance in terms of both in-sample and out-of-sample criteria. For example, from Table 3.2 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.

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 yields 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 that 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% better

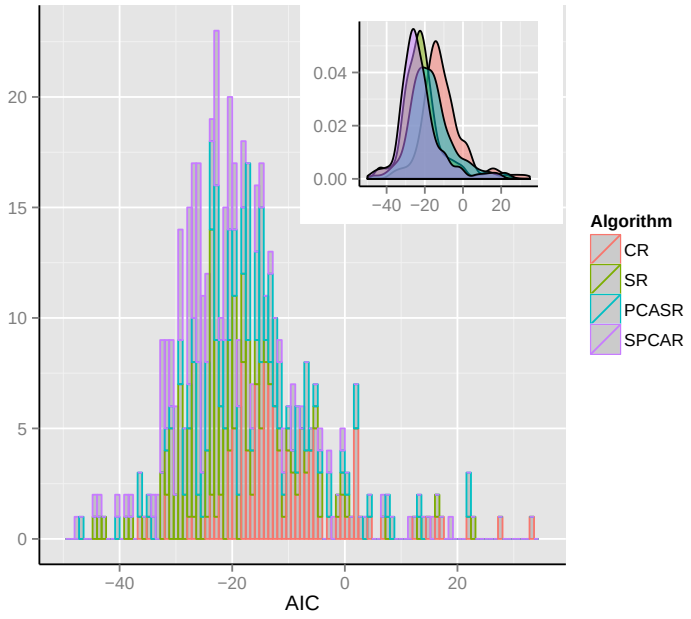


Figure 3.2: Histogram and Estimated Density of AICs for Thtree Algorithms. "CR" the represents classical regression method, "SR" is the screening regression algorithm, "PCASR" is the PCA screening regression method, and "SPCAR" is the screening PCA regression algorithm.

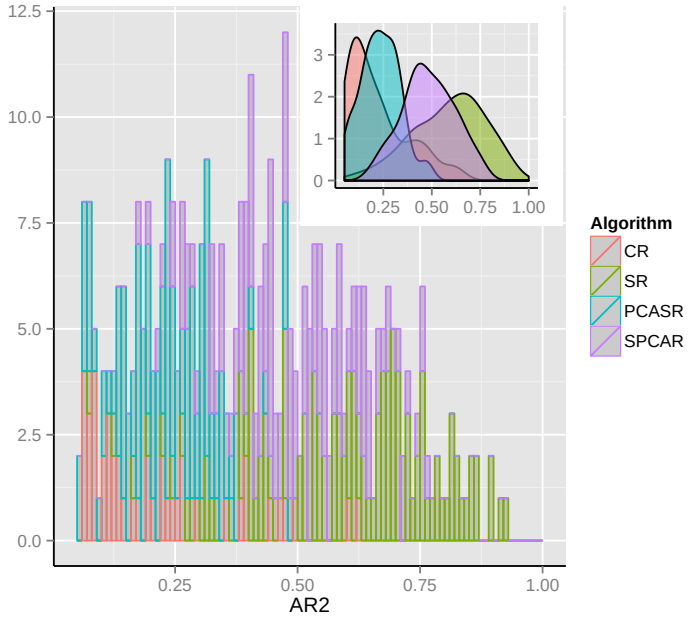


Figure 3.3: Histogram and Estimated Density of AR2 for Three Algorithms. "CR" the represents classical regression method, "SR" is the screening regression algorithm, "PCASR" is the PCA screening regression method, and "SPCAR" is the screening PCA regression algorithm.

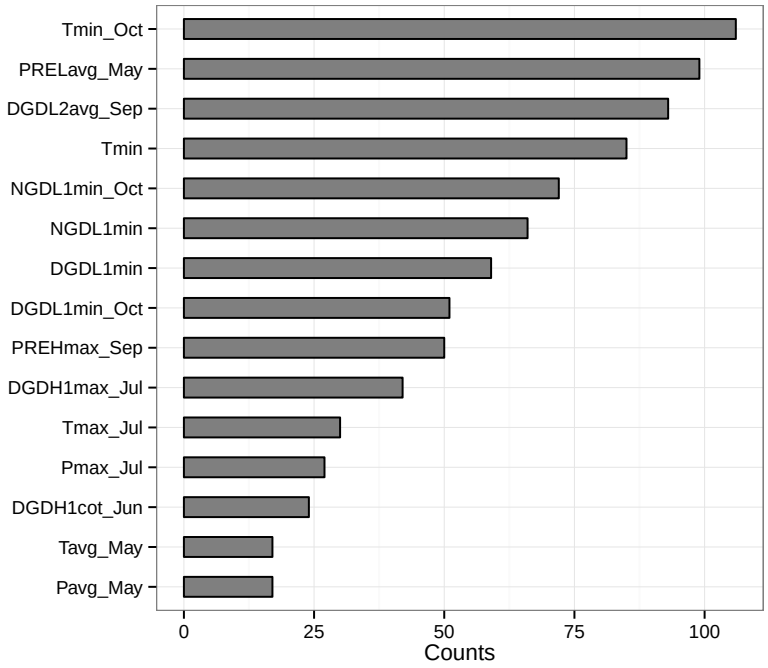
than PCASR (0.0142 v.s. 0.0219) while the out-of-sample of SPCAR is 4% worse than PCASR (0.0342 v.s. 0.0329). Considering that the out-of-sample forecasting capability is of greater importance for forecasting yields, we propose that the SPCAR is of more interest in yield forecasting from reducing the out-of-sample forecasting error point of view.

A main advantage of using the SPCAR is reducing 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 3.4. "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.

Table 3.2: Summary of Forecasting Results. Important statistics are summarized in this table including the mean, median, standard deviation (SD), and maximum errors (Max).

In-Sample Errors (MSE)				
Statistic	CR	SR	PCASR	SPCAR
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)				
Statistic	CR	SR	PCASR	SPCAR
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

Figure 3.4: Count of "Dominant" Weather Indexes for Each Municipality.



3.3 Pricing Framework

In this section, a reinsurance pricing framework is developed, which includes a new crop yield forecasting estimator to improve out-of-sample forecasting ability, as well as closed form pricing formulas. Within this framework, the proposed forecasting estimator is constructed using an adjusted regression credibility estimator to take into consideration the geographical heterogeneity is taken into consideration in the reinsurance pricing. 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.

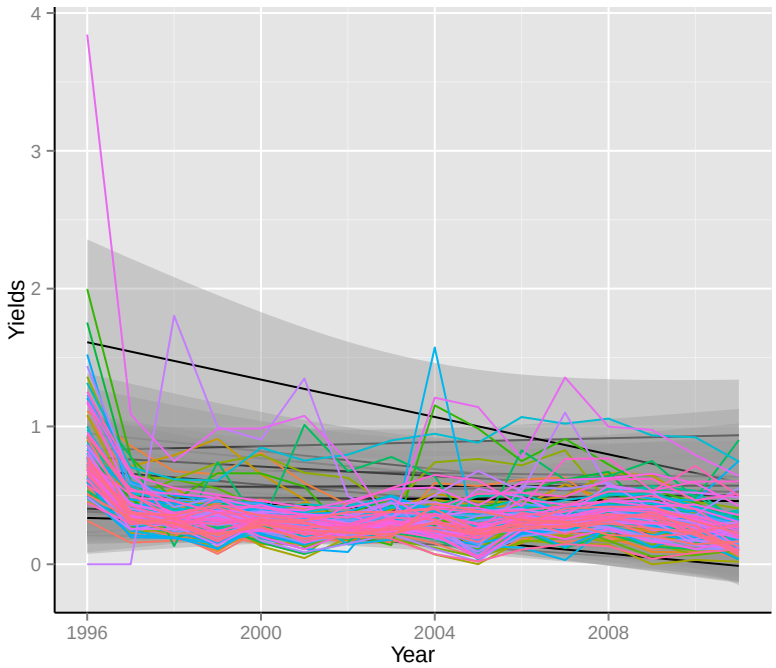
3.3.1 Geographical Heterogeneity

Geographical heterogeneity has been studied on a very limited basis in literature, yet, plays an important role in agricultural reinsurance pricing. Actuarial principles dictate that the 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 faced.

Unfortunately, the risk portfolios of agricultural reinsurers are usually time and spatially dependent and highly heterogeneous. Figure 3.5 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 3.2.4 in Table 3.3, which summarizes the correlations of yields in 122 municipalities and the "key" weather indices. Take "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 some municipalities while it may have negative impact on others. This geographical heterogeneity indicates that the traditional pricing method for crop reinsurance may fail to consider the spatial differences in the reinsurer's risk portfolio, therefore a new reinsurance pricing framework should be considered. In the next subsection, a new credibility estimator that integrates weather information and considers the geographical heterogeneity is proposed.

Figure 3.5: Crop Yields for 122 Municipalities in Manitoba, Canada (1996-2011).



3.3.2 New Credibility Estimator

Let us asssume that there are d risk categories, and let i.i.d random variable Θ_i describes 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. 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:

$$Y(\theta_i) = \mu_Y(\theta_i) + \sigma_Y(\theta_i)\varepsilon_i, \tag{3.5}$$

where ε_i is a standardized random variable with mean zero 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 and Kahn (1975) formulate 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 is 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),

Table 3.3: Summary of Correlation Coefficients of important Weather Indices. Important statistics are summarized in this table, including the mean ($\mathsf{E}(\rho)$), median, standard deviation ($\sqrt{\mathsf{Var}(\rho)}$), maximum correlations ($\rho_{\max}$), and minimum correlations ($\rho_{\min}$).

Index	$\rho_{\min}$	$\rho_{\max}$	$\mathsf{E}(\rho)$	median(ρ)	$\sqrt{\mathsf{Var}(\rho)}$
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

can be expressed as

$$\hat{\mu}_Y^C = W_i^C[A_i^C\beta^C(\theta_i) + (\mathbb{I} - A_i^C)\beta_0^C], \tag{3.6}$$

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

$$\begin{aligned}
\beta_0^Y &= E(\beta^C(\theta_i)), \\
\sigma_Y^2 &= E(\sigma_Y^2(\theta_i)), \\
\Sigma_i^Y &= \sigma_Y^2\big((W_i^Y)'W_i^Y\big)^{-1}, \\
\Gamma_i^Y &= Cov\big(\beta^C(\theta_i), (\beta^C(\theta_i))'\big), \\
A_i^Y &= \Gamma_i^Y(\Gamma_i^Y + \Sigma_i^Y)^{-1}.
\end{aligned}$$

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 $Z(\theta_i)$ by assuming that it can be fully specified by the following regression model so that there is no model risk:

$$Z(\theta_i) = \mu_Z(\theta_i) = W_i\beta(\theta_i) + \sigma_Z(\theta_i)\epsilon. \tag{3.7}$$

where W_i is a $1 \times p$ design matrix, and $\beta(\theta_i)$ is $p \times 1$ vector of regression coefficients. We further assume that $Z(\theta_i)$ is related to $Y(\theta_i)$ through the correlation coefficients defined as $\rho_{YZ} = \dfrac{Cov(\mu_Y(\theta_i)\mu_Z(\theta_i))}{\sqrt{Var(\mu_Y(\theta_i))Var(\mu_Z(\theta_i))}}$.

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 according to (see, for example Bühlmann and Gisler, 2005):

$$\begin{aligned}
\widehat{\mu_Y(\theta_i)} &= \mathsf{E}\big(\mu_Y(\theta)|L(1, Z)\big), \\
&= \mathsf{E}\Big(Proj\big(\mu_Y(\theta)|L(1, Z, \mu_Z(\theta_i))\big)|L(1, Z)\big)\Big),
\end{aligned} \tag{3.8}$$

where $L(1, X_1, \ldots, X_n)$ denotes the linear space spanned by $1, X_1, \ldots, X_n$. Since $L(1, Z) \subset L(1, Z, \mu_Z(\theta_i))$, and $\varrho\big(L(1, \mu_Z(\theta_i))\big) = \varrho\big(L(1, Z, \mu_Z(\theta_i))\big)$ ($\varrho(X)$ denotes the related σ-field

generated by random variable X), therefore, by defining

$$\begin{aligned}\widetilde{\mu_Y(\theta_i)} &= \mathsf{E}\Big(\mu_Y(\theta_i)|L(1,Z,\mu_Z(\theta_i)))\Big)\\ &= \mathsf{E}\Big(\mu_Y(\theta_i)|L(1,\mu_Z(\theta_i)))\Big),\end{aligned}$$

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

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

Thus, it needs to satisfy the normal equations (see Bühlmann and Gisler, 2005; Corolary 3.17)

$$\begin{aligned}E(\widetilde{\mu_Y(\theta_i)}) &= a + bE(\mu_Z(\theta_i)) = E(\mu_Y(\theta_i)),\\ Cov(\widetilde{\mu_Y(\theta_i)},\mu_Z(\theta_i)) &= bVar(\mu_Z(\theta_i)) = Cov(\mu_Y(\theta_i),\mu_Z(\theta_i)).\end{aligned}$$

Solving the above normal equations yields

$$\begin{aligned}a &= E(\mu_Y(\theta_i)) - bE(\mu_Z(\theta_i)),\\ b &= \frac{Cov(\mu_Y(\theta_i),\mu_Z(\theta_i)}{Var(\theta_Z(\theta_i))}.\end{aligned}$$

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

$$\begin{aligned}&\mu_Y = E(\mu_Y(\theta_i)), \quad \tau_Y = \sqrt{Var(\mu_Y(\theta_i))}, \quad \tau_Z = \sqrt{Var(\mu_Z(\theta_i))},\\ &\sigma_Z^2 = E(\sigma_Z^2(\theta_i)), \quad M_i = \sigma_Z^2(W_i^TW_i)^{-1}, \quad \beta_0 = E(\beta(\theta_i)),\\ &A_i = \Gamma_i(\Gamma_i + M_i)^{-1}, \mu_Z = E(\mu_Z(\theta_i)) = E(W_i\beta(\theta_i)),\\ &\Gamma_i = Cov\big(\beta(\theta_i),\beta^T(\theta_i)\big), \tau_{YZ} = Cov(\mu_Y(\theta_i),\mu_Z(\theta_i)).\end{aligned}$$

Given the regression credibility estimator of $Z(\theta_i)$, we can derive the new forecasting esti-

mator $\hat{\mu}_Y^N$ as,

$$\begin{aligned}
\hat{\mu}_Y^N &= b\mu_Z^C + a, \\
&= \mu_Y - \frac{\tau_{YZ}}{\tau_Z^2}\mu_Z + \frac{\tau_{YZ}}{\tau_Z^2}W_i\Big(A_i\beta(\theta_i) + (\mathbb{I} - A_i)\beta_0\Big) \\
&= \mu_Y + \frac{\tau_{YZ}}{\tau_Z^2}\Big(W_iA_i(\beta(\theta_i) - \beta_0))\Big), \\
&= \mu_Y + \rho_{YZ}\frac{\tau_Y}{\tau_Z}\Big(W_iA_i(\beta(\theta_i) - \beta_0))\Big). \qquad (3.9)
\end{aligned}$$

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 in 1976 (Vylder, 1976a;b) and has been applied in health insurance for adjusting large claims. A closer investigation of the expression of equation (3.9) 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 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.

Proposition 3.3.1. *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 (3.9) is no greater than that of $\hat{\mu_Y^C}$ defined in equation (3.6). In other words,*

$$E\big(\hat{\mu}_Y^N - \mu_Y(\theta_i)\big)^2 \leq E\big(\hat{\mu}_Y^C - \mu_Y(\theta_i)\big)^2. \qquad (3.10)$$

Proof. Details are provided in Appendix 3A.1. □

Table 3.4 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, and the classical regression credibility estimator. The second column in Table 3.4 shows the forecasting results for the SPCAR, the third column is for classical regression credibility estimator, which is denoted as "RegCred", and the last column shows the results for the new credibility estimator, denoted as "NewCred". We can see that from both in-sample and out-of-sample points of view, the new credibility estimator proposed in this chapter has better forecasting abilities, largely due to the fact that the new credibility

estimator adjusts the model risk by using parameter ρ_{YZ}.

Table 3.4: Summary of Credibility Forecasting Results. 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.

In-Sample Errors (MSE)			
Statistic	SPCAR	RegCred	NewCred
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)			
Statistic	SPCAR	RegCred	NewCred
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

3.3.3 Reinsurance Pricing Formula

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).

To be more specific, the liability and indemnity for the ith risk category can be expressed as:

$$L(\theta_i) = c \cdot E\big(\mu_Y(\theta_i)\big); \tag{3.11}$$

$$I(\theta_i) = \max(0, L(\theta_i) - Y(\theta_i)), \tag{3.12}$$

where c is the coverage level, which in practice is usually $65\%, 75\%, 85\%$, 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 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:

$$X(\theta_i) = \frac{I(\theta_i)}{L(\theta_i)} = \frac{\max(0, L(\theta_i) - Y(\theta_i))}{L(\theta_i)}. \tag{3.13}$$

While theoretically there are various designs for the reinsurance coverage, the excess of loss (XoL) reinsurance policy is the most common. In addition, XoL reinsurance policy is found to be optimal (Tan et al., 2011) reinsurance design, including for agricultural reinsurance (Porth et al., 2013). For these reasons, this chapter will focus on the XoL reinsurance policy. An XoL with a $A \times B$ 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

$$\pi(\theta_i) = \min(A, \max(0, X(\theta_i) - B)). \tag{3.14}$$

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

$$\widehat{\pi^N(\theta_i)} = E(\pi(\theta_i)) = E(\min(A, \max(0, X(\theta_i) - B))). \tag{3.15}$$

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 chapter, 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 $\widehat{\pi^{SD}(\theta_i)}$):

$$\widehat{\pi^E(\theta_i)} = E(\pi(\theta_i))(1 + \eta_1); \tag{3.16}$$

$$\widehat{\pi^{SD}(\theta_i)} = E(\pi(\theta_i)) + \eta_2 \sqrt{Var(\pi(\theta_i))}. \tag{3.17}$$

where η_1 and η_2 are the respective loading coefficients. Proposition 3.3.2 provides reinsurance pricing formulas for the two premium principles. Not only do the closed form formulas facilitate agricultural reinsurance pricing, it may also assist the reinsurance companies make scientific weather management strategies.

Proposition 3.3.2. *Let $\mu = E\big(\mu_Y(\theta_i)\big)$, $\sigma^2 = Var\big(\mu_Y(\theta_i)\big)$, and further assume that in Equation (3.5), ε, is normally distributed. Let $K_1 = c(1-B), K_2 = c(1-A-B)$, $\Phi(\cdot)$ and $\phi(\cdot)$ be the c.d.f and p.d.f. 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

$$\begin{aligned}
\widehat{\pi^N(\theta_i)} &= \left(\frac{\sigma}{c\mu}\right)[\phi\left(\frac{(K_1-1)\mu}{\sigma}\right)+\left(\frac{(K_1-1)\mu}{\sigma}\right)\Phi\left(\frac{(K_1-1)\mu}{\sigma}\right) \\
&- \phi\left(\frac{(K_2-1)\mu}{\sigma}\right)-\left(\frac{(K_2-1)\mu}{\sigma}\right)\Phi\left(\frac{(K_2-1)\mu}{\sigma}\right)], \qquad (3.18)\\
&:= M_1, \qquad (3.19)\\
\widehat{\pi^E(\theta_i)} &= M_1(1+\eta_1); \qquad (3.20)\\
\widehat{\pi^{SD}(\theta_i)} &= M_1+\eta_2\sqrt{M_2-(M_1)^2}, \qquad (3.21)
\end{aligned}$$

where

$$\begin{aligned}
M_2 := & \left(\frac{\sigma}{c\mu}\right)^2\left\{[\left(\frac{(K_1-1)\mu}{\sigma}\right)^2+1]\Phi\left(\frac{(K_1-1)\mu}{\sigma}\right)+\left(\frac{(K_1-1)\mu}{\sigma}\right)\phi\left(\frac{(K_1-1)\mu}{\sigma}\right)\right. \\
+ & [\left(\frac{(K_2-1)\mu}{\sigma}\right)^2-2\left(\frac{(K_1-1)\mu}{\sigma}\right)\left(\frac{(K_2-1)\mu}{\sigma}\right)-1]\Phi\left(\frac{(K_2-1)\mu}{\sigma}\right) \\
+ & \left.\left(\frac{(K_2-1)\mu}{\sigma}-2\frac{(K_1-1)\mu}{\sigma}\right)\phi\left(\frac{(K_2-1)\mu}{\sigma}\right)\right\}, \qquad (3.22)
\end{aligned}$$

Proof. Details are provided in Appendix 3A.2. □

3.4 Conclusion Remarks

Weather risks are cited to be one of the major risks that are unmanaged and threatening the success of agricultural reinsurance business. Faced with new challenges, such as tightening markets, weather sensitive industries such as reinsurers with specialized agricultural business units must look to improve their weather risk modeling and management platforms. Based on the crop yield forecasting and reinsurance pricing framework proposed in this chapter, we conclude this study by proposing an *Division to Integration* weather risk management procedure for agricultural reinsurers, including an exhaustive risk exposure analysis and a strategic risk management method.

***Division*: Identifying Weather Risks**

The *Division* step helps to identify the critical weather risk variables, as well as the corresponding impact of these variables on incomes, revenues, margins, and profits. The crop yield forecasting framework developed in the previous sections of this chapter can be used to achieve these objectives for reinsurers.

For example, the model selection algorithms proposed in Section 3.2.4, are able to effectively identify important weather indices, while the analytical pricing formulas in Section 3.3 facilitate the sensitivity analysis of the impact of important weather indices. Let us recall the selected important weather indices in Section 3.2.3, Figure 3.4. Some indices indicate extreme events (e.g., Tmin_Oct describes the lowest temperature in October). Some indices are average measures (e.g., PRELavg_May calculates the average PREL during May and provides information about the average precipitation level). Some indices count extreme days (e.g., DGDH1cot_Jun counts the nonzero days during June for DGDH and indicates the number of days that the crops can grow). Identifying these important weather variables is critical for developing sound weather risk management strategies. Also recall that the reinsurance premium is defined as a function of weather indices, $\pi(\mathbf{W}) = \pi(W_1, \ldots, W_p) : \mathbb{R}^p \mapsto \mathbb{R}$, where $\mathbf{W}$ is a p-dimensional weather index. The gradient vector of $\pi(\mathbf{W})$ is expressed as:

$$\nabla\pi(\mathbf{W}) \quad = \quad (\frac{\partial\pi}{\partial W_1}, \ldots, \frac{\partial\pi}{\partial W_p})'.$$

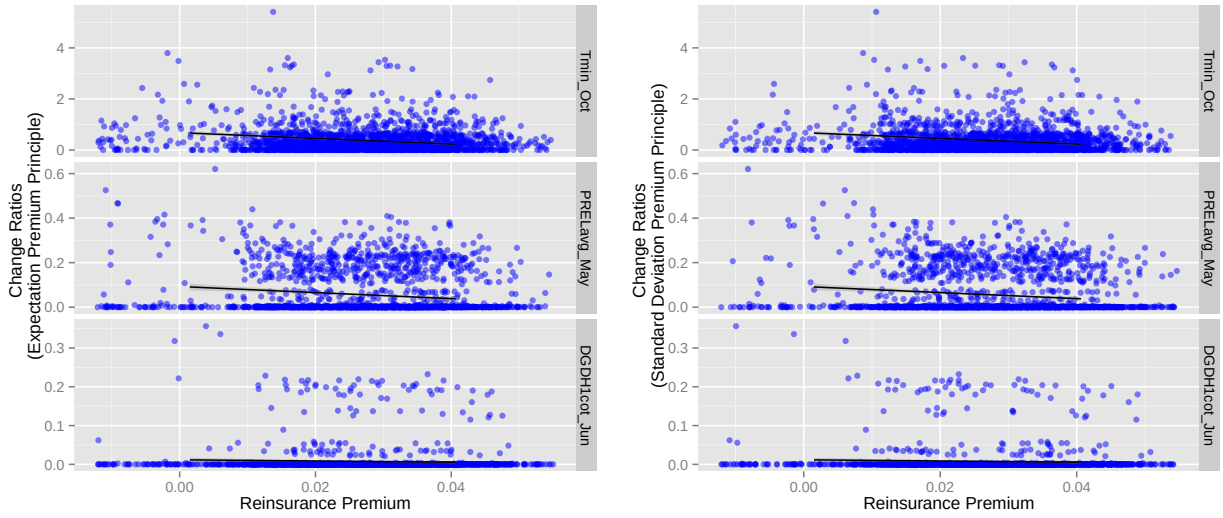
Then it follows from the first order Taylor expansion at an arbitrary point $\mathbf{W}_0$ that

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

Under the above linear approximation, the change in premium, as triggered by the weather indexes, is proportional to the gradient vector $\nabla\pi(\mathbf{W})$. Consequently, the gradient vector is a measure of the sensitivity of premium with respect to weather indices.

***Integration*: Managing Weather Risks**

Through detailed measurement of the "key" weather variables in the *division* step, the reinsurance companies can help control extreme weather risks and proceed to the *integration* step, which aims at risk/return optimization and value creation (Ingram, 2009). The *integration* step seeks to develop a comprehensive platform to sustainably interpret and control all risks on some comparable basis for higher level decision making. With a wealth of research in the *devision* step, the reinsurers have the expertise to provide tailored programs for different weather-sensitive parts of the risk portfolio. For example, Figure 3.6 shows the relationship of the weather risk sensitivities with the reinsurance premiums. Note that this analysis is based on the entire risk portfolio in Manitoba and hence the strategic management approaches will be promoted at a macro level. We observe that lower reinsurance premiums are more sensitive to the weather risks and this observation applies to all three types of weather indices, extreme events, average measures and extreme days. This suggests that the reinsurance companies require adequate capital budgeting to fulfill proposed



(a) Expectation premium principle.

(b) Standard deviation premium principle.

Figure 3.6: Relationships of weather sensitivities to reinsurance premiums based on two premium principles. Results for three weather indices, Tmin_Oct, PRELavg_May and DGDH1cot_Jun, are displayed.

risk management objectives which can be achieved by carefully selecting appropriate risk loadings.

The ultimate objective of the *Division to Integration* procedure is to improve the reinsurance firm value. Taking into consideration all critical weather risk factors of the corporation, the risk managers will make high level strategic balance between risk and return, achieving value creation for the company. Based on the informative and exhaustive "*division to integration*" analysis, agricultural reinsurance companies will construct a comprehensive protection for the total risk portfolio with diversified retention level and coverage limit level for each risk and achieve an optimal and self-interactive risk management framework.

In summary, this chapter provides a new crop yield forecasting model, which is able to provide better in-sample and out-of-sample forecasting results, throughout the integration of weather and geographical correlated regions. A closed form pricing formula for reinsurance policy is provided, based on a newly proposed forecasting estimator. In the empirical analysis, a new crop mix restatement algorithm is shown to restate the farming program over the historical experience to the current level, and dimension reduction and model selection algorithms are proposed to select the best design matrix in crop yield forecasting. Finally a *Division to Integration* procedure is provided for agricultural reinsurers to manage weather risks.

3A Appendix: Proofs

3A.1 Proof of Proposition 3.3.1

Let us start the proof by defining

$$\begin{aligned} QL^N &= E\big(\hat{\mu}_Y^N - \mu_Y(\theta_i)\big)^2 \\ QL^C &= E\big(\hat{\mu}_Y^C - \mu_Y(\theta_i)\big)^2, \end{aligned}$$

then,

$$\begin{aligned} QL^N &= E\big(\mu_Y - \mu_Y(\theta_i) + \lambda(\hat{\mu}_Y^C - \mu_Z)\big)^2 \\ &= E\big(\hat{\mu}_Y^C - \mu_Y(\theta_i) + \lambda(\hat{\mu}_Y^C - \mu_Z) + (\mu_Y - \hat{\mu}_Y^C)\big)^2 \\ &= E\big(\hat{\mu}_Y^C - \mu_Y(\theta_i)\big)^2 + \lambda^2 E\big(\hat{\mu}_Y^C - \mu_Z\big)^2 \\ &\quad + E\big(\mu_Y - \hat{\mu}_Y^C\big)^2 + 2\lambda E\big(\hat{\mu}_Y^C - \mu_Y(\theta_i)\big)\big(\hat{\mu}_Y^C - \mu_Z\big) \\ &\quad + 2\lambda E\big(\hat{\mu}_Y^C - \mu_Z\big)\big(\mu_Y - \hat{\mu}_Y^C\big) + 2E\big(\hat{\mu}_Y^C - \mu_Y(\theta_i)\big)\big(\mu_Y - \hat{\mu}_Y^C\big) \end{aligned}$$

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

$$\begin{aligned} E\big(\hat{\mu}_Y^C - \mu_Z\big)^2 &= Var\big(\hat{\mu}_Y^C\big) \\ E\big(\mu_Y - \hat{\mu}_Y^C\big)^2 &= Var\big(\hat{\mu}_Y^C\big) + (\mu_Y - \mu_Z)^2 \end{aligned}$$

also,

$$\begin{aligned} & 2\lambda\Big(E\big(\hat{\mu}_Y^C - \mu_Y(\theta_i)\big)(\hat{\mu}_Y^C - \mu_Z) + E(\hat{\mu}_Y^C - \mu_Z)(\mu_Y - \hat{\mu}_Y^C)\Big) \\ = \; & 2\lambda E\big(\hat{\mu}_C^Y - \mu_Z\big)\big(\mu_Y - \mu_Y(\theta_i)\big) \\ = \; & 2\lambda E\Big(\big(\mu_Y - \mu_Y(\theta_i)\big)E\big(\hat{\mu}_Y^C - \mu_Z|\theta_i\big)\Big) \\ = \; & 0, \end{aligned}$$

and,

$$\begin{aligned} & E\big(\hat{\mu}_Y^C - \mu_Y(\theta_i)\big)\big(\mu_Y - \hat{\mu}_Y^C\big) \\ = \; & E\big(\mu_Y(\theta_i) - \mu_Y\big)\big(\hat{\mu}_Y^C - \mu_Y\big) + E\big(\mu_Y - \hat{\mu}_Y^C\big)\big(\hat{\mu}_Y^C - \mu_Y\big) \\ = \; & -E\big(\hat{\mu}_Y^C - \mu_Y\big)^2 \\ = \; & -Var\big(\hat{\mu}_Y^C\big) - (\mu_Y - \mu_Z)^2. \end{aligned}$$

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

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

Namely, if $\lambda = \rho_{YZ}\dfrac{\sigma_Y}{\sigma_Z} \in [0,1]$, we have $QL^N < QL^C$, and relation 3.10 holds.

3A.2 Proof of Proposition 3.3.2

Since

$$\begin{aligned}\pi(\theta_i) &= \min(A, \max(0, X(\theta_1) - B));\\ &= \max(X(\theta_i) - B, 0) - \max(X(\theta_i) - (A+B), 0),\end{aligned}$$

where $X(\theta_i)$ is defined as Equation 3.13. Therefore,

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

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

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

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

Therefore,

$$\begin{aligned}
E(\pi(\theta_i)) &= (1-B)E\left(\mathbb{1}_{[\frac{Y(\theta_i)}{L(\theta_i)}<1-B]}\right) - E\left(\frac{Y(\theta_i)}{L(\theta_i)}\mathbb{1}_{[\frac{Y(\theta_i)}{L(\theta_i)}<1-B]}\right) \\
&- (1-A-B)E\left(\mathbb{1}_{[\frac{Y(\theta_i)}{L(\theta_i)}<1-A-B]}\right) - E\left(\frac{Y(\theta_i)}{L(\theta_i)}\mathbb{1}_{[\frac{Y(\theta_i)}{L(\theta_i)}<1-A-B]}\right); \\
&= \left\{\frac{\sigma}{c\mu}\phi\left(\frac{(K_1-1)\mu}{\sigma}\right) - \frac{1-K_1}{c}\Phi\left(\frac{(K_1-1)\mu}{\sigma}\right)\right\} \\
&- \left\{\frac{\sigma}{c\mu}\phi\left(\frac{(K_2-1)\mu}{\sigma}\right) - \frac{1-K_2}{c}\Phi\left(\frac{(K_2-1)\mu}{\sigma}\right)\right\},
\end{aligned}$$

which is exactly equation 3.18. Meanwhile,

$$\begin{aligned}
E[(\pi(\theta_i))^2] &= E\left((1-B-\frac{Y(\theta_i)}{L(\theta_i)})^2\mathbb{1}_{[\frac{Y(\theta_i)}{L(\theta_i)}<1-B]}\right) + E\left((1-A-B-\frac{Y(\theta_i)}{L(\theta_i)})^2\mathbb{1}_{[\frac{Y(\theta_i)}{L(\theta_i)}<1-A-B]}\right) \\
&- 2E\left((1-B-\frac{Y(\theta_i)}{L(\theta_i)})(1-A-B-\frac{Y(\theta_i)}{L(\theta_i)})\mathbb{1}_{[\frac{Y(\theta_i)}{L(\theta_i)}<1-A-B]}\right), \\
&= \left(\frac{\sigma}{c\mu}\right)^2\{[\left(\frac{(K_1-1)\mu}{\sigma}\right)^2+1]\Phi\left(\frac{(K_1-1)\mu}{\sigma}\right) + \left(\frac{(K_1-1)\mu}{\sigma}\right)\phi\left(\frac{(K_1-1)\mu}{\sigma}\right) \\
&+ [\left(\frac{(K_2-1)\mu}{\sigma}\right)^2 - 2\left(\frac{(K_1-1)\mu}{\sigma}\right)\left(\frac{(K_2-1)\mu}{\sigma}\right) - 1]\Phi\left(\frac{(K_2-1)\mu}{\sigma}\right) \\
&+ \left(\frac{(K_2-1)\mu}{\sigma} - 2\frac{(K_1-1)\mu}{\sigma}\right)\phi\left(\frac{(K_2-1)\mu}{\sigma}\right)\},
\end{aligned}$$

Denote:

$$\begin{aligned}
M_1 &= \left(\frac{\sigma}{c\mu}\right)[\phi\left(\frac{(K_1-1)\mu}{\sigma}\right) + \left(\frac{(K_1-1)\mu}{\sigma}\right)\Phi\left(\frac{(K_1-1)\mu}{\sigma}\right) \\
&- \phi\left(\frac{(K_2-1)\mu}{\sigma}\right) - \left(\frac{(K_2-1)\mu}{\sigma}\right)\Phi\left(\frac{(K_2-1)\mu}{\sigma}\right)], \\
M_2 &= \left(\frac{\sigma}{c\mu}\right)^2\{[\left(\frac{(K_1-1)\mu}{\sigma}\right)^2+1]\Phi\left(\frac{(K_1-1)\mu}{\sigma}\right) + \left(\frac{(K_1-1)\mu}{\sigma}\right)\phi\left(\frac{(K_1-1)\mu}{\sigma}\right) \\
&+ [\left(\frac{(K_2-1)\mu}{\sigma}\right)^2 - 2\left(\frac{(K_1-1)\mu}{\sigma}\right)\left(\frac{(K_2-1)\mu}{\sigma}\right) - 1]\Phi\left(\frac{(K_2-1)\mu}{\sigma}\right) \\
&+ \left(\frac{(K_2-1)\mu}{\sigma} - 2\frac{(K_1-1)\mu}{\sigma}\right)\phi\left(\frac{(K_2-1)\mu}{\sigma}\right)\},
\end{aligned}$$

then, $Var(\pi(\theta_i)) = M_2 - (M_1)^2$, and equation 3.21 holds obviously.

Chapter 4

A Copula-based Model for Spatial Dependence & Aggregation in Weather Risk Hedging

4.1 Introduction

Weather risk, described as the operational and financial variabilities caused by adverse meteorological conditions, is a major environmental issue and a key economic factor. Possible climate change also brings concerns of more frequent and severe extreme natural hazards over larger areas and affecting more people (Hellmuth et al., 2009; IPCC, 2007). The agriculture sector is one of the most exposed industries to weather related risks, with some estimates stating that adverse weather may be responsible for at least 70% of agricultural loss, including crop and livestock production (USDA, 2014). A major challenge facing the agricultural sector is that weather risk is systematic and undiversifiable in the sense that it is outside the control of human management, and at times weather risk can be widespread and spatially correlated, impacting many farms within a region (Porth et al., 2014a). Therefore, weather risk will not be eliminated by pooling (Doherty and Dionne, 1993), and must be managed through various risk transfer techniques. Agricultural insurance schemes have played an important role in helping to stabilize a producer's income by minimizing the economic effects caused by adverse weather events.

The main objective of this chapter is to develop and compare different weather risk hedging strategies for agricultural insurers and reinsurers. This topic is of great importance since hedging weather risk effectively is critical for the long-term sustainability of the agricultural

sector. The Property & Casualty (P & C) insurance sector is highly focused on managing catastrophic losses due to disasters compared to those insurance dealing with life coverages (Dong et al., 1996; Kleindorfer et al., 2012; Priest, 1996). Further, agricultural insurers and reinsurers bear higher loss ratios compared to other lines of business in the P&C sector (Woodard and Garcia, 2008b). Moreover, agricultural insurers may face additional exposure to weather risk due to the increase of climate variability and uncertainty. As such, managing weather risks with financial instruments such as weather derivatives has emerged, and over the past decades has shown more success. While most of the weather derivative (WD) market transactions are tailor-made in the Over the Counter (OTC) market, the organized markets are becoming more successful and many types of WDs are traded at the Chicago Mercantile Exchange (CME).

To help manage the insurers exposure to losses, reinsurance is often an important component of the risk management strategy. A study from Qatar Re shows that almost 80% of the global downside risk for agricultural insurers are reinsured (Schneider and Roth, 2013). In Canada, provincial crop insurance companies can choose to participate in a unique Federal-Provincial Reinsurance Fund and/or purchase reinsurance from the private market. Similarly, in the US the Standard Reinsurance Agreement (SRA) and private reinsurance provide significant risk transfer, helping insurers manage extreme events (i.e., low frequency-high severity). In general, the various large international agricultural reinsurance companies are high aggregators of risks, and are therefore particularly exposed to catastrophic events. Kunreuther et al. (1993) study how uncertainty affects the decisions of actuaries, insurers and reinsurers and suggest improving risk assessment and creating new risk-sharing arrangements to address the issues related to uncertainty involving natural and technological hazard. As such, agricultural reinsurers also require advanced methods to manage the systemic part of their risk exposures (Turvey et al., 1988).

In some cases hedging weather risks with financial instruments may be advantageous over traditional reinsurance, in terms of potential reduced cost, and improved market efficiency. For example, financial instruments do not require loss checking and adjusting, thereby saving administration costs. Further, financial weather instruments may reduce information asymmetry, including adverse selection and moral hazard, which was previously mentioned as a classical crop (re)insurance challenge (Goodwin, 2001; Quiggen et al., 1994). This is because indemnities of financial weather instruments are triggered based on a specific weather event rather than actual farm loss, which is a more transparent approach that is not subject to manipulation, etc. Furthermore, from a statistical inference viewpoint the modeling and pricing of financial weather instruments may be more appealing compared to reinsurance pricing, since large volumes of reliable and extensive weather data records are typically available in

daily frequency. In contrast, reinsurance for agriculture is usually faced with the challenge of shortness of data, where the loss experience can be very short and at times may suffer from missing data (Porth et al., 2014b). Therefore, weather hedging via financial engineering tools may be an ideal complement to agricultural reinsurance.

The second objective of this chapter is to refine the statistical weather variable modeling. This is an essential, yet, challenging task for financial weather instrument pricing and hedging, owing to the nonstationarity, seasonality and multidimensionality of the weather data (Dischel and Barrieu, 2002), as well as the incomplete nature of the market (Alexandridis and Zapranis, 2013). Unfortunately, existing stochastic weather models are typically designed for modeling only a single region (with the exception of the work by Okhrin et al. (2013a)). However, failure to consider the dependence structure for weather variable modeling and weather derivative pricing may lead to substantial basis risk in the resulting hedging strategy if the spatial correlations are not taken into account. Therefore, in this chapter, we use a wavelet technique that allows detailed analysis of the nonstationarity and seasonality of the data, together with a non-Gaussian general hyperbolic (GH) distribution family to capture the heavy tail property of the data. We model the dependence structure of the weather data with the copula approach.

The construction and estimation of high dimensional copulas are challenging problems, yet these are critical and essential for risk management (Kole et al., 2007; Patton, 2009). Therefore, investigating the theoretical properties and empirical applications of high dimensional modelling with copulas have attracted much attentions in the literature. Elliptical copula models are in adequate to capture the nonlinear dependence in the financial returns (Embrechts et al., 2002). In addition, the number of parameters in the elliptical copula grows quadratically with dimension. Vine copula also known as pair copula construction (PCC), facilitate extensions from bivariate copulas to higher dimensions through conditioning using a handy graphical tool for labelling high-dimensional dependence structures[1]. Although extremely flexible, there are still some outstanding issues still need to be adequately addressed for vine copula models. These include testing the simplifying assumptions and overcoming the potential problems due to these assumptions, selecting appropriate bivariate models from the huge number of potential candidates, designing spatial vines and goodness-of-fit tests for high dimensional vine copulas, etc.

Archimedean copulas (AC), though have a very small number of parameters irrespective of dimensions, suffer from the exchangeable structures, which makes AC inadequate to model

[1]A detailed introduction to vine copulas can be found in Aas et al. (2009), and the estimation of vine copulas is introduced in Kurowicka and Cooke (2006) for Gaussian vines and Aas et al. (2009) for non-Gaussian vines.

complex dependence structures (Weiß and Scheffer, 2015). In an attempt to overcome the exchangeability issue of the Archimedean copula, the hierarchical Archimedean copula (HAC) has been proposed (Joe, 1997). This approach partially overcomes the exchangeability by "nesting" two or more Archimedean copulas with appropriate groupings. Therefore, HACs provide a more flexible framework by allowing different distribution properties between each subgroup with a relatively small set of parameters. Despite their advantages, there are compatible conditions which the generators need to be satisfied to ensure that the resulting HAC yields a valid multivariate distribution. These conditions, however, can be difficult to verify and hence also restrict its practical applications.

In this chapter, we advocate Lévy subordinated hierarchical Archimedean copulas (LSHAC) for the modelling of the geographical dependence of weather risks. Hering et al. (2010) introduce the construction and simulation of LSHACs, while Mai and Scherer (2012) discuss LSHAC within a h-extendibly copula framework. LSHAC model is general enough to comprise all HACs whose generators are compatible (Hering et al., 2010). In other words, by inducing dependence within each group with Lévy subordinators, the hard-to-check *compatible conditions* are conveniently overcome, leading to more flexible and tractable parametric models that have huge application potential to benefit empirical modelling.

Despite the advantages, LSHAC models have never been employed in empirical application, mostly because of the difficulties in on determining the hierarchy structure as well as estimating the parameters for LSHACs. The recursive multi-stage maximum likelihood (ML) estimation procedure proposed by Okhrin et al. (2013b) is efficient for HACs with the same generator functions such as Gumbel generator or Clayton generator (hereafter, we call these models All-GM-HACs or All-CL-HACs in short), but will be computationally demanding for general HAC models with different generators. Moreover, the technique by Okhrin et al. (2013b) provides a sub-optimal structure as well as ML estimators because of its recursive nature, and also LSHAC models are constructed in such a way that the parameters in the outer layer in the hierarchy should not be estimated later than the inner layer parameters, meaning the bottom-up recursive procedure by Okhrin et al. (2013b) is not applicable for LSHACs. Motivated by these observations, this chapter attempts to fill up these gaps by providing a comprehensive study of some of the outstanding issues in the construction and estimation of LSHACs. In doing so, we explicitly construct a multi-level LSHAC in a fully general setting by developing a notation system, an integral representation (Hofert, 2008; Joe, 1997; Marshall and Olkin, 1988; McNeil, 2008; Whelan, 2004) and the corresponding sampling algorithm. In addition, we propose to exploit the hierarchical clustering analysis to efficiently determine the grouping structure of LSHACs with three dissimilarity metrics, Euclidean, Kendall's τ and τ-Euclidean. We also use a simulation study to indicate that

τ-Euclidean metric provides the best grouping reliability on correctly identifying the true structure. Once the optimal structure of LSHAC is identified, an augmented inference for margin (AIFM) method is used to estimate the remaining LSHAC parameters.

Empirical results show that the proposed LSHAC model has better estimation performance compared to the classical Gaussian copula and traditional hierarchical Archimedean copulas (HAC). To the best of our knowledge, this is the first time that the LSHAC is employed for modeling the geographical dependence of weather events. Finally, we propose a pricing framework based on the conditional Esscher transform method (Bühlmann et al., 1996; Gerber and Shiu, 1994) to address the challenge of instrument pricing in an incomplete market. The empirical analysis of this paper uses temperature data from eight provinces in Canada. The focus is on temperature, rather than precipitation because previous studies argue that temperature has a higher correlation with crop production compared to precipitation, and is better suited for crop insurance hedging (Lobell and Burke, 2008; Woodard and Garcia, 2008a). Using the refined statistical modeling of the weather data proposed in this paper, four hedging strategies are developed and compared. In assessing the effectiveness of the various hedging strategies, we are interested in the following three problems: (1) the necessity of hedging weather risk; (2) the importance of the assumed underlying dependence structure; (3) the geographical aggregation effect on hedging effectiveness. The results indicate that hedging weather risk is an important risk management approach to stabilize cash flows and reduce losses. The importance of capturing the appropriate dependence structure of weather risk is also highlighted, and the LSHAC is shown to improve the hedging performance. Moreover, the results reveal significant geographical aggregation benefits in weather risk hedging, which means that more effective hedging may be achieved as the spatial aggregation level increases.

Main Contributions in This Chapter

This chapter contributes the literature from the following perspectives. First, we we explicitly constructs a multi-level LSHAC in a fully general setting by developing a notation system, an integral representation and the corresponding sampling algorithm (Section 4.2.2). In addition, we propose to exploit the hierarchical clustering analysis to efficiently determine the grouping structure of LSHACs with a new proposed dissimilarity metrics, τ-Euclidean metric (Equation (4.17)). We compare the efficiency of this new metric with Euclidean and Kendall's τ metric. We also use a simulation study to indicate that τ-Euclidean metric provides the best grouping reliability on correctly identifying the true structure (Section 4.3). Moreover, an augmented inference for margin (AIFM) estimating procedure (Section 4.2.3) is proposed to estimate the remaining LSHAC parameters. This chapter refines the statistical modeling

of Canadian temperature processes with wavelet analysis and LSHAC model. Finally, this chapter proposes and compares different weather risk hedging strategies for agricultural insurers and reinsurers.

The remainder of this paper proceeds as follows. Section 4.2 introduces the methodology of this paper. In particular, a general framework of LSHAC together with some theoretical results are developed. After presenting the theoretical model, a three-stage estimation procedure by an AIFM method for LSHACs is proposed. In particular, a new dissimilarity metric based on the hierarchical clustering analysis is used to determine the optimal structure of a LSHAC. Section 4.3 provides a simulation study to investigate the efficiency of the proposed grouping method. In the empirical study in Section 4.4, using Canadian daily temperature data. Section 4.5 discusses different hedging strategies for the insurers to hedge the weather risk. The paper concludes with future research directions in Section 4.6. Appendix of the paper collects the proofs.

4.2 Methodology

4.2.1 Hierarchical Archimedean Copulas (HACs)

A function $C : [0,1]^d \to [0,1], C(u_1, u_2, \ldots, u_d) = \psi\big(\psi^{-1}(u_1)+, ..., \psi^{-1}(u_d)\big)$ defines a d-dimensional Archimedean copula (AC) if $\psi \in \mathcal{G} = \{\psi : [0,\infty) \to [0,1] \,|\, \psi_{\lim u\to\infty}(u) = 0, \psi(0) = 1, (-1)^k \frac{d^k}{du^k}\psi(u) \geq 0, k \in \mathbb{N}\}$ (Kimberling, 1974; Nelsen, 2006). Functions in the class of $\mathcal{G}$ is known as *completely monotonic* (c.m.). ψ is called the generator of the corresponding Archimedean copula and ψ^{-1} is its general inverse, defined by $\psi^{-1}(u) = \inf\{t : \psi(t) \leq u\}$. According to the Bernstein's Theorem (Feller, 2008), the class of c.m. functions coincides with the class of Laplace-Stieltjes transforms on $[0,\infty)$. Hence, copulas defined by the c.m. generators are also known as the Laplace-Stieltjes transform AC (LT-AC).

The advantage of the AC family is that it simplifies the modelling of dependence in high dimension with only one parameter. The drawback of such simplification is that the resulting distribution leads to the exchangeability phenomenon; i.e. the distribution of random variables $(u_1, u_2, ..., u_d)$ is invariant under permutation. To address this problem, the Hierarchical Archimedean Copula (HAC) models have been proposed by nesting the random variables into a hierarchy. HAC was first introduced by Joe (1997), and discussed within a more general framework by Savu and Trede (2010). In that paper, the authors derived recursive formulas for general HACs and provided simulation techniques. Sampling algorithms

were also discussed by Whelan (2004), McNeil (2008) and Hofert (2012).

The HAC model is best illustrated with an example. Assuming that a six-dimensional HAC is given by

$$C(u_1,\ldots,u_6) \;=\; C_{\psi_0}(\, C_{\psi_{1,1}}(\, C_{\psi_{2,1}}(u_1,u_2),u_3\,),C_{\psi_{1,2}}(\, C_{\psi_{2,2}}(u_4,u_5),u_6\,)\,). \tag{4.1}$$

Note that (4.1) is a three-level HAC with five generators. The copula C_{ψ_0} with generator ψ_0 is known as the outer copula while copulas $C_{\psi_{1,1}}$ and $C_{\psi_{1,2}}$ ($C_{\psi_{2,1}}$ and $C_{\psi_{2,2}}$), with generators $\psi_{1,1}$ and $\psi_{1,2}$ ($\psi_{2,1}$ and $\psi_{2,2}$), are the inner copulas at level 1 (level 2), respectively. Thus, $\{u_1,u_2\}$ ($\{u_4,u_5\}$) are first nested by $C_{\psi_{2,1}}$ ($C_{\psi_{2,2}}$), grouped together with u_3 (u_6) under $C_{\psi_{1,1}}$ ($C_{\psi_{1,2}}$), and hence C_{ψ_0}. Besides that ψ_0 and $\psi_{i,j}\,(i,j=1,2)$ should be c.m., to ensure (4.1) is a valid copula, the conditions ψ_0 and $\psi_{i,j} \in \mathcal{G}$ and $(\psi_0^{-1}\circ\psi_{i,j})'$ and $(\psi_{1,k}^{-1}\circ\psi_{2,k})' \in \mathcal{G}\,(i,j,k=1,2)$, called *compatible conditions*, need to be satisfied. Note that the notation "$\circ$" denotes function composition. The compatible conditions cause the construction of HACs more challenging. If all the generators of a HAC are from the same AC family (e.g. Gumbel family), these conditions are not too difficult to verify, since in most cases the copula parameters should be monotonic from top to deeper levels (Embrechts et al., 2003; Okhrin et al., 2013b). However, if HACs are constructed from mixed generators involving different families, one has to verify the compatible conditions on a case-by-case basis (Hofert, 2012; Savu and Trede, 2010). For this reason most empirical studies on HAC models have focused on either All-GM-HACs or All-CL-HACs (Chorośś-Tomczyk et al., 2013; Okhrin et al., 2013a;b; Savu and Trede, 2010). Hering et al. (2010) circumvented this hard-to-check *compatible conditions* by constructing two-level HACs via Lévy subordinators (LSHAC) and provided a stochastic representation using a probability construction. Relying on the fact that Lévy subordinators are stable under (independent) subordination, Mai and Scherer (2012) considered an h-extendible framework in which LSHAC is one of the special cases. They provided a stochastic representation of three-level LSHAC models and explained that that the stochastic representation can be extended to higher levels in an iterative way.

4.2.2 General Framework of the LSHAC

Hering et al. (2010) delicately constructed c.m. generators for HAC with Lévy subordinators. In this subsection, we extend the model in Hering et al. (2010) by introducing a multi-level LSHAC in a fully general setting with a comprehensive notation system, stochastic representation and sampling algorithm. Specifically, let $\{S_t : 0 \le t \le T\}$ be a Lévy subordinator, i.e., a stochastically continuous non-decreasing Lévy process, which has zero start, stationary

and independent increments (See Tankov, 2004; Proposition 3.10). The Laplace transform of S_t satisfies the following equation:

$$E(e^{-\omega S_t}) = \exp\left(-t\Psi(\omega)\right), \qquad \forall \omega > 0, \tag{4.2}$$

where the non-decreasing function $\Psi : [0, \infty) \to [0, \infty)$ is called the *Laplace exponent* of the Lévy subordinator. As shown in Theorem 2.1 in Hering et al. (2010), given a c.m. generator ψ_0, the generator defined by $\psi_0 \circ \Psi$ is a c.m. generator and satisfies the *compatible conditions*. Mai and Scherer (2012) provided a three-level LSHAC and demonstrated that this construction via Lévy subordinators could be iterated to higher-level LSHACs. Indeed, it is theoretically demanding to construct a LSHAC in a fully general setting. To this end, we provide a notation system and an integral representation (Hofert, 2008; Joe, 1997; Marshall and Olkin, 1988; McNeil, 2008; Whelan, 2004) of a general multi-level LSHAC.

We now describe a general L-level LSHAC exhibited in Figure 4.1 by introducing the following notation:

- For $l = 0, 1, \ldots, L$, let l denote the index level of LSHAC and J_l denote the number of copulas at level l.
- At level 0:
  - There is only one copula, denoted by $C_{0,1}^{(0)}$, and hence by construction $J_0 = 1$. This is also known as the outer copula.
  - There is a random time variable $V_{0,1}^{(0)}$ at which the Lévy subordinators for all subsequent groups are evaluated. We denote its corresponding cumulative distribution function (c.d.f.) as $G_{0,1}^{(0)}(v)$ and its LT-AC generator as $\psi_{0,1}^{(0)}$.
- At level l:
  - For $l = 1, \ldots, L$ and $j = 1, \ldots, J_{l-1}$, let $D_j^{(l)}$ be the number of copulas at level l "emanating" from the j-th copula in the previous level $l - 1$. Note that the following condition must hold:

$$\sum_{j=1}^{J_{l-1}} D_j^{(l)} \quad = \quad J_l, \qquad l = 1, \ldots, L, \tag{4.3}$$

    and $J_L = d$.
  - Let $C_{j,k}^{(l)}$ with generator $\psi_{j,k}^{(l)}$ be the k-th copula in the j-th cluster with size $D_j^{(l)}$. It is emanated from the j-th copula at level $l-1$, for $l = 1, \ldots, L, j = 1, \ldots, J_{l-1}$,

and $k = 1, \ldots, D_j^{(l)}$;

- The m-th adjacent copula emanated from $C_{j,k}^{(l)}$ at level $l+1$ is denoted as $C_{s,m}^{(l+1)}$, where $s = 1, \ldots, J_l$, is the position of $C_{s,m}^{(l+1)}$, satisfying

$$s = \left(\sum_{q=1}^{j-1} D_q^{(l)} \right) \mathbb{I}_{\{j>1\}} + k, \tag{4.4}$$

where $\mathbb{I}_{\{.\}}$ is the indicator function.

- At level L :
  - We partition $(u_1, \ldots, u_d)$ into J_{L-1} groups and define

$$C_{j,k}^{(L)} = u_s, s = 1, \ldots, J_L, j = 1, \ldots, J_{L-1}, k = 1, \ldots, D_j^{(L)}, \tag{4.5}$$

where s satisfies (4.4).

- Additional definitions:
  - Let $X(t)$ denote a Lévy subordinator evaluated at time t, with corresponding c.d.f $\widetilde{G}(x;t)$ and Laplace exponent $\widetilde{\Psi}$.
  - Define function $F_{s_{l-1},j_l}^{(l)}(u)$ as

$$F_{s_{l-1},j_l}^{(l)}(u) = \exp(-\psi_{s_{l-1},j_l}^{(l)-1}(u)). \tag{4.6}$$

Given the above definition of $F_{s_{l-1},j_l}^{(l)}$, the following function

$$\left(F_{s_{l-1},j_l}^{(l)}(u)\right)^v \;=\; \exp\big(-v\psi_{s_{l-1},j_l}^{(l)-1}(u)\big) \tag{4.7}$$

is a valid c.d.f. for any positive v (Marshall and Olkin, 1988). Let $\Psi_{s_{l-1},j_l}^{(l)} = \psi_{s_{l-2},j_{l-1}}^{(l-1)-1} \circ \psi_{s_{l-1},j_l}^{(l)}$ be the Laplace exponent of a Lévy subordinator, $X_{s_{l-1},j_l}^{(l)}$, with c.d.f. $G_{s_{l-1},j_l}^{(l)}$, then the generator given by

$$\begin{aligned} \tilde{\psi}_{s_{l-1},j_l}^{(l)}(u;v) &= \left(F_{s_{l-2},j_{l-1}}^{(l-1)}\big(\psi_{s_{l-1},j_l}^{(l)}(u)\big)\right)^v = \exp\big(-v\psi_{s_{l-2},j_{l-1}}^{(l-1)-1} \circ \psi_{s_{l-1},j_l}^{(l)}(u)\big) \\ &= \exp\big(-v\Psi_{s_{l-1},j_l}^{(l)}(u)\big), \end{aligned} \tag{4.8}$$

is also a c.m. LT-AC generator, where $l = 2, \ldots, L-1$ (Feller, 2008; Nelsen, 2006).

Give the above notation and definitions, Theorem 4.2.1 provides the integral representation

of the multi-level LSHAC depicted in Figure 4.1 in terms of Laplace transform.

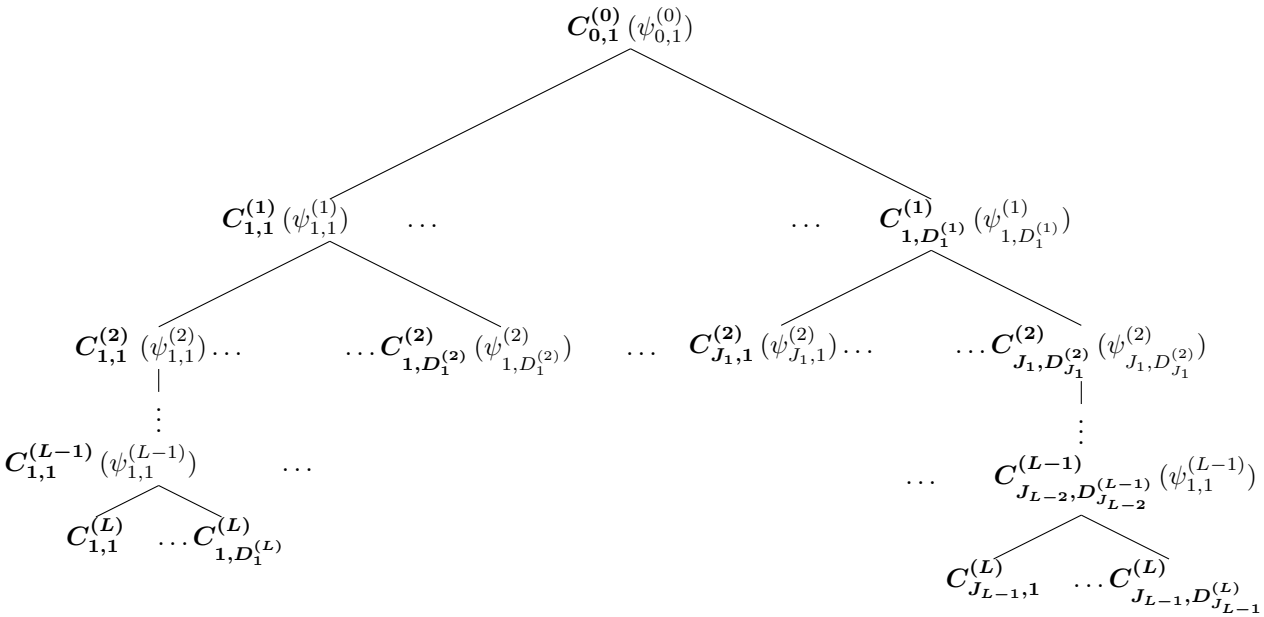


Figure 4.1: General Framework of a LSHAC Model

Theorem 4.2.1. *Given the structure of a LSHAC in Figure 4.1, the copula function,* $C\big(u_1,\ldots,u_d\big)$*, can be constructed as*

$$\int_0^\infty \prod_{j_1=1}^{D_{J_0}^{(1)}} \int_0^\infty \prod_{j_2=1}^{D_{s_1}^{(2)}} \int_0^\infty \cdots \int_0^\infty \prod_{j_{L-1}=1}^{D_{s_{L-2}}^{(L-1)}} \int_0^\infty \prod_{j_L=1}^{D_{s_{L-1}}^{(L)}} \left(F^{(L-1)}_{s_{L-2},j_{L-1}}(C^{(L)}_{s_{L-1},j_L})\right)^{v^{(L-1)}_{s_{L-2},j_{L-1}}} \big(dG\big)^{(L-1)}_{j_{L-1}}, \tag{4.9}$$

where

$$\big(dG\big)^{(0)}_{j_0} = dG^{(0)}_{0,1}(v^{(0)}_{0,1}),$$

and

$$\big(dG\big)^{(l)}_{j_l} = d\widetilde{G}^{(l)}_{s_{l-1},j_l}(v^{(l)}_{s_{l-1},j_l}; v^{(l-1)}_{s_{l-2},j_{l-1}}) \ldots d\widetilde{G}^{(l)}_{s_0,j_1}(v^{(1)}_{s_0,j_1}; v^{(0)}_{0,1}) dG^{(0)}_{0,1}(v^{(0)}_{0,1}).$$

Proof. Proof of Theorem 4.2.1 is provided in 4A.1 □

It follows from above theorem that the following corollary provides an expression of inner generators.

Corollary 4.2.1. *At level* l*, where* $1 \le l \le L$*, the* j_l*-th copula generator in position* s_{l-1}*:*

$\psi^{(l)}_{s_{l-1},j_l}$, *can be expressed as*

$$\psi^{(l)}_{s_{l-1},j_l}(u) \quad = \quad \psi^{(0)}_{0,1} \bigodot_{i=1}^{l} \widetilde{\Psi}^{(i)}_{s_{i-1},j_i}(u), \tag{4.10}$$

where $\bigodot_{k=1}^{n} f_k := f_1 \circ \ldots \circ f_n$, *and* $\psi^{(l)}_{s_{l-1},j_l}$ *is c.m..*

Corollary 4.2.1 states that at each level of a LSHAC, the generator can be constructed from composing an outer AC generator and a sequence of Laplace exponents of Lévy subordinators. Tables 4.1 and 4.2 list, respectively, examples of c.m. Archimedean generators and Lévy subordinators.[2] These AC generators and Lévy subordinators are used in the simulation study in Section 4.3 and the empirical analysis in Section 4.4. In addition, the tail dependences of each Archimedean copula are listed in Table 4.1. In contrast to Gaussian copula with no tail dependence, the LSHAC models can provide both upper tail dependence (e.g., GM family) and lower tail dependence (e.g., CL family). For weather risk, where the extreme events usually happen asymmetrically, LSHAC may have the potential advantage to achieve more flexibility in modelling the tail dependence of the data.

Table 4.1: Archimedean Copula (AC) generators. CL: Clayton family, GM: Gumbel family.

Family	$\boldsymbol{\psi(u)}$	$\boldsymbol{C(u_1,\ldots,u_d)}$	$\boldsymbol{\lambda_l}$	$\boldsymbol{\lambda_u}$	**Parameter**
GM	$\psi_{GM}(u) = \exp\big(-x^{\frac{1}{\theta}}\big)$	$\exp\big(-(\sum_{i-1}^{d}(-\log u_i))^{\frac{1}{\theta}}\big)$	0	$2-2^{\frac{1}{\theta}}$	$\theta \geq 1$
CL	$\psi_{CL}(u) = (1+u)^{-\frac{1}{\theta}}$	$\big(1+\sum_{i=1}^{d}(u_i^{-\theta}-1)\big)^{-\frac{1}{\theta}}$	$2^{-\frac{1}{\theta}}$	0	$\theta > 0$

Table 4.2: Lévy Subordinators. G: Gamma process, GM: Stable process, IG: the Inverse Gaussian process.

Subordinator	$\boldsymbol{\Psi(u)}$	**Parameters**
G	$\Psi_G = a\log\big(1+\frac{u}{b}\big)$	$a>0, b>0$
GM	$\Psi_{GM} = u^a$	$0<a<1$
IG	$\Psi_{IG} = a\sqrt{2u+b^2} - ab$	$a>0, b>0$

It follows immediately from Corollary 4.2.1 that the All-GM-HAC model, which is the most commonly used HAC in the empirical analysis, is a special case of LSHAC. This property is expressed in Corollary 4.2.2 below, and it is also mentioned with a three level HAC example in Mai and Scherer (2012).

[2]In the expressions of Laplace exponents, a and b are parameters of the corresponding Lévy measures (see Tankov (2004) for more information). In addition, Stable process is denoted as GM because it is the distribution of the Gumbel family.

Corollary 4.2.2. *For an All-GM-HAC, the l-th level copula generator $\psi^{(l)}(u)$ can be expressed as ($l \geq 1$):*

$$\psi^{(l)}(u) = \psi^{(0)} \bigodot_{k=1}^{l-1} \widetilde{\Psi}^{(k)}(u) = \exp\big(- u^{\prod_{k=1}^{l-1} \frac{1}{\theta_k}} \big).$$

From the parameterization in Table 4.1 and Table 4.2, $\psi^{(0)}$ represents a GM generator with $\theta = \theta_0, \theta_0 \geq 1$, and $\widetilde{\Psi}^{(k)}$ denotes the k-th GM subordinator with $a = 1/\theta_k, \theta_k \geq 1$.

Random samples from a LSHAC can be simulated relatively easily by recognizing that $\big(F^{(L-1)}_{s_{L-2},j_{L-1}}(C^{(L)}_{s_{L-1},j_L})\big)^{v^{(L-1)}_{s_{L-2},j_{L-1}}}$, where $j_L = 1, \ldots, D^{(L)}_{s_{L-1}}$, is a valid c.d.f. for any positive $v^{(L-1)}_{s_{L-2},j_{L-1}}$ (see (4.7), (4.8) and Theorem 4.2.1). More specifically, if Y_{s_{L-1},j_L} is a uniform random variable on $(0,1)$, then given $V^{(L-1)}_{s_{L-2},j_{L-1}}$ with c.d.f. $\widetilde{G}^{(L-1)}_{s_{L-2},j_{L-1}}(x;t)$, a random sample of $C^{(L)}_{s_{L-1},j_L}$ can be obtained via inverse transform as

$$C^{(L)}_{s_{L-1},j_L} \;=\; \psi^{(L-1)}_{s_{L-2},j_{L-1}}\bigg(- \frac{\log(Y_{s_{L-1},j_L})}{V^{(L-1)}_{s_{L-2},j_{L-1}}} \bigg). \tag{4.11}$$

In summary, for a multi-level LSHAC with a general structure displayed in Figure 4.1, the random samples can be simulated by a sequential procedure formally described in Algorithm 4.2.1.

Algorithm 4.2.1 (**Sampling an L-level LSHAC**).

Step 1: *Generate a random variable $V^{(0)}_{0,1}$ with c.d.f. $G^{(0)}_{0,1}(x)$.*

Step 2: *For $l = 1, \ldots, L-1$, $s_{l-1} = 1, \ldots, J_{l-1}$, $j_l = 1, \ldots, D^{(l)}_{s_{l-1}}$, generate a random variable $V^{(l)}_{s_{l-1},j_l}$ with c.d.f. $\widetilde{G}^{(l)}_{s_{l-1},j_l}(x; V^{(l-1)}_{s_{l-2},j_{l-1}})$.*

Step 3: *Generate a series of independent uniform random variables: $Y_{s_{L-1},j_L}, j_L = 1, \cdots, D^{(L)}_{s_{L-1}}$.*

Step 4: *Return $\bar{U}_{s_{L-1},j_L} = \psi^{(L-1)}_{s_{L-2},j_{L-1}}\left(-\dfrac{log(Y_{s_{L-1},j_L})}{V^{(L-1)}_{s_{L-2},j_{L-1}}}\right) = \psi^{(0)}_{0,1} \bigodot_{i=1}^{L-1} \widetilde{\Psi}^{(i)}_{s_{i-1},j_i}\left(-\dfrac{log(Y_{s_{L-1},j_L})}{V^{(L-1)}_{s_{L-2},j_{L-1}}}\right)$.*

Then $(\bar{U}_{1,1}, \cdots, \bar{U}^{(L)}_{s_{L-1},j_L}, \cdots, \bar{U}^{(L)}_{s_{L-1},J_L})$ is a sample from copula $C(u_1, \cdots, u_d)$.

Note that when $L = 2$ and $d = J$, Algorithm 4.2.1 reduces to the sampling algorithm of a two-level LSHAC proposed by Hering et al. (2010).

4.2.3 Structure and Estimation of a LSHAC

In this section, we discuss an estimation procedure for LSHACs, with a special focus on the determination of the hierarchical structure. Given a d-dimensional sample data with T observations, $\boldsymbol{x}_T = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d)_{T\times d}$, the log-likelihood function of the sample is defined by

$$L(\boldsymbol{\theta}) = \sum_{j=1}^{d}\sum_{t=1}^{T} \log f_j\big(x_{t,j}|\mathcal{F}_{t-1};\boldsymbol{\theta}^M\big) + \sum_{t=1}^{T} \log\big(c\big(F_1(x_{t,1}),\ldots,F_d(x_{t,d})|\mathcal{F}_{t-1};\boldsymbol{\theta}\big)\big) \quad (4.12)$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}^M, \boldsymbol{\theta}^C, \mathcal{S})$ is the parameter vector to be estimated, including the marginal parameter set, $\boldsymbol{\theta}^M$, the copula parameter set, $\boldsymbol{\theta}^C$, and the hierarchy structure $\mathcal{S}$; $\mathcal{F}_t$ is the information available up to time t; c is the corresponding copula density; F_j is the marginal c.d.f. of $\boldsymbol{x}_j$ with density f_j, where $j = 1, \ldots, d$.

The classical IFM estimation for copulas, in which $\boldsymbol{\theta}^M$ and $\boldsymbol{\theta}^C$ is calibrated in a two-step estimation procedure, is widely used and yields asymptotically efficient estimates (Joe, 1997; Patton, 2006). However, the ML estimation can only be employed to an HAC with a known hierarchical structure. Consequently, we propose an augmented IFM (AIFM) method with a three-stage procedure, which additionally determines the hierarchical structure of a LSHAC by using hierarchical clustering method. Our estimation procedure comprises of three stages: the first stage focuses on marginal distribution, second stage determines the optimal structure of a LSHAC, and finally by combining results from the first two stages, the third stage globally obtains the required ML estimators. We now describe these stages in greater details.

In the first stage we obtain ML estimator of each margin's parameter set, $\boldsymbol{\theta}_j^M$, $j = 1, \ldots, d$ from

$$\widehat{\boldsymbol{\theta}_{\boldsymbol{j}}^{\boldsymbol{M}}} \;=\; \underset{\boldsymbol{\theta}_j^M}{\operatorname{argmax}} \sum_{t=1}^{T} \log f_j\big(x_{t,j}|\mathcal{F}_{t-1};\boldsymbol{\theta}_j^M\big), \quad (4.13)$$

and produce the *pseudo-sample* $\boldsymbol{u} = (\boldsymbol{u_1}, \ldots, \boldsymbol{u_d})'$ by probability transformation with the estimated marginal distribution functions, namely

$$\boldsymbol{u} = (\boldsymbol{u_1}, \ldots, \boldsymbol{u_d})' \;=\; (\widehat{F}_1(\boldsymbol{x_1};\widehat{\boldsymbol{\theta}_{\boldsymbol{1}}^{\boldsymbol{M}}}), \ldots, \widehat{F}_d(\boldsymbol{x_d};\widehat{\boldsymbol{\theta}_{\boldsymbol{d}}^{\boldsymbol{M}}}))', \quad (4.14)$$

where $\widehat{F}_1(\boldsymbol{x_1};\widehat{\boldsymbol{\theta}_{\boldsymbol{1}}^{\boldsymbol{M}}}), \ldots, \widehat{F}_d(\boldsymbol{x_d};\widehat{\boldsymbol{\theta}_{\boldsymbol{d}}^{\boldsymbol{M}}})$ represent the estimates of the marginal probability transformations.

Let $\mathcal{S}$ be the true hierarchical structure that underlies the LSHAC. Given that $\mathcal{S}$ is unknown in practice, the objective of the second stage is to determine $\widehat{\mathcal{S}}$ that closely resemblances $\mathcal{S}$. As noted earlier, determining the optimal structure of a LSHAC is one of the key issues that has largely been ignored in the literature on LSHACs, despite its critical role on dependence modelling. Here we propose to determine the optimal structure of a LSHAC by resorting to the hierarchical clustering analysis (Jr., 1963; Székely and Rizzo, 2005; Zhang et al., 2013).

The hierarchical clustering procedure entails choosing an appropriate metric of dissimilarity between each pair of the pseudo sample, $(\boldsymbol{u_1}, \ldots, \boldsymbol{u_d})'$, obtained from the first estimation stage, where $\boldsymbol{u_j} = (u_{1,j}, \ldots, u_{T,j})'$ and $j = 1, \ldots, d$. The dissimilarity metric is used to construct a symmetric proximity matrix $\zeta = [d_{i,j}]$, where $d_{i,j}$ denotes a proximity index between the i-th and the j-th variables. Larger $d_{i,j}$ represents a higher level of dissimilarity. In hierarchical clustering, the Euclidean metric is one of the most commonly used dissimilarity metrics, where $d_{i,j}$ is given by

$$d_{i,j}^{E} = \sqrt{\sum_{t=1}^{T}(u_{t,i} - u_{t,j})^2}. \tag{4.15}$$

For the grouping of HAC models, Okhrin et al. (2013b) determine an All-GM-HAC and an All-CL-HAC by grouping the two variables with the largest Kendall's τ at each level of binary hierarchy. Along this line, the second metric is to employ the association between the variables defined as

$$d_{i,j}^{\tau} = 1 - \tau_{i,j}, \tag{4.16}$$

where $\tau_{i,j}$ is a dependence association (e.g., Kendall's τ) between u_i and u_j. This metric is widely used in the partition of HAC models. However, the drawback of this approach is that it fails to take into consideration the dissimilarity resulting from the distance, i.e., $d_{i,j}^{E}$, between variables. It is possible, for example, that $d_{i,j}^{\tau}$ is high but $d_{i,j}^{E}$ is low. To alleviate this problem, we advocate a new dissimilarity metric as follows:

$$d_{i,j}^{\tau\text{-}E} = \frac{d_{i,j}^{E}}{1 + \tau_{i,j}}. \tag{4.17}$$

We refer this metric as *τ-adjusted-Euclidean* (hereafter, τ-Euclidean) metric. Note that the proposed new metric integrates both Euclidean metric and Kendall's τ metric in such a way that a lower $d_{i,j}^{E}$ and a higher dependence (i.e., a larger $\tau_{i,j}$) lead to a smaller $d_{i,j}^{\tau\text{-}E}$.

As a result, the τ-Euclidean metric has the capability of simultaneously reflecting both the dissimilarity resulting from the Euclidean distance and the association between each variables. The simulation experiment to be conducted in Section 4.3 further confirms the superiority of the τ-Euclidean metric in correctly identifying the structure of LSHACs.

Given the calibrated margins $\widehat{F}_1(\boldsymbol{x_1}; \widehat{\boldsymbol{\theta_1^M}}), \ldots, \widehat{F}_d(\boldsymbol{x_d}; \widehat{\boldsymbol{\theta_d^M}})$ obtained in the first stage and the estimated hierarchical structure, $\widehat{\mathcal{S}}$, from the second stage, the final stage is to determine the ML estimator of copula parameter set $\boldsymbol{\theta}^C$ according to

$$\widehat{\boldsymbol{\theta}^C} = \underset{\boldsymbol{\theta}^C}{\operatorname{argmax}} \sum_{t=1}^{T} \log \left(c\big(F_1(x_{t,1}), \ldots, F_d(x_{t,d}) | \mathcal{F}_{t-1}; \widehat{\boldsymbol{\theta}^M}, \boldsymbol{\theta}^C, \widehat{\mathcal{S}}\big) \right). \tag{4.18}$$

The resulting AIFM estimator is denoted by $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\theta}^M}, \widehat{\boldsymbol{\theta}^C}, \widehat{\mathcal{S}})$, with an optimal hierarchical structure, $\widehat{\mathcal{S}}$, obtained from hierarchical clustering analysis.

It is important to distinguish our proposed estimation procedure from that of Okhrin et al. (2013b). The key difference is that Okhrin et al. (2013b) uses a multistage ML method to determine the structure of LSHACs as well as the copula parameters. The recursive nature of their proposed procedure implies that the final estimator of their LSHAC is sub-optimal. Furthermore, their recursive procedure applies only to generators with specified "separable" property (such as All-GM-HAC, see also Corollary 4.2.2, and All-CL-HAC). Consequently, this severely limits the application of their proposed strategy. In contrast, our proposed estimation procedure requires us to first determine the optimal structure of a LSHAC before we estimate the necessary parameters. The optimal structure is determined using the hierarchical clustering analysis involving some commonly used metrics as well as our proposed metric.

4.3 Simulation Analysis

In Section 4.2.3, we propose using the hierarchical clustering analysis for determining the optimal structure of LSHAC. In particular, three plausible grouping metrics based on Euclidean, Kendall's τ, and τ-Euclidean are described. Resorting to a simulation study, this subsection provides an in-depth analysis on the relative efficiency of these hierarchical clustering metrics. In our benchmark example, we assume a LSHAC model with the following known structure $\mathcal{S}$

$$C(u_1, \cdots, u_6) = C_{0,1}^{(0)}(C_{1,1}^{(1)}(C_{2,1}^{(2)}(C_{1,1}^{(3)}(u_1, u_2), u_3), u_4), C_{1,2}^{(1)}(u_5, u_6)), \tag{4.19}$$

and generators

$$\begin{aligned}
\psi_{0,1}^{(0)}(u) &= \psi_{GM}(u) = \exp(-u^{1/\theta}), && (4.20)\\
\psi_{1,1}^{(1)}(u) &= \psi_{GM\circ G}(u) = \exp\big(-\big(a_{1,1}\log(1+\frac{u}{b_{1,1}})\big)^{\frac{1}{\theta}}\big), && (4.21)\\
\psi_{1,2}^{(1)}(u) &= \psi_{GM\circ IG}(u) = \exp\big(-\big(a_{1,2}\sqrt{2u+b_{1,2}^2}-a_{1,2}b_{1,2}\big)^{\frac{1}{\theta}}\big), && (4.22)\\
\psi_{2,1}^{(2)}(u) &= \psi_{GM\circ G\circ IG}(u) = \exp\big(-\big(a_{1,1}\log(1+\frac{a_{2,1}}{b_{1,1}}(\sqrt{2u+b_{2,1}^2}-b_{2,1}))\big)^{\frac{1}{\theta}}\big), && (4.23)\\
\psi_{1,1}^{(3)}(u) &= \psi_{GM\circ G\circ IG\circ GM}(u)\\
&= \exp\big(-\big(a_{1,1}\log(1+\frac{a_{2,1}}{b_{1,1}}(\sqrt{2\exp(-u^{1/\theta_{3,1}})+b_{2,1}^2}-b_{2,1}))\big)^{\frac{1}{\theta}}\big). && (4.24)
\end{aligned}$$

Here the subscripts denote the outer generator and the Lévy subordinators used to construct the corresponding inner generators. For example, $\psi_{GM\circ G\circ IG\circ GM}(u)$ is an inner generator constructed by a GM outer generator and three sequential Lévy subordinators, namely, G, IG, and GM. This structure provides a four-level, six-dimensional copula with five generators. The parameter set of this LSHAC model is $\boldsymbol{\theta^C} = (\theta, a_{11}, b_{11}, a_{12}, b_{12}, a_{21}, b_{21}, \theta_{31}) = (1.3, 1.3, 10, 0.3, 9, 0.08, 9, 0.5)$. Using Algorithm 4.2.1, Figure 4.2 depicts the pairwise scatter plots of a simulated sample from this LSHAC.

Recall that the objective of the simulation study is to evaluate the efficiency of the various proximity metrics at identifying correctly the underlying structure of our benchmark LSHAC. This can be accomplished by first simulating samples from the LSHAC model, then estimating the structure of the simulated copula using the corresponding hierarchical clustering metric, and finally comparing to the true underlying structure. The step-by-step procedure is given as follows:

Step 1: Sample N sets of copula parameters $\widetilde{\boldsymbol{\theta}_n^C}$, $n = 1, \ldots, N$, from uniform distributions with range $[\boldsymbol{\theta}^C(1-\pi), \boldsymbol{\theta}^C(1+\pi)]$.

Step 2: For each n-th set of copula parameters, where $n = 1, \ldots, N$, sample M independent batches each of sample size T from a LSHAC with parameters $\widetilde{\boldsymbol{\theta}_n^C}$ and structure $\mathcal{S}$.

Step 3: For each m-th simulated batch of sample size T, where $m = 1, \ldots, M$, estimate $\widehat{\mathcal{S}_{n,m}}$ using the hierarchical clustering analysis.

Step 4: Calculate the reliability ratio, ρ_n, which measures the relative proportion of the estimated structures $\widehat{\mathcal{S}_{n,m}}, m = 1, \ldots, M$, that correctly identify the true structure

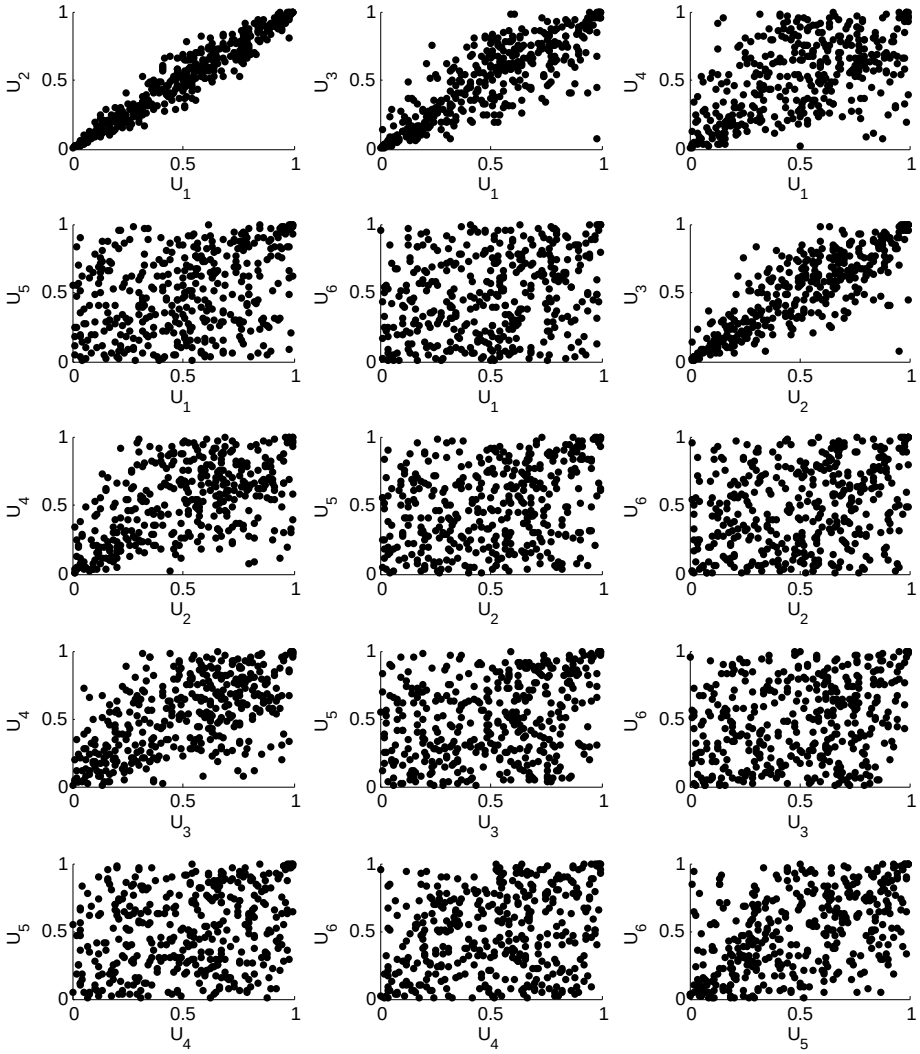


Figure 4.2: Scatter plots of a generated six-dimensional LSHAC with its structure expressed in (4.19) to (4.24) and parameter set $\boldsymbol{\theta}^{\boldsymbol{C}} = (\theta, a_{11}, b_{11}, a_{12}, b_{12}, a_{21}, b_{21}, \theta_{31}) = (1.3, 1.3, 10, 0.3, 9, 0.08, 9, 0.5)$.

$\mathcal{S}$; i.e.,

$$\rho_n = \sum_{m=1}^{M} \frac{\mathbb{I}_{\widehat{\mathcal{S}_{n,m}}=\mathcal{S}}}{M}, \ n = 1, \ldots, N,$$

where $\mathbb{I}_{\widehat{\mathcal{S}_{n,m}}=\mathcal{S}}$ is an indicator variable with value equals to one if the estimated structure coincides with the true structure, zero otherwise.

In our simulation experiments, we assume $N = 1000$, $M = 100$, and $T = 1000$. The hierarchical clustering analysis based on three proximity metrics is used to optimally determining the structures, $\widehat{\mathcal{S}_{n,m}}$, where $n = 1, \ldots, N$ and $m = 1, \ldots, M$. We also use $\pi = 10\%, 15\%$, and 20% to reflect the parameter uncertainty in the LSHAC. The mean and variance of the reliability ratio (ρ_n) over 1000 independent replications are summarized in Table 4.3.

Table 4.3: Mean and Variance of the Reliability Ratio (ρ_n)

Proximity Measure	**Statistics**	$\boldsymbol{\pi = 10\%}$	$\boldsymbol{\pi = 15\%}$	$\boldsymbol{\pi = 20\%}$
Euclidean	Mean	0.6536	0.6492	0.6426
	Variance	0.0108	0.0207	0.0318
Kendall's τ	Mean	0.7480	0.7425	0.7331
	Variance	0.0077	0.0148	0.0234
τ-Euclidean	Mean	0.8384	0.8323	0.8241
	Variance	0.0052	0.0099	0.0159

We draw the following conclusions based on the results in Table 4.3:

- The results clearly highlight the superiority of our newly proposed τ-Euclidean metric. This metric is able to correctly identify the true structure with at least 82% chance. Not only that this metric yields the highest reliability ratio, its variability (as measured by its sample variance) is also the smallest. While the Euclidean metric is the worst among the three metrics, it is comforting to know that it still has a success rate of at least 64%.

- As the degree of parameter uncertainty increases (i.e. by increasing π from 10% to 20%), the reliability ratio deteriorates slightly with increasing variability. This phenomenon is consistent for all three proximity metrics. It is, however, worth pointing out that while the performance declines with increasing parameter uncertainty, the changes are quite small and hence this provides some indication of the robustness of the underlying proximity metric at identifying the true structure of the underlying LSHAC.

Figure 4.3 plots the empirical cumulative distribution functions (eCDF) of the reliability ratios for the three proximity metrics for $\pi = 10\%, 15\%$, and 20%. It is also of interest to note that the eCDF of the reliability ratios based on τ-Euclidean metric lies under those of the other two metrics. According to the definition of stochastic ordering (See, for example, Hadar and Russell, 1969), the results of τ-Euclidean metric is first order stochastic dominance over the Euclidean and the association metric. As a result, τ-Euclidean metric achieves advantages of both Euclidean metric and dependence metric.

4.4 Empirical Analysis of Weather Risk in Canada

In this section, we analyze the systemic weather risk in Canada following the methodology in Section 4.2. First, the dataset used in this study is described in Section 4.4.1. Next, the marginal dynamics and spatial dependence of the data are analyzed in Section 4.4.2 and Section 4.4.3, respectively.

The general modeling framework of the empirical analysis in this chapter is summarized in Figure 4.4. The multivariate daily average temperature (DAT) model is constructed involving two steps. First, the marginal dynamic for each region i is analyzed with a wavelet technique from both time and frequency scales in order to obtain thorough information about the marginal dynamics of weather processes. Second, the dependence structure between different regions are constructed with the new proposed LSHAC model, which is shown to have better estimation performance compared to the traditional Gaussian copula and hierarchical Archimedean copula (HAC) models. Next, the weather index data are simulated according to the estimated joint distribution, and the corresponding weather derivatives are priced under a risk neutral measure. Finally, various weather hedging strategies are developed and the most efficient approach is identified.

4.4.1 Data

The data used in this paper includes the Adjusted and Homogenized Canadian Climate (AHCC) data, obtained from Environment Canada covering the years from 2001 to 2011. This dataset contains daily temperature series for eight provinces in Canada, including Alberta (AB), Saskatchewan (SK), British Columbia (BC), Manitoba (MB), Ontario (ON), New Brunwick (NB), Nova Scotia (NS), and Quebec (QC). The geographical locations of these provinces are pictured in Figure 4.5. These eight provinces were selected because they contain 98.72% of the farms and 99.26% of the aggregate farm incomes in Canada, and

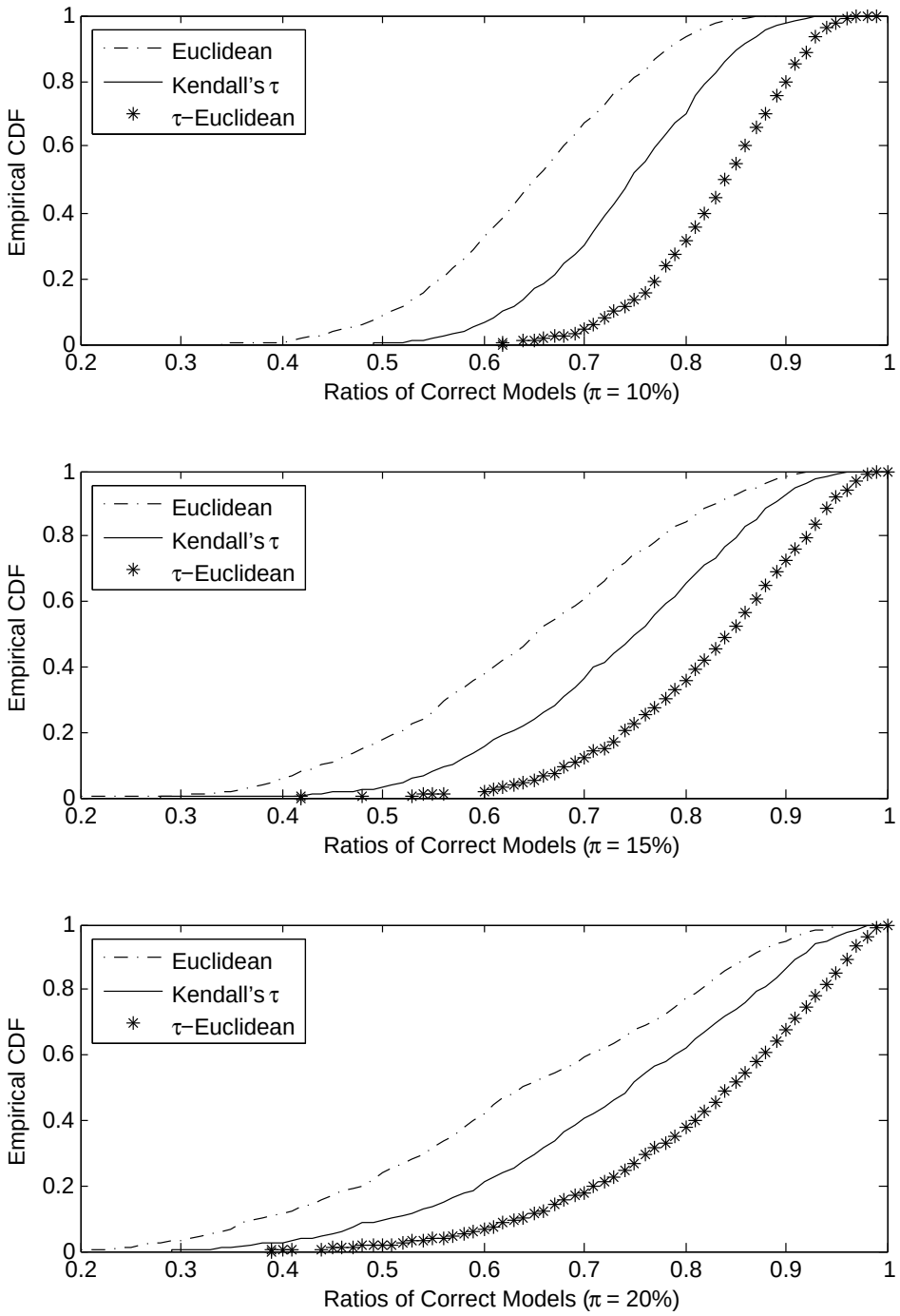


Figure 4.3: Empirical Cumulative Distribution Functions (eCDF) of the reliability ratios with $\pi = 10\%, 15\%$, and 20% under the three proximity metrics.

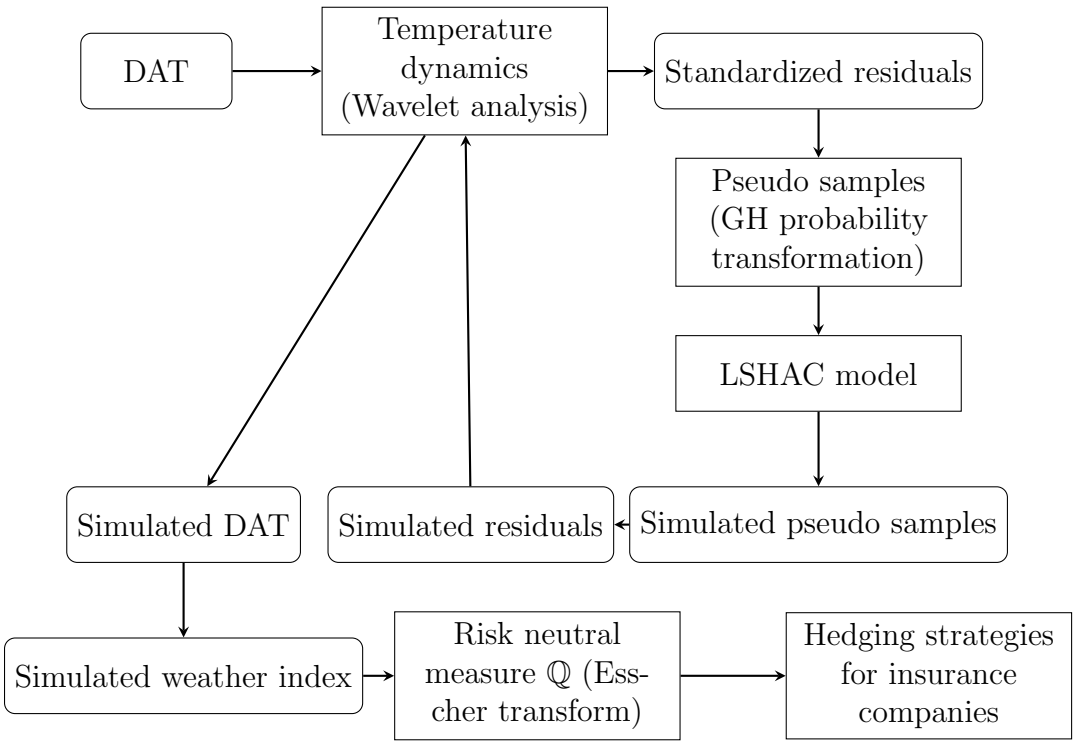


Figure 4.4: Flow chart of the general modeling framework in Chapter 4.

include most agricultural insurance programs in Canada. In addition, there are six weather derivative trading cities among these eight provinces, including Calgary (AB), Edmonton (AB), Vancouver (BC), Toronto (ON), Montreal (QC), and Winnipeg (MB).

Figure 4.5: Map of Canada by provinces

The descriptive statistics of the data are summarized in Table 4.4. The data displays obvious heavy tail properties, with the temperature series from all provinces showing negative skewness, and most large kurtosis. In addition, extreme risks appear for several provinces, such as AB, BC, NB, and QC, with the 1% quantile having temperature lower than -20 $^{\circ}C$. Moreover, the weather risk conditions are also found to vary in different regions. For example, the lowest historical temperature is -36.75 $^{\circ}C$ in Quebec, while in Ontario it is -19.8 $^{\circ}C$. A good understanding of the heterogeneity of weather risks across provinces provides an opportunity for insurers to diversify their risk portfolios and develop efficient hedging strategies. As an example, the time series and histogram of historical temperature data for Manitoba is displayed in Figure 4.6.

Table 4.4: Descriptive Statistics of Weather Data in Canada, including Alberta (AB), Saskatchewan (SK), British Columbia (BC), Manitoba (MB), Ontario (ON), New Brunwick (NB), Nova Scotia (NS), and Quebec (QC). The statistics include mean, standard deviation (SD), skewness, kurtosis, and 5% and 1% left quantiles ($Q_{0.05}$ and $Q_{0.01}$). The temperatures are recorded in $^{\circ}C$.

	AB	BC	MB	NB	NS	ON	QC	SK
Mean	-0.79	5.05	1.44	4.00	5.57	2.33	-0.49	-0.65
SD	6.07	5.16	13.81	9.10	7.39	8.57	6.07	8.56
Skewness	-1.58	-1.42	-0.34	-0.38	-0.36	-0.91	-1.05	-1.17
Kurtosis	6.19	5.72	2.00	2.84	2.48	2.99	3.93	4.13
$Q_{0.05}$	-13.77	-5.18	-22.73	-12.80	-7.40	-14.88	-12.83	-19.07
$Q_{0.01}$	-22.13	-11.64	-27.72	-18.80	-12.50	-20.44	-18.16	-26.33

4.4.2 Marginal Dynamics with Wavelet Analysis

In order to describe the nonstationarity and seasonality nature of the temperature data, many statistical models propose to decompose the DAT dynamic as follows (Alexandridis and Zapranis, 2013; Okhrin et al., 2013a),

$$Y_i(t) \quad = \quad \Gamma_i(t) + \Pi_i(t) + \Upsilon_i(t), \tag{4.25}$$

where $Y_i(t), \Gamma_i(t), \Pi_i(t)$, and $\Upsilon_i(t)$ are the DAT process, trend component, seasonality component, and adjusted temperature (i.e., residual part) at time t in area $i, i = 1, \ldots, d$, respectively.

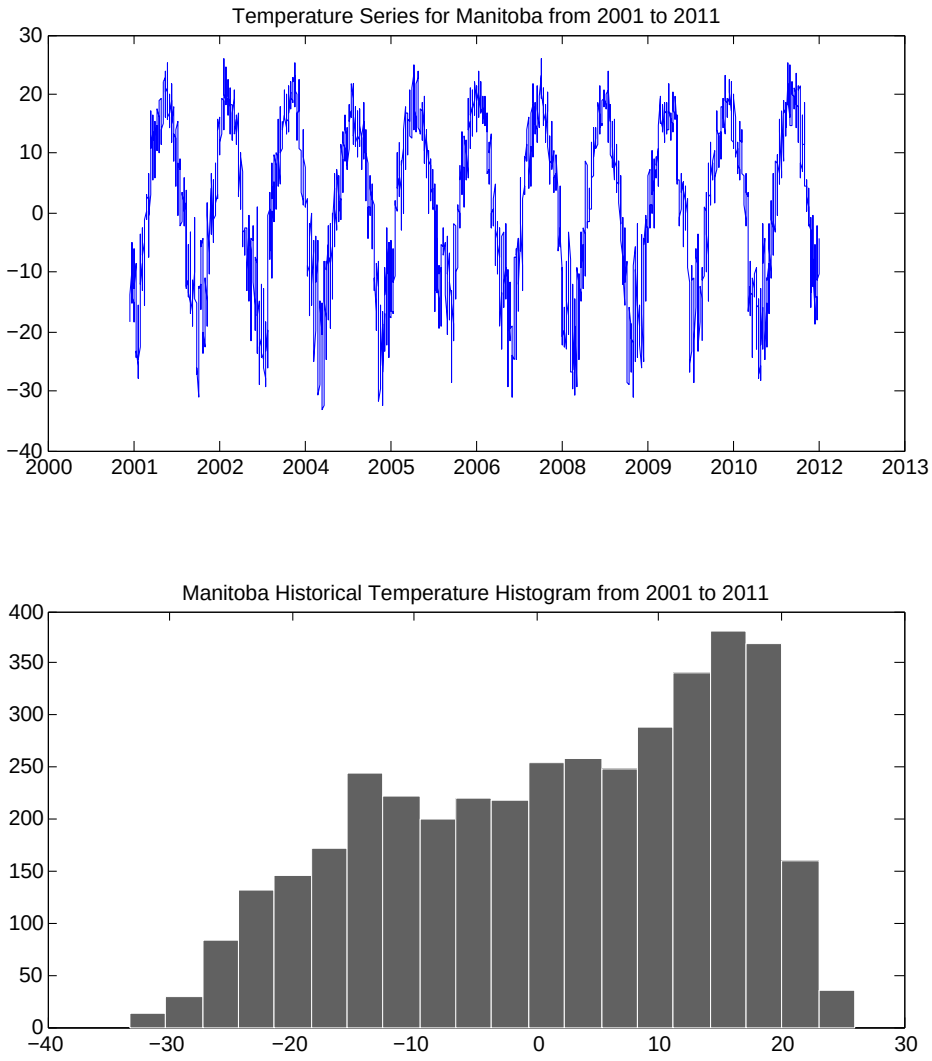


Figure 4.6: Time series and histogram of daily temperature data for Manitoba (2001-2011).

In studying the Canadian DAT, however, we find although Equation (4.25) illustrates some stylized general properties of daily temperature data, such as cyclical and seasonal trending, it does not capture the distinctive characteristics of Canadian DAT, which includes low temperatures that appear with higher frequency and more extreme values in the winter. Therefore, we propose to add a "shock" term, i.e. $\Delta_i(t)$, to emphasize this nature. Hence, the DAT decomposition becomes:

$$Y_i(t) = \Gamma_i(t) + \Delta_i(t) + \Pi_i(t) + \Upsilon_i(t) \tag{4.26}$$

In order to justify the $\Gamma_i(t), \Delta_i(t)$, and $\Pi_i(t)$ parts and to determine the compositions of the seasonal parts, wavelet analysis is performed. The wavelet transform decomposes certain time series into a time-frequency space, providing detailed analysis of the variability in the data. Power and Turvey (2010) apply wavelet analysis to study the long-range dependence in the volatility of commodity futures prices. Alexandridis and Zapranis (2013) use wavelet technique to study temperature process and price weather derivatives. A more thorough introduction of wavelet analysis and its application can be found in Daubechies (1990), Daubechies (1992), and Lau and Weng (1995). In general, a wavelet transform writes a real-valued signal $S(t)$ with respect to the complex-valued wavelet function $\psi(t)$ as

$$S(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \tilde{\psi}(\frac{t-b}{a}) S(t) dt, \tag{4.27}$$

where function $\tilde{\psi}(t)$ is the complex conjugate of the wavelet function $\psi(t)$, and $\psi(t)$ satisfies the following two conditions:

$$\int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty, \tag{4.28}$$

$$2\pi \int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty, \tag{4.29}$$

where $\Psi(\omega)$ is the Fourier transform of $\psi(t)$. There are different choices of wavelet functions $\psi(t)$, such as Haar wavelet, Meyer wavelet, Morlet wavelet, and Daubechies wavelet. Among these wavelet functions, Daubechies 10 wavelets are the most commonly used discrete wavelet transforms. Therefore, we use Daubechies 10 to decompose the Daily Average Temperature (DAT) from each region i. The scaling function and wavelet of Daubechies 10 are displayed in Figure 4.7. According to Wavelet analysis, the trend term, $\Gamma_i(t), shockterm, \Delta_i(t)$ and seasonality term, $\Pi_i(t)$ are modelled according to Equation (4.30), (4.31) and Equa-

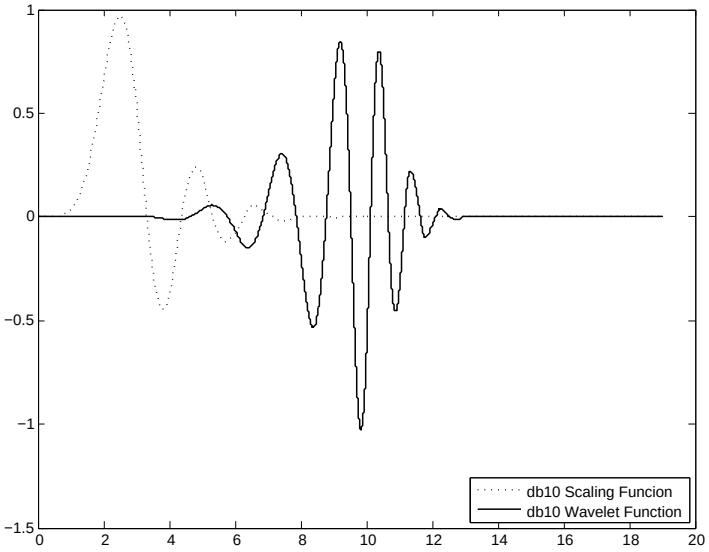


Figure 4.7: The scaling function and wavelet function for Daubechies 10 (The dotted line is the scaling function and the solid line is the wavelet function)

tion (4.32).

$$\Gamma_i(t) = \gamma_0 + \gamma_1 \frac{t}{365} \tag{4.30}$$

$$\begin{aligned} \Delta_i(t) &= \delta_1 \mathbb{I}_{M_1} + \delta_2 \mathbb{I}_{M_2} + \delta_3 \mathbb{I}_{M_3} + \delta_4 \mathbb{I}_{M_{11}} + \delta_5 \mathbb{I}_{M_{12}} \\ &+ \delta_6 \mathbb{I}_{M_{11-3}||T_i(t)<Q_{t_{-1},0.05}} + \delta_7 \mathbb{I}_{M_{11-3}||T_i(t)<Q_{t_{-1},0.01}} \end{aligned} \tag{4.31}$$

$$\begin{aligned} \Pi_i(t) &= \sum_{k=1}^{K} a_k \sin\left(2\pi\left(\frac{t-\psi_{a_k}}{k\cdot 365}\right)\right) + \sum_{s=1}^{S} b_k \cos\left(2\pi\left(\frac{t-\psi_{b_s}}{s\cdot 365}\right)\right) \\ &+ \sum_{v=1}^{V} c_v[1-\cos\left(2\pi\left(\frac{t-\psi_{c_v}}{v\cdot 365}\right)\right)] \end{aligned} \tag{4.32}$$

First, for the trend term $\Gamma_i(t)$ we observe significant upward linear trends in the temperature series as represented by the parameter γ_1 in Equation (4.30), and this is consistent with previous work by Alexandridis and Zapranis (2013). Second, it is important to recognize that extreme low temperatures are prevalent in Canada, and in many cases impact the success of agricultural production. Therefore, it is critical to model the shock term $\Delta_i(t)$. The winter period temperature, from November to March, for each region i is carefully studied and the extreme low temperatures are identified as a series of indicator functions $\mathbb{I}_A$, which is equal to 1 when event A happens and equal to 0 otherwise. To be more specific, the

shock term can be mathematically expressed in Equation (4.31), where $\mathbb{I}_{M_1}$ (or analogously, $\mathbb{I}_{M_2}, \mathbb{I}_{M_3}, \mathbb{I}_{M_{11}}, \mathbb{I}_{M_{12}}$) is the indicator of the event for the time in January (or February, March, November, December, respectively). $\mathbb{I}_{M_{11-3}||T_i(t)<Q_{t_{-1},0.05}}$ is the indicator for the event that the temperatures during the winter period are lower than the 5% past year left quantile, and $\mathbb{I}_{M_{11-3}||T_i(t)<Q_{t_{-1},0.01}}$ is for the 1% left quantile. Finally, we express the seasonality term $\Pi_i(t)$ as Equation (4.32). In particular, in addition to the regular sinusoid functions, $\Pi_i(t)$ also contains the quadratic terms of sinusoids. This indicates that long-term cycles and complex periodic dynamics of temperature process are captured by the wavelet analysis.

The coefficients ($\gamma_0, \gamma_1, \delta_u, a_k, b_s$, and c_v), the phases (ψ_{a_k}, ψ_{b_s}, and ψ_{c_v}), and the optimal orders (K, S, and V) in Equation (4.30), (4.31), and (4.32), where $u = 1, \ldots, 7$, $k = 1, \ldots, K$, $s = 1, \ldots, S$, and $v = 1, \ldots, V$, are estimated by least squares and selected based on the Bayesian information criterion (BIC). Maximum Likelihood (ML) method is used to estimate the optimal orders (m, n, p, and q) and coefficients ($c_i, \omega_i, \phi_{i,j}, \theta_{i,j}, \eta_{i,j}$, and $\xi_{i,j}$) in the ARMA (m, n)-GARCH (p, q) model (Equation (4.33) to (4.35)), and the best distribution for standard residuals $z_i(t)$ are determined based on BIC, where $i = 1, \ldots, 8$.

The residual parts $\Upsilon_i(t)$ in Equation (4.26) are estimated with a heteroskedasticity model with the general hyperbolic (GH) family, which has been shown to provide superior fit for the empirical data in the temperature residual modeling (Alexandridis and Zapranis, 2013; Bellini, 2005; Benth and Benth, 2005). More specifically, the autoregressive moving average-generalized autoregressive conditional heteroskedasticity (ARMA (m, n)-GARCH (p, q)) models (Bollerslev, 1986; Engle, 1982) are estimated to analyze the time-varying correlations of $\Upsilon_i(t)$. Further, the generalized hyperbolic (GH) distribution family (Barndorff-Nielsen, 1997) is used to model the volatility in order to capture the heavy tail and leptokurtic properties (Yang, 2011).

The ARMA (m, n)-GARCH (p, q) model with GH residuals are more specifically defined by

$$\Upsilon_i(t) = c_i + \sum_{j=1}^{m} \phi_{i,j} \Upsilon_i(t-j) + \sum_{j=1}^{n} \theta_{i,j} \epsilon_i(t-j) + \epsilon_i(t), \tag{4.33}$$

$$\epsilon_i(t) = \sqrt{h_i(t)} z_i(t), \tag{4.34}$$

$$h_i(t) = \omega_i + \sum_{j=1}^{p} \eta_{i,j} h_i(t-j) + \sum_{j=1}^{q} \xi_{i,j} \epsilon_i^2(t-j). \tag{4.35}$$

where $\Upsilon_i(t)$ is the residual part of the decomposed temperature series in the i-th province at time t; $\epsilon_i(t)$, $z_i(t)$, and $h_i(t)$ are the residual, standard residual, and the conditional variance

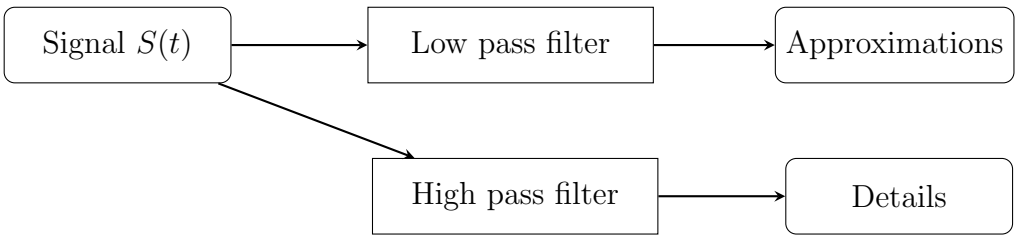


Figure 4.8: Discrete wavelet transform (DWT) for the signal to produce approximation coefficients and detail coefficients that contain information of the original signal

of the i-th province at time t, respectively; and $c_i, \omega_i, \phi_{i,j}, \theta_{i,j}, \eta_{i,j}, \xi_{i,j}$ are corresponding parameters.

The standard residuals $z_i(t)$ are modelled with the General Hyperbolic (GH) family, with the corresponding density function of $GH(\alpha, \beta, \delta, \gamma, \lambda)$ law

$$f_{GH}(x|\alpha,\beta,\delta,\gamma,\lambda) = e^{\beta(x-\lambda)} \frac{\left(\frac{\sqrt{\alpha^2-\beta^2}}{\delta}\right)^{\gamma}}{\sqrt{2\pi}\left(K_{\gamma}\left(\delta\sqrt{\alpha^2-\beta^2}\right)\right)} \frac{K_{\gamma-1/2}\left(\alpha\sqrt{\delta^2+(x-\lambda)^2}\right)}{\left(\frac{\sqrt{\delta^2+(x-\lambda)^2}}{\alpha}\right)^{1/2-\gamma}}, \quad (4.36)$$

where K_γ is the modified Bessel function of the second kind, α, β, and γ are parameters determining the shape of the GH distribution satisfying $\alpha > |\beta| \geq 0$ and $\gamma \in \mathbb{R}$, δ is the scale parameter, and λ is the shift parameter. The GH distribution becomes a hyperbolic distribution when $\gamma = 1$, and Normal Inverse Gaussian (NIG) distribution when $\gamma = -0.5$. The Student's t distribution and Variance Gamma (VG) distribution are included within the GH family as its limiting cases.

To illustrate, the wavelet decomposition of Manitoba is taken as an example. We perform discrete wavelet transform (DWT) with Daubechies 10 wavelet at level 11 to the temperature series from Manitoba. The outputs of the DWT provide the detail coefficients and approximation coefficients at 11 levels containing information regarding the original temperature series. The DWT procedure can be more clearly illustrated in Figure 4.8, and the decomposition results are displayed in Figure 4.9 and Figure 4.10. A slight upward trend in temperature process is observed as shown in the decomposed series, such as a_{11} and d_{11}. Canadian provinces tend to be more impacted by extreme low temperature during the winter period, and this property is captured by the wavelet analysis, where the first five details (d_1 to d_5) show great turbulences during the winter. For the seasonality, we can observe a one year period circle from a_1 to a_7 and d_8 (i.e., the first seven approximations and the the eighth detail). From the sixth detail (d_6), we find a 0.5 year period circle. Additionally, a_8, a_9, and d_9 show a two year circle, while a_{10} and d_{10} display a circle with a period of four.

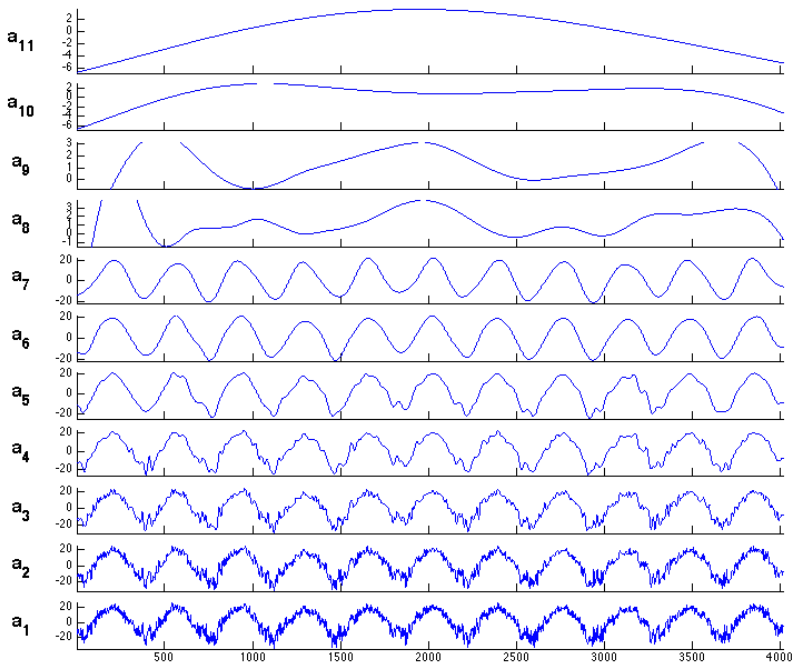


Figure 4.9: Wavelet Analysis of Historical Temperature in Manitoba, 2001 to 2011 (Approximations)

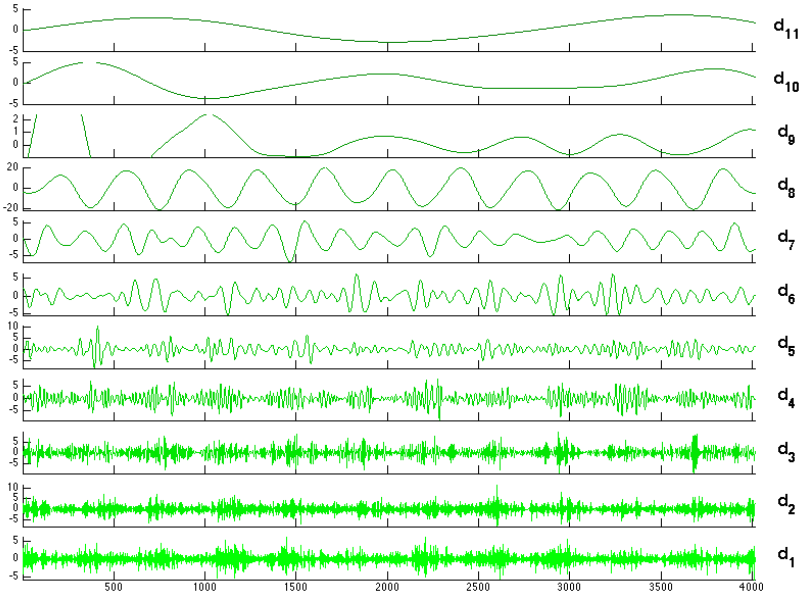


Figure 4.10: Wavelet Analysis of Historical Temperature in Manitoba, 2001 to 2011 (Details)

The estimating results are listed in Table 4.5 to Table 4.7. Table 4.5 displays the estimating results for the trend and shock parts. Positive trends ($\gamma_1 > 0$) exist in the historical temperature process of all eight provinces, indicating a statistically significant climate change effect. The winter season shocks ($\Delta_i(t)$) are significant at the 0.01 level for all provinces, except NB. This implies that low temperature during the winter effects most Canadian provinces. Shocks during January, for example, have a large negative effect on the time series of the temperature in the provinces of AB, ON, QC, and SK. The extreme low temperature during the winter season makes it important for crop insurance companies, who are usually weather risk takers, to hedge their weather risks in order to stabilize their risk portfolios and generate profits. Table 4.6 shows the estimating results for seasonality ($\Pi_i(t)$). Based on the BIC, seasonalities in AB, BC, MB, and ON are modelled with a summation of four $\cos^2$ terms, while the provinces of NS, QC, and SK model their seasonalities with a summation of four sin terms, and the seasonality in NB is modelled with a summation of four cos terms. This shows that the temperature process from the province of NB contains longer expanded cycles. Table 4.7 displays the estimating results for the ARMA(1,1)-GARCH(1,1) models with the GH residuals. The optimal residual distribution for NB is Variance Gamma (VG), and for the other seven provinces, the optimal marginal distribution for the residuals are Normal Inverse Gaussian (NIG).

Table 4.5: Estimating results for trends ($\Gamma_i(t)$) and shocks ($\Delta_i(t)$) of temperature processed from eight provinces in Canada. Values of t statistic are displayed in the brackets. "***" means significant at 0.01 level, "**" means significant at 0.05 level, and "*" means significant at 0.1 level.

		AB	BC	MB	NB	NS	ON	QC	SK
Trend	γ_0	-0.7882	5.2416	2.1918	2.5206	-2.7289	3.1544	-1.7258	-7.0516
		(-25.5611)***	(283.2865)***	(54.3637)***	(1.0849)	(-36.7273)***	(40.2966)***	(-31.1911)***	(-123.4392)***
	γ_1	0.0332	0.0054	0.1374	0.2219	0.1901	0.0925	0.2613	0.0856
		(1.4884)**	(28.6987)***	(154.1652)***	(0.4800)	(41.7917)***	(25.8648)***	(43.3680)***	(64.7446)***
Shock	δ_1	-7.2288	-5.4385	-5.5125	-2.5117	-5.0642	-9.0005	-7.6912	-8.9139
		(-2.3393)***	(-75.5364)***	(-25.3525)***	(-0.5568)	(-52.9220)***	(-13.4708)***	(-30.0510)***	(-88.5591)***
	δ_2	-5.6189	-2.8430	-5.3406	-2.2019	-3.1094	-7.9048	-7.2951	-7.3550
		(-2.7215)***	(-48.4545)***	(-11.6197)***	(-1.0133)	(-24.2533)***	(-4.8473)***	(-24.9135)***	(-65.9974)***
	δ_3	-3.0857	-1.1075	-2.8064	-0.9054	-0.0233	-4.1082	-5.9609	-4.5754
		(-1.0642)	(-38.4857)***	(-49.2319)***	(-1.8473)**	(-23.1348)***	(-32.2007)***	(-22.9898)***	(-65.3638)***
	δ_4	4.4745	2.5838	-1.8516	4.1377	-2.0467	1.2974	2.6413	2.9775
		(3.1169)***	(41.7153)***	(-38.4180)***	(0.4285)	(-17.5830)***	(16.3835)***	(16.6558)***	(34.4551)***
	δ_5	6.6845	4.4799	-4.5805	4.8370	-0.6492	2.4828	5.1586	5.0942
		(3.8455)***	(57.9552)***	(-30.3718)***	(1.5165)*	(-31.9479)***	(7.8462)***	(46.8879)***	(54.0337)***
	δ_6	0.8292	0.6818	1.1110	0.0036	0.3245	1.3499	-0.0898	2.0812
		(2.4616)***	(10.5084)***	(74.3215)***	(0.6418)	(19.3349)***	(2.3137)***	(-8.1510)***	(21.2005)***
	δ_7	1.8020	0.6721	0.5774	0.3241	-0.3889	-0.7021	-0.1904	1.3706
		(3.3684)***	(22.2028)***	(53.9478)***	(0.3171)	(-7.3731)***	(-3.6020)***	(-5.9030)***	(23.1412)***

Table 4.6: Estimating results for seasonality ($\Pi_i(t)$) of temperature processed from eight provinces in Canada. Values of t statistic are displayed in the brackets. "***" means significant at 0.01 level, "**" means significant at 0.05 level, and "*" means significant at 0.1 level. [a]

	AB	BC	MB	NB	NS	ON	QC	SK
$a_1/b_1/c_1$	-2.8574	-4.7873	-15.5948	-11.3118	8.2976	-7.6888	3.2052	7.5575
	(-5.5152)***	(-84.3172)***	(-127.4434)***	(-12.5873)***	(275.9175)***	(-11.9757)***	(197.2147)***	(177.1714)***
$a_2/b_2/c_2$	0.5178	-0.0196	0.8523	0.7481	-0.6810	0.6811	-0.8264	-0.6874
	(3.5575)***	(-42.3831)***	(939.1281)***	(0.8365)	(-37.1366)***	(8.2525)***	(-24.6348)***	(-38.1103)***
$a_3/b_3/c_3$	-0.6067	-0.5487	0.3317	0.1669	-0.0929	-0.3330	-0.3576	0.2838
	(-1.6267)*	(-18.1991)***	(41.1471)***	(0.4725)	(-126.0970)***	(-2.5213)***	(-131.3185)***	(68.5091)***
$a_4/b_4/c_4$	-0.3559	-0.4216	0.8913	1.0131	0.6871	0.8956	-1.0711	-0.3970
	(-3.7448)***	(-46.7319)***	(27.5787)***	(1.3156)*	(50.5850)***	(14.2602)***	(-14.0708)***	(-68.1067)***
$\psi_{a_1}/\psi_{b_1}/\psi_{c_1}$	-59.2484	-60.0032	-70.5054	16.8743	59.4274	-50.4186	11.8413	22.4791
	(-1.4851)*	(-144.2460)***	(-673.9152)***	(1.9613)**	(295.4596)***	(-160.3534)***	(19.9947)***	(75.3143)***
$\psi_{a_2}/\psi_{b_2}/\psi_{c_2}$	-55.8884	9.0365	-202.8639	47.2855	81.8202	-139.0177	119.9941	73.8153
	(-1.6387)*	(62.1956)***	(-9.9053)***	(1.4091)*	(45.9719)***	(-7.6449)***	(52.6440)***	(46.8854)***
$\psi_{a_3}/\psi_{b_3}/\psi_{c_3}$	-82.6372	-73.4053	234.1601	-154.9447	241.9686	233.5305	-124.4196	20.6872
	(-3.1977)***	(-59.0252)***	(18.8922)***	(-0.9070)	(41.0943)***	(2.2288)***	(-67.7938)***	(55.0413)***
$\psi_{a_4}/\psi_{b_4}/\psi_{c_4}$	327.1032	342.5273	-177.0218	217.5468	-530.4357	-231.8575	117.6337	200.1544
	(2.4239)***	(124.0579)***	(-184.7719)***	(0.7870)	(-34.0278)***	(-8.2881)***	(36.2485)***	(126.5108)***

[a] Optimal models are $a_k(\sin\text{terms})$ for provinces NS, QC, and SK, $b_s(\cos\text{terms})$ for provinces NB, and $c_i(\cos^2\text{terms})$ for provinces AB, BC, MB, and ON. $(k, s, i = 1, 2, 3, 4)$

Table 4.7: Estimating results for residuals ($\Upsilon_i(t)$) of temperature processed from eight provinces in Canada. Values of t statistic are displayed in the brackets. "***" means significant at 0.01 level, "**" means significant at 0.05 level, and "*" means significant at 0.1 level. [3]

	AB	BC	MB	NB	NS	ON	QC	SK
c	0.4716	0.1049	0.1978	0.0141	0.2627	0.0972	0.2627	0.4364
	(9.0714)***	(3.3002)***	(2.3683)***	(0.4236)	(5.0305)***	(2.3231)**	(5.9512)***	(6.6811)***
ω	0.2041	0.2253	0.2040	0.1593	-0.0402	0.0321	0.1423	0.1593
	(9.2429)***	(124.0579)***	(8.1196)***	(5.2108)***	(-34.0278)***	(1.5995)*	(4.6270)***	(6.6881)***
ϕ_1	0.5754	0.1479	0.1552	0.2150	0.2354	0.0952	0.1491	0.2181
	(5.1166)***	(3.3002)***	(2.6728)***	(2.3246)***	(5.0305)***	(2.2276)***	(2.7712)***	(1.9701)**
θ_1	0.8712	0.9773	0.9834	0.6774	0.9783	0.9773	0.8741	0.9528
	(79.9812)***	(254.6255)***	(281.2559)***	(33.3144)***	(247.5172)***	(239.6713)***	(80.1137)***	(6.6811)***
η_1	0.6708	0.7428	0.8923	0.9112	0.8331	0.9013	0.8826	0.8653
	(15.5867)***	(15.0908)***	(40.1683)***	(38.0292)***	(21.4820)***	(33.9433)***	(31.9240)***	(17.3746)***
ξ_1	0.2644	0.1854	0.0944	0.0625	0.1221	0.0776	0.0884	0.0959
	(6.6870)***	(5.0871)***	(5.0441)***	(4.4450)***	(4.5109)***	(4.1252)***	(4.5254)***	(2.9078)***
α	1.0964	0.9818	1.8781	2.4477	1.4350	1.7215	1.3112	1.2440
	(16.1451)***	(17.1439)***	(10.1084)***	(13.4747)***	(13.3460)***	(11.6934)***	(14.2693)***	(15.8703)***
β	-0.1715	-0.1841	-0.2043	-0.0459	-0.1750	-0.3696	-0.1837	-0.3264
	(-4.1894)***	(-5.0194)***	(-2.7459)***	(-0.5583)	(-3.0475)***	(-5.0818)***	(-3.3527)***	(-6.0781)***
Distribution	NIG	NIG	NIG	VG	NIG	NIG	NIG	NIG

4.4.3 Spatial Dependence

After removing the serial correlations from the DAT data and performing the GH probability transformation according to the model in Section 4.4.2, the next step is to appropriately model the joint distribution of the resulting pseudo sample, $\boldsymbol{u} = (u_1, \ldots, u_d)'$. This is a critical step because failing to capture an accurate spatial dependence structure of the weather variables may lead to large basis risk in the resulting hedging strategies. In this section, the spatial dependence of the temperature process of eight Canadian provinces are modelled and analyzed with the LSHAC model. Following the modelling procedure and notation system in Section 4.2.3, the resulting structure is displayed in Figure 4.11 and can be described as follows:

- The structure emanates from the outer copula ($\boldsymbol{C}_{0,1}^{(0)}$) at level 0.
- At level 1, the structure is classified into two subgroups.
  - The first subgroup contains five provinces in the west and middle territories of Canada (including BC, AB, SK, MB, and ON), and is nested together into the inner copula $\boldsymbol{C}_{1,1}^{(1)}$.
  - The second subgroup contains three provinces in the east (including NB, NS and QC), and is nested into $\boldsymbol{C}_{1,2}^{(1)}$.
- At level 2, provinces from the west, middle and east parts of Canada are grouped in to different subgroups.
  - Western provinces (including BC, AB and SK) are nested together by $\boldsymbol{C}_{1,1}^{(2)}$.
  - Middle provinces (including MB and ON) are grouped together by $\boldsymbol{C}_{1,2}^{(2)}$.
  - Eastern provinces NB and NS are first nested together by $\boldsymbol{C}_{2,1}^{(2)}$, and then grouped into $\boldsymbol{C}_{1,2}^{(1)}$ with QC.
- At level 3, AB and SK are nested together into the inner copula $\boldsymbol{C}_{1,1}^{(3)}$.

We find that the hierarchal structure in Figure 4.11 resembles the geographical positioning of the eight provinces as shown in Figure 4.5. For example, BC, AB and SK are three neighbouring provinces in the western part of Canada, and are grouped together. The two provinces in the middle of the Canadian territory, MB and ON, are in the same subgroup. The three provinces in the eastern part of Canada, namely, QC, NB, and NS, are grouped together into another subgroup. This hierarchical structure provides information about weather risks in different geographical regions. In addition, this structure also indicates associative relationships between different provinces. The Kendall's τ matrix is displayed

in Table 4.8. We can see that weather risks in regions within the same subgroup are more associative compared to regions in other subgroups. This implies that it is important for insurance companies to consider the dependence structure of the risk portfolio they hold in order to develop targeted hedging strategies.

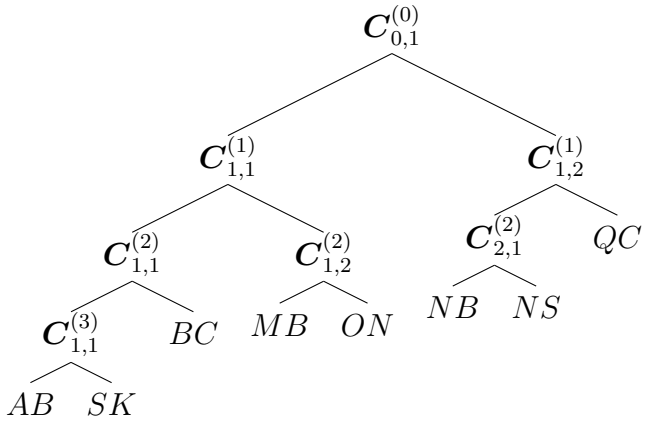


Figure 4.11: Hierarchical structure of the temperature process for eight Canadian provinces

Table 4.8: Kendall's τ of Temperature Data Between Each Provinces

	AB	BC	MB	NB	ON	NS	QC	SK
AB	1.00	0.51	0.25	0.27	0.30	0.35	0.43	0.46
BC	0.51	1.00	0.56	0.55	0.57	0.61	0.45	0.67
MB	0.25	0.56	1.00	0.65	0.52	0.65	0.36	0.70
NB	0.27	0.55	0.65	1.00	0.60	0.62	0.52	0.60
ON	0.30	0.57	0.52	0.60	1.00	0.68	0.44	0.55
NS	0.35	0.61	0.65	0.62	0.68	1.00	0.51	0.65
QC	0.43	0.45	0.36	0.52	0.44	0.51	1.00	0.46
SK	0.46	0.67	0.70	0.60	0.55	0.65	0.46	1.00

According to Theorem 2.1 in Hering et al. (2010), the copulas at each node can be constructed by composing an outer copula to the Lévy subordinator. Therefore, the LSHAC are highly flexible models with a large number of candidate models. For example, in the modeling of this paper the outer copulas of the LSHAC are selected as a Gumbel copula (GM) or Clayton copula (CL) as listed in Table 4.1. The Lévy subordinators are chosen from the three processes listed in Table 4.2, including Gamma process (G), Stable process (GM) and the Inverse Gaussian process (IG). As a consequence, to calibrate the eight dimensional LSHAC model with seven AC generators in Figure 4.11, we have $2 \times 3^6 = 1458$ models to chose from. Instead of going through all of the combinations, the estimation begins from the second level

of the structure (i.e., start from estimating the optimal copulas of $\boldsymbol{C}_{1,1}^{(2)}, \boldsymbol{C}_{1,2}^{(2)}, \boldsymbol{C}_{2,1}^{(2)}$, and $\boldsymbol{C}_{1,1}^{(3)}$), and the GM copula is found to provide the best fit. Therefore, the Lévy subordinators for $\boldsymbol{C}_{1,1}^{(2)}, \boldsymbol{C}_{1,2}^{(2)}, \boldsymbol{C}_{2,1}^{(2)}$, and $\boldsymbol{C}_{1,1}^{(3)}$ are fixed as a GM, reducing the candidate model to 18. It is important to emphasize that the GM and CL generators are selected for the outer copula in order to model asymmetric tail dependence of the data. For example, conditional on an extreme low temperature in Manitoba, the neighbouring province of Ontario is highly probable to have a very low temperature too, indicating a lower tail dependence property. Similarly, an upper tail dependence means that extreme high temperatures tend to appear together for neighbouring provinces. In contrast to the Gaussian copula, which does not have tail dependence, the GM copula has upper tail dependence and the CL copula has lower tail dependence, as shown in the third and forth columns in Table 4.1. Therefore, the LSHAC models potentially have the advantage of capturing the clusters of the extreme values in the data.

The estimating results are displayed in Table 4.9. The first seven columns describe the LSHAC model, including how to choose the Archimedean copulas at each node. The log-likelihood function values (LLF), as well as the BIC values (BIC) for each model, and the improvement in BIC of the LSHAC models compared to the Gaussian copula model (BIC Imp.) are shown. In particular, a positive sign indicates better performance compared to the Gaussian copula, while a negative sign indicates worse performance. The last column in the table shows the number of parameters in each copula model (No. Para). The first 9 LSHAC models are constructed with the GM as their outer generators (denoted as GM-LSHAC), and the last 9 LSHAC models are constructed with the CL as their outer generators (denoted as CL-LSHAC).

We obtain the following information based on the estimation results.

- First, based on BIC more than half of the estimated LSHAC models perform better than the Gaussian copula, and they are highlighted in bold with a "$\star$" in the parentheses. In particular, the best LSHAC model has a 88.76 improvement in BIC, and is constructed with the CL copula as the outer generator, G Lévy subordinator in the first level and GM Lévy subordinator in the second and the third level (the model in the last row in Table 4.9).
- Second, the LSHAC models are more efficient in the sense that they have better fitting abilities with smaller sets of parameters. To be more specific, for the particular structure in Figure 4.11, the LSHAC models have at most 9 parameters, compared to the Gaussian copula, which has 28 parameters to estimate.
- The first LSHAC model in the table, which is highlighted in bold with a "†", is con-

structed in such a way that the copula at each node in the structure in Figure 4.11 is a GM generator, denoted as All-GM-HAC. This is currently the most common HAC model used in empirical analysis, such as Okhrin et al. (2013a), which uses an All-GM-HAC model to estimate the dependence structure of the temperature process in China. However, as is shown in Table 4.9, the estimation ability of this model is limited, mainly because it restricts the copula at each node of the structure as a GM copula. Therefore, this paper introduces a more flexible LSHAC model with a large number of candidates to improve the estimation abilities.

- Finally, comparing the results of the GM-LSHAC models and the CL-LSHAC models, the CL-LSHAC models are found to perform slightly better. This may be explained by the difference in the tail dependence properties. To be more specific, the CL copula, as shown in Table 4.1, has lower tail dependence, meaning that it can capture low temperatures that appear together. In contrast, the GM copula only has upper tail dependence, which models the clusters of extreme high temperatures. The results show that lower tail dependence models (i.e., CL-LSHAC models) have better fitting results, indicating that the clustering of extreme low temperatures may be more important than extreme high temperature in Canada.

4.4.4 Esscher Transform and Pricing Formulas

When the market is complete, a unique risk neutral measure can be obtained by changing the process of the underlying asset into a martingale, and the securities can be priced as the expectation of the discounted derivative payoff under the risk neutral measure. However, the weather market is incomplete and there exists more than one equivalent risk neutral measures (Tankov, 2004). Therefore, traditional arbitrage-free theory cannot be applied in pricing securities written on weather indices, since the underlying assets cannot be traded.

The pricing methodology employed in this chapter uses a martingale measure based on the conditional Esscher transform (Bühlmann et al., 1996; Gerber and Shiu, 1994), which has been widely used in financial and insurance securities pricing in incomplete markets (Li et al., 2010; Siu et al., 2004; Yang, 2011). Note that the dependence structure does not change under both measures. We define a $\mathcal{F}_t$-adapted stochastic process $\{\zeta_t | t = 1, 2, \ldots, T\}$ as follows:

$$\zeta_T = \prod_{t=1}^{T} \frac{\exp\big(\theta Y(t)\big)}{E^{\mathbb{P}}\big(\exp(\theta Y(t))|\mathcal{F}_{t-1}\big)}, \tag{4.37}$$

Table 4.9: LSHAC estimating results for the eight dimensional hierarchical structure in Figure 4.11. The first seven columns describe copulas at each node in the LSHAC model. The next column refers to the log-likelihood function values of different models (LLF), the Bayesian Information Criterion (BIC) is shown next, followed by the improvement in BIC of the LSHAC models compared to the Gaussian copula model (BIC Imp.). In particular, a positive sign indicates better performance relative to the Gaussian copula, while a negative sign indicates worse performance. The last column shows the number of parameters in each copula model (No. Para). The first 9 LSHAC models are constructed with the GM as their outer generators (denoted as GM-LSHAC), and the last 9 LSHAC models are constructed with the CL as their outer generators (denoted as CL-LSHAC).

Gaussian Copula							**LLF**	**BIC**	**BIC Imp.**	**No. Para**
							2743.27	-2628.43	-	28
LSHAC Model										
$C_{0,1}^{(0)}$	$C_{1,1}^{(1)}$	$C_{1,2}^{(1)}$	$C_{1,1}^{(2)}$	$C_{1,2}^{(2)}$	$C_{2,1}^{(2)}$	$C_{1,1}^{(3)}$	**LLF**	**BIC**	**BIC Imp.**	**No. Para**
GM	**GM**	**GM**	**GM**	**GM**	**GM**	**GM**	2247.98	**-2194.66** (†)	-433.77	7
GM	**GM**	**IG**	**GM**	**GM**	**GM**	**GM**	2346.85	-2293.54	-334.89	8
GM	**GM**	**G**	**GM**	**GM**	**GM**	**GM**	2377.09	-2323.77	-304.66	8
GM	**IG**	**GM**	**GM**	**GM**	**GM**	**GM**	2595.82	-2542.51	-89.52	8
GM	**IG**	**IG**	**GM**	**GM**	**GM**	**GM**	2694.62	**-2641.31** ($\star$)	+12.88	9
GM	**IG**	**G**	**GM**	**GM**	**GM**	**GM**	2723.77	**-2670.45** ($\star$)	42.02	9
GM	**G**	**GM**	**GM**	**GM**	**GM**	**GM**	2608.84	-2555.52	-72.91	8
GM	**G**	**IG**	**GM**	**GM**	**GM**	**GM**	2708.01	**-2654.70** ($\star$)	+26.27	9
GM	**G**	**G**	**GM**	**GM**	**GM**	**GM**	2738.12	**-2684.80** ($\star$)	+56.37	9
CL	**GM**	**GM**	**GM**	**GM**	**GM**	**GM**	2605.90	-2552.58	-75.85	7
CL	**GM**	**IG**	**GM**	**GM**	**GM**	**GM**	2619.31	-2565.99	-62.44	8
CL	**GM**	**G**	**GM**	**GM**	**GM**	**GM**	2620.65	-2567.34	-61.09	8
CL	**IG**	**GM**	**GM**	**GM**	**GM**	**GM**	2725.76	**-2672.44** ($\star$)	+44.01	8
CL	**IG**	**IG**	**GM**	**GM**	**GM**	**GM**	2718.17	**-2664.86** ($\star$)	+36.43	9
CL	**IG**	**G**	**GM**	**GM**	**GM**	**GM**	2722.18	**-2668.87** ($\star$)	+40.44	9
CL	**G**	**GM**	**GM**	**GM**	**GM**	**GM**	2731.76	**-2678.45** ($\star$)	+50.02	8
CL	**G**	**IG**	**GM**	**GM**	**GM**	**GM**	2765.19	**-2711.88** ($\star$)	+83.45	9
CL	**G**	**G**	**GM**	**GM**	**GM**	**GM**	2770.51	**-2717.19** ($\star$)	+88.76	9

where θ is the parameter of the Esscher transform representing the market price of risk (MPR) charged for the weather derivatives. Usually θ describes the risk preferences of policy holders. Hence, a new martingale measure with respect to θ, $\mathbb{Q}_\theta$, can be defined as

$$\frac{d\mathbb{Q}_\theta}{d\mathbb{P}} \mid_{\mathcal{F}_T} = \zeta_T. \tag{4.38}$$

The Esscher transform has several advantages. First and foremost, as a generalization of the Girsanov transform, the Esscher transform changes the jump size (i.e., the price of jump risk) of the process under $\mathbb{Q}_\theta$ (Hubalek and Neilsen, 2006; Tankov, 2004). Second, the Esscher transform leads to a minimal entropy martingale measure, which is closest to the original physical measure (Frittelli, 2000; Hubalek and Neilsen, 2006; Tankov, 2004). Finally, many distributions stay invariant under the Esscher transform in the sense that their density functions retain their original form. This makes the Esscher transform easy to obtain and apply for pricing in a practical sense. The GH family has this invariable property, namely, the distribution with $GH(\alpha, \beta, \delta, \gamma, \lambda)$ law becomes $GH(\alpha, \beta + \theta, \delta, \gamma, \lambda)$ under the marginal measure $\mathbb{Q}_\theta$.

4.5 Hedging Weather Risks

A hedging example is developed in this section, in which weather exposures are hedged with index-based instruments. The purpose of this example is twofold. First, by applying the statistical model proposed in Section 4.4, we assess the potential benefits of the LSHAC dependence assumption in reducing basis risk and improving weather risk hedging. Second, four hedging strategies are developed to investigate the geographical aggregation levels on the effectiveness of weather risk hedging performance.

A financial weather contract is a weather contingent contract that pays claims based on future realization of weather events determined from certain weather indices. It can take the form of either a weather derivative (WD) or a weather index-based insurance (WIBI) product. Both are triggered by the underlying weather index, which is the common feature that is a main focus of this chapter from a risk management viewpoint. The differences between WD and WIBI are primarily a concern for regulators and policy makers (Dischel and Barrieu, 2002). Therefore, in this example, we do not identify the differences between the two unless necessary and refer to both as WDs.

Indices based on temperature have been shown to exhibit strong correlation with crop yield

(Parodi, 2104), and temperature derivatives have been good contracts to hedge weather risks that are traded on Chicago Mercantile Exchange (CME) (Woodard and Garcia, 2008a;b). It follows, therefore, that temperature indices serve as feasible proxies to assess the weather risk exposures of the insurance company. The most popular weather index, heating degree days (HDD), is defined as the difference between the daily average temperatures (DAT) and the base temperature ($\widetilde{T}$) if DAT falls below $\widetilde{T}$; otherwise it is assigned zero. Other popular temperature indices include Cooling Degree Days (CDD) and Cumulative Average Temperature (CAT). CDD is assigned zero if the DAT is smaller than $\widetilde{T}$; otherwise it is the difference between the DAT and the base temperature $\widetilde{T}$. CAT is calculated by summing the DAT over the contract period. In this chapter, we focus on HDD weather derives to construct a proxy portfolio of the insurance company.

4.5.1 Hedging Strategies

In this section, we develop four hedging strategies and assess the effectiveness of the hedging performance. The following assumptions are made:

- We assume that the farmers from 8 provinces in Canada buy WDs to protect their crop yield losses. Hence the weather risk exposure of the insurance company is a collection of WD contracts.
- The WDs are based on seasonal accumulated HDD (AccHDD) over the growing season (May - October) with the form: $(P_i - K_i)_+ = \max(0, P_i - K_i)$, where P_i is the AccHDD in province i, defined as

$$P_i \quad = \quad \int_{t_1}^{t_2} HDD_t, \tag{4.39}$$

  where $[t_1, t_2]$ represents the growing season.
- To simplify, we assume that the risk portfolio of the insurance company has equal weights in each province. Hence, the total exposure of the insurance company can be expressed as

$$X^{\text{Exp}} = \sum_{i=1}^{d} \frac{1}{d}(P_i - K_i)_+. \tag{4.40}$$

- The weather risk exposure is modeled according to a portfolio totaling $1000 million, which is consistent with Agriculture and Agri-Food Canada (2015) and the Word Bank

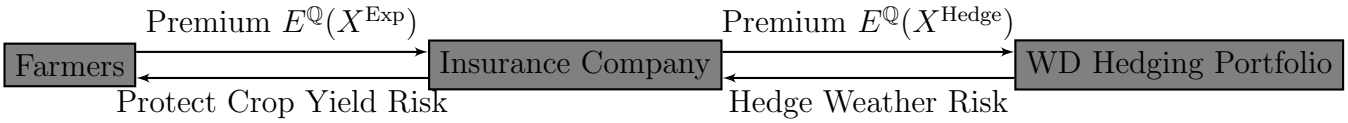


Figure 4.12: Flow chart of the transactions.

survey (Mahul and Stutley, 2010), which reports the agricultural insurance premium in Canada in 2008 as $1090 million.

- The objective of the insurance company is to find an appropriate hedging portfolio, denoted as X^{Hedge}, to hedge against the weather risk.
- Prices of WDs are calculated under $\mathbb{Q}$ measure with Esscher transform, while hedging performances are examined under $\mathbb{P}$ measure.
- Only non-linear hedging strategies with call option type WDs are considered.
- The hedging strategy is developed based on a budget constraint, such that the price of the hedging portfolio is no more than the price of the risk exposure of the insurance company.

Figure 4.12 shows the transactions between each stakeholder including the farmers, the insurance company and the WD traders. As stated before, this insurance company holds business across 8 provinces in Canada, i.e., the "business set" of this company can be expressed as

$$\mathcal{B} = \Big\{\{\text{AB}\}, \{\text{SK}\}, \{\text{BC}\}, \{\text{MB}\}, \{\text{ON}\}, \{\text{NB}\}, \{\text{NS}\}, \{\text{QC}\}\Big\}.$$

The hedging portfolio can be any subset of $\mathcal{B}$. For example, the company can use the weather index from any individual province to hedge its weather risk exposure; or it can also use the weather indices from several provinces (e.g., $\Big\{\{AB\}, \{SK\}\Big\}$ or $\Big\{\{AB\}, \{BC\}, \{MB\}\Big\}$) to hedge its weather risk. Theoretically, without any prior information, the insurance company has 2^8 choices to construct the hedging portfolio. With the LSHAC approach proposed in Section 4.4, in contrast, the company is able to develop an appropriate hedging strategy according to the dependence structure information from the LSHAC model.

Strategy 1: *Local hedging strategy*

The idea of local hedging is that the insurance company buys WD contracts from only one province with the form

$$X_{1,g}^{\text{Hedge}} = (P_g - K_g)_+, \tag{4.41}$$

where $g \in \mathcal{B}$. In other words, the hedging portfolio consists of γ_g shares of WD contract from province g. Hence, the portfolio after hedging, denoted as $X_{1,g}^{\text{HP}}$, can be expressed as

$$X_{1,g}^{\text{HP}} = \gamma_g X_{1,g}^{\text{Hedge}} - X^{\text{Exp}} + E^{\mathbb{Q}}(X^{\text{Exp}}) - E^{\mathbb{Q}}(X_{1,g}^{\text{Hedge}}). \tag{4.42}$$

In this case, the objective of the insurance company is to decide γ_g, such that

$$\min_{\gamma_g} \sqrt{Var^{\mathbb{P}}(X_{1,g}^{\text{HP}})},$$

$$\text{subject to } E^{\mathbb{Q}}(X_{1,g}^{\text{Hedge}}) \leq E^{\mathbb{Q}}(X^{\text{Exp}}). \tag{4.43}$$

In contrast to local hedging, the remaining three strategies are global hedging. In theory, the higher the geographical aggregation level, the more offsetting of risks in the portfolio (i.e., natural diversification), therefore, the remaining risk is more systematic. This relationship leads to the following hypothesis:

H_0: Hedging strategies with higher geographical aggregation levels are more effective.

As a result, global hedging strategies are proposed to test the hypothesis of the geographical aggregation effect by introducing different levels of spatial aggregation into the hedging portfolios.

Strategy 2: *Three parts global hedging strategy*

According to the hierarchical structure in Figure 4.11, this strategy divides the hedging portfolio into three parts,

$$\begin{aligned} X_2^{\text{Hedge}} &= X_{2,g_{2,1}}^{\text{Hedge}} + X_{2,g_{2,2}}^{\text{Hedge}} + X_{2,g_{2,3}}^{\text{Hedge}}, & (4.44)\\ X_{2,g_{2,j}}^{\text{Hedge}} &= (\sum_{g_j \in g_{2,j}} \omega_{g_j} P_{g_j} - \sum_{g_j \in g_{2,j}} \delta_{g_j} K_{g_j})_+, \quad j = 1,2,3, & (4.45) \end{aligned}$$

where $g_{2,1} = \Big\{\{AB\},\{SK\},\{BC\}\Big\}, g_{2,2} = \Big\{\{MB\},\{ON\}\Big\}, g_{2,3} = \Big\{\{NB\},\{NS\},\{QC\}\Big\}$, and δ_{g_j} is defined as:

$$\delta_{g_j} = \begin{cases} \frac{1}{3} & j = 1, \\ \frac{1}{2} & j = 2, \\ \frac{1}{3} & j = 3. \end{cases}$$

Figure 4.13 displays the geographical location of the three parts.

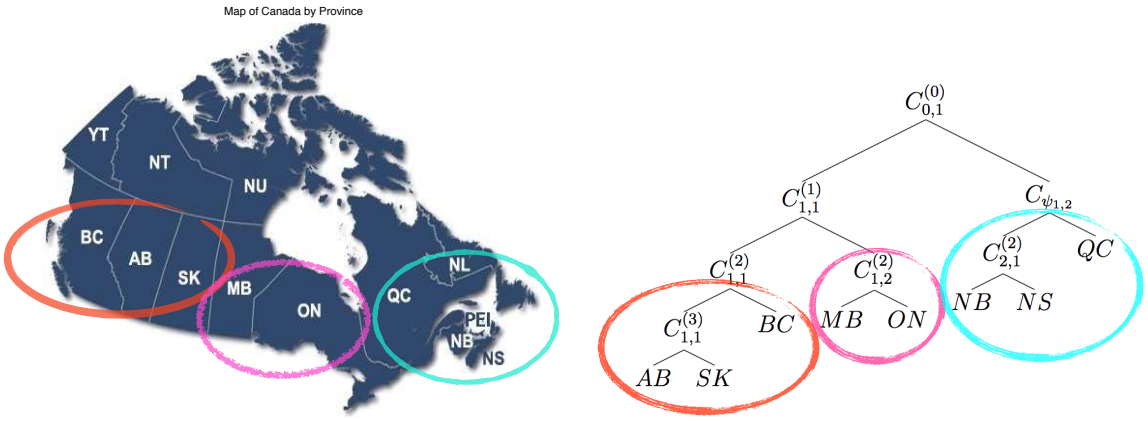


Figure 4.13: Illustration of the three parts global hedging strategy, where the eight provinces are aggregated into three parts based on the hierarchical structure in Figure 4.11 and neighbour provinces are put into the same hedging portfolios

Hence, the corresponding hedged portfolio is $X_2^{\text{HP}} = X_2^{\text{Hedge}} - X^{\text{Exp}} + E^{\mathbb{Q}}(X^{\text{Exp}}) - E^{\mathbb{Q}}(X_2^{\text{Hedge}})$. The objective of this hedging strategy is to solve the following optimization problem:

$$\min_{\omega_{g_j}, j=1,2,3} \sqrt{\sum_{j=1}^{3} Var^{\mathbb{P}}(X_2^{\text{HP}})},$$

$$\text{subject to} \quad E^{\mathbb{Q}}(X^{\text{Exp}}) = E^{\mathbb{Q}}(X_2^{\text{Hedge}}), \tag{4.46}$$

$$\sum_{j=1}^{3} \sum_{g_j \in g_{2,j}} \omega_{g_j} = 1. \tag{4.47}$$

Strategy 3: *Two parts global hedging strategy*

The two parts global hedging strategy increases the geographical aggregation level by dividing the eight provinces into two parts based on the hierarchical structure from Figure 4.11.

$$X_3^{\text{Hedge}} = X_{3,g_{3,1}}^{\text{Hedge}} + X_{3,g_{3,2}}^{\text{Hedge}}, \tag{4.48}$$

$$X_{3,g_{3,k}}^{\text{Hedge}} = \Big(\sum_{g_k \in g_{3,k}} \omega_{g_k} P_{g_k} - \sum_{g_k \in g_{3,k}} \delta_{g_k} K_{g_k} \Big)_+, \quad k = 1, 2, \tag{4.49}$$

where $g_{3,1} = \Big\{\{AB\},\{SK\},\{BC\},\{MB\},\{ON\}\Big\}, g_{3,2} = \Big\{\{NB\},\{NS\},\{QC\}\Big\}$, and δ_{g_k} is defined as

$$\delta_{g_k} = \begin{cases} \frac{1}{5} & j = 1, \\ \frac{1}{3} & j = 2. \end{cases}$$

Figure 4.14 displays the geographical location of the two parts, where the first part contains five provinces and the second part contains three provinces.

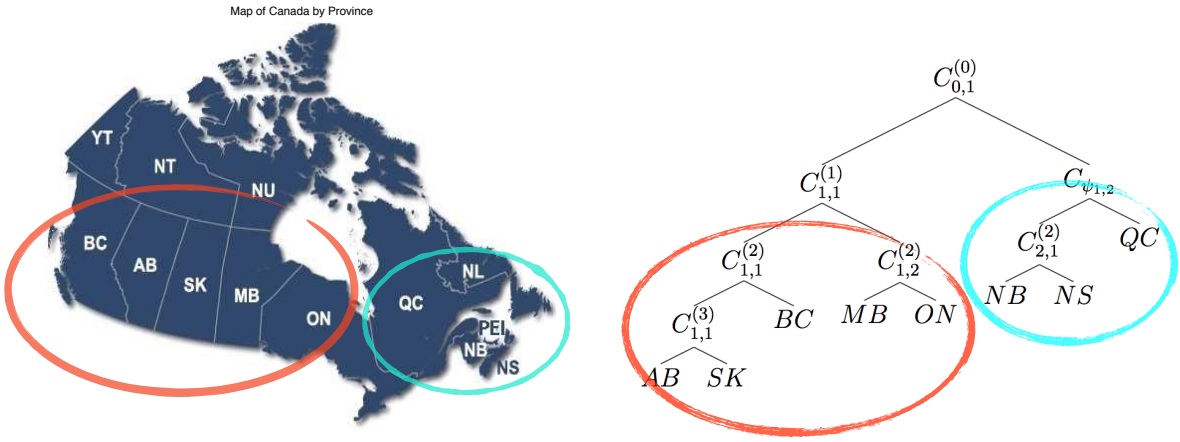


Figure 4.14: Illustration of the two parts global hedging strategy, where the eight provinces are aggregated into two parts based on the hierarchical structure in Figure 4.11, and neighbour provinces are put into the same hedging portfolios

The optimization problem of this hedging strategy becomes

$$\min_{\omega_{g_k}, k=1,2} \sqrt{\sum_{k=1}^{2} Var^{\mathbb{P}}(X_3^{\text{HP}})},$$

$$\text{subject to} \quad E^{\mathbb{Q}}(X_3^{\text{Hedge}}) \leq E^{\mathbb{Q}}(X^{\text{Exp}}), \tag{4.50}$$

$$\sum_{k=1}^{2} \sum_{g_k \in g_{3,k}} \omega_{g_k} = 1. \tag{4.51}$$

Strategy 4: *One part global hedging strategy*

The one part global hedging strategy aggregates all eight provinces into one hedging portfolio. Therefore, this hedging strategy has the highest geographical aggregation level, which may

have the most natural diversification effect.

$$X_4^{\text{Hedge}} = (\sum_{g\in\mathcal{B}} \omega_g P_g - \sum_{g\in\mathcal{B}} \delta_g K_g)_+, \tag{4.52}$$

and solves the following optimization problem

$$\min_{\omega_g, g\in\mathcal{B}} \sqrt{Var^{\mathbb{P}}(X_4^{\text{HP}})},$$

$$\text{subject to} \quad E^{\mathbb{Q}}(X_4^{\text{Hedge}}) \leq E^{\mathbb{Q}}(X^{\text{Exp}}), \tag{4.53}$$

$$\sum_{g\in\mathcal{B}} \omega_g = 1. \tag{4.54}$$

4.5.2 Hedging Effectiveness

In this section, we discuss the results of different hedging strategies. We are interested in the following problems: (1) the implication of a hedged vs unhedged portfolio (i.e. the necessity of hedging weather risk); (2) the importance of the assumed underlying dependence structure; (3) the geographical aggregation effect on hedging effectiveness. First, we define hedging effectiveness based on three criteria:

1. *Weather risk variance reduction:* Following Li and Hardy (2011), we define the hedging efficiency of certain hedging strategy $*$, Ef_*, as its risk reduction effect. To be more specific,

$$Ef_* = 1 - \frac{Var^{\mathbb{P}}(X_*^{\text{HP}})}{Var^{\mathbb{P}}(X^{\text{UHP}})}. \tag{4.55}$$

   This implies that better hedging strategies have Ef_* values closer to one. On the contrary, low Ef_* values indicate poor hedging performances. Obviously, without hedging, which we can take it as a "do nothing" strategy, the efficiency is zero.

2. *Weather risk value-at-risk (VaR):* For each hedged portfolio, X_*^{HP}, we calculate the VaR at 1% level, defined as $VaR_{0.01} = F^{-1}_{X_*^{\text{HP}}}(0.01)$, where the subscription "*" denotes a certain hedging strategy. As the 1% quantile of the hedged portfolio, $VaR_{0.01}$ characterizes the left tail of the hedged portfolio distribution. Therefore, a high value of $VaR_{0.01}$ indicates a better hedging strategy. To compare the hedging effectiveness, we also calculate $VaR_{0.01}$ of the unhedged portfolio, denoted as X^{UHP}.

3. *Weather risk conditional tail expectation (CTE):* The 1% level CTE, also called Expected Shortfall (ES), of a random variable X, defined as $CTE(0.01) = E(X|X < VaR(0.01))$, calculates the average losses that have exceeded $VaR_{0.01}$, providing more information about the extreme scenarios. As a result, CTE is sometimes preferred by risk managers in practice (Acerbi and Tasche, 2002). Therefore, we calculate and compare the $CTE_{0.01}$ of each hedged (and unhedged) portfolio to see the hedging effectiveness.

The first criteria measures the weather hedging efficiency in terms of the variance reduction effect, while the second and the third focus on the reduction in the downside risk, i.e., the worst-case scenario of the portfolio. We compare the hedging performances for both Independent and LSHAC copula assumptions. Our analysis is under different MPR parameters assuming θ to be $\{0, 0.1, 0.3, 0.5\}$. The results of the different hedging strategies are displayed in Tabler 4.10, Table 4.12, Table 4.11, and Figure 4.15.

Necessity of Hedging Weather Risk

An efficient hedging strategy should achieve a large reduction in risk, and help the insurance company maintain stable future cash flows. The hedging error distributions for the original portfolio without any hedges are shown in Table 4.10. In Figure 4.15, the simulated hedging error densities of the best local hedging strategy (i.e., local-BC strategy), and the best global hedging strategy (i.e., one part global hedging strategy), are displayed. It is obvious that all strategies are able to reduce the portfolio risks significantly under both dependence structure assumptions, since they have reduced the dispersion of the portfolios. We can also observe a significant risk reduction effect of the weather hedging from Table 4.11, which displays the hedging efficiencies of different hedging strategies. In general, all hedges reduce the variance. The best hedge, i.e., one part global hedge, has hedging efficiency of more than 96% for both dependence structure assumptions and all MPR assumptions. In fact, even the hedge with the worst performance among all strategies, i.e., the local-ON strategy, can reduce the variance by 35% or more.

In addition to variance reduction, the weather hedge also reduces the downside risk for insurance companies. Table 4.12 displays the $VaR_{0.01}$ and $CTE_{0.01}$ of simulated unexpected cash flows for he risk portfolio with no hedge. From a risk management perspective, $VaR_{0.01}$ represents the quantile of extreme losses, and $CTE_{0.01}$ is the expected value of extreme losses. Therefore, weather hedging is effective in reducing both the probability of loss and the severity of losses, since all hedges reduce both the $VaR_{0.01}$ and $CTE_{0.01}$ compared to the original portfolios. For example, under the LSHAC copula assumption and $\theta = 0.1$, the local-AB strategy can reduce the $VaR_{0.01}$ by 82.07% ($\frac{(-2359.68)-(-423.07)}{-2359.68}$) and $CTE_{0.01}$

by 89.22%($\frac{(-4112.45)-(-443.51)}{-4112.45}$). Similarly, one part global hedging strategy can reduce the $VaR_{0.01}$ by 92.21% ($\frac{(-2359.68)-(-183.85)}{-2359.68}$) and $CTE_{0.01}$ by 95.34% ($\frac{(-4112.45)-(-191.56)}{-4112.45}$). Similar results can be obtained with other MPR assumptions. In summary, weather risk hedges play an essential role for insurance companies to stablize incomes and reduce losses, which is necessary for helping to ensure a sustainable firm structure.

Importance of Dependence Structures

It is important to understand the impact of introducing the dependence structure in the statistical modeling of temperature with respect to improving hedging performance. In particular, under all MPR assumptions the LSHAC model has better hedging performance compared to the independent assumption. In Table 4.12, we can see that the LSHAC model can reduce the downside risk of the portfolio further than the independent assumption for the insurance company. As an example, when $\theta = 0.1$, one part global hedging can reduce the $VaR_{0.01}$ by $ 2175.83 million and $CTE_{0.01}$ by $ 3920.89 million under the LSHAC model assumption. In contrast, the $VaR_{0.01}$ and $CTE_{0.01}$ reductions are $2186.89 million and $3599.28 million for the independent assumption. Therefore, by comparing the CTE reduction, we can see that the LSHAC model reduces extreme weather downside risk by more than $ 321.61 million compared to the independent model. Similarly, in Table 4.11, under each MPR assumption the LSHAC models achieve better hedging efficiencies relative to the independent assumption.

Geographical Aggregation Effect

The empirical results support the hypothesis that the effectiveness of the hedging strategy is dependent on the geographical aggregation effect. More specifically, we find that hedging strategies with higher levels of aggregation have superior performance in hedging systematic weather risk. The local hedging strategy has the lowest level of geographical aggregation among all hedging strategies, and also has the worst performance compared to the global hedging strategies. It is interesting to note that the one part global hedging strategy which has the highest spatial aggregation level among the three global hedging strategies, is most effective in hedging weather risk. These results are consistent with previous work by Woodard and Garcia (2008a;b), which showed that agricultural hedging can be more effective as the spatial aggregation in the risk exposure and hedging instrument increases.

We first compare the weather risk VaR and CTE in Table 4.12. As an example, under the LSHAC model assumption when $\theta = 0.1$, the $VaR_{0.01}$ and $CTE_{0.01}$ of the unexpected cash flows under the best local hedging strategy (i.e., local-BC strategy) are -$684.30 million and -$1496.53 million, respectively. In contrast, global hedging strategies have better per-

formance. In particular, the $VaR_{0.01}$ and $CTE_{0.01}$ of the hedging errors under three parts (two parts) global hedging strategy are -$304.47 (-$297.19) million and -$317.69 (-$307.82), respectively. The one part global hedging strategy has the highest geographical aggregation level, with $VaR_{0.01}$ and $CTE_{0.01}$ of the hedging errors of -$183.85 and -$191.56. The results in Table 4.11 also show supporting evidence regarding spatial aggregation, where under the LSHAC model and $\theta = 0.1$ assumptions the local-BC strategy has hedging efficiency of 0.8877, while the one part global hedging strategy increases the hedging efficiency to 0.9680. Similar observations are also found with the other dependence structure assumptions and MPR assumptions.

Table 4.10: Summary of statistics of simulated distributions of unexpected cash flows with no hedge. Independent and LSHAC dependent structure assumptions are compared. In addition, results for different MPR assumptions are displayed.

	Independent				LSHAC			
	$\theta = 0$	$\theta = 0.1$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0$	$\theta = 0.1$	$\theta = 0.3$	$\theta = 0.5$
μ	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
σ	694.87	649.98	580.47	504.99	764.79	696.61	595.06	502.17
VaR	-2565.15	-2374.44	-1984.92	-1537.56	-2556.11	-2359.68	-1958.02	-1519.87
CTE	-4202.92	-3796.43	-3115.51	-2355.27	-4649.17	-4112.45	-3210.29	-2339.93

Table 4.11: Hedging efficiencies of four strategies. Local hedging strategies have eight choices, i.e., the insurance company can select to use the HDD from eight provinces to hedge the weather risks. Independent and LSHAC assumptions are compared. In addition, results for different MPR assumptions are displayed.

		Independent				LSHAC			
		$\theta = 0$	$\theta = 0.1$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0$	$\theta = 0.1$	$\theta = 0.3$	$\theta = 0.5$
Local	AB	0.8331	0.8331	0.8331	0.8331	0.8680	0.8680	0.8680	0.8680
	BC	0.9026	0.9025	0.9010	0.8984	0.8881	0.8877	0.8853	0.8816
	MB	0.7393	0.7393	0.7393	0.7393	0.8011	0.8011	0.8011	0.8011
	NB	0.8318	0.8340	0.8323	0.8287	0.8095	0.8113	0.8087	0.8042
	NS	0.8191	0.8191	0.8191	0.8191	0.8597	0.8597	0.8597	0.8597
	ON	0.3498	0.3498	0.3498	0.3498	0.4568	0.4568	0.4568	0.4568
	QC	0.8257	0.8257	0.8257	0.8257	0.8628	0.8628	0.8628	0.8628
	SK	0.7459	0.7459	0.7459	0.7459	0.8068	0.8068	0.8068	0.8068
Global	3 Parts	0.9117	0.9117	0.9117	0.9117	0.9283	0.9283	0.9283	0.9283
	2 Parts	0.9143	0.9143	0.9139	0.9140	0.9284	0.9283	0.9283	0.9282
	1 Part	0.9609	0.9609	0.9609	0.9609	0.9680	0.9680	0.9680	0.9680

Table 4.12: $VaR_{0.01}$ and $CTE_{0.01}$ of simulated distributions of unexpected cash flows for four strategies. Local hedging strategies have eight choices, i.e., the insurance company can select to use the HDD from eight provinces to hedge the weather risks. Independent and LSHAC dependent structure assumptions are compared. In addition, results for different MPR assumptions are displayed.

			Independent				LSHAC			
			$\theta = 0$	$\theta = 0.1$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0$	$\theta = 0.1$	$\theta = 0.3$	$\theta = 0.5$
Local	AB	VaR	-420.29	-421.20	-431.07	-438.52	-412.81	-423.07	-433.50	-439.65
		CTE	-442.17	-445.52	-450.97	-460.09	-438.36	-443.51	-453.32	-461.21
	BC	VaR	-727.45	-674.45	-587.96	-509.06	-764.78	-684.30	-585.96	-496.81
		CTE	-1501.70	-1347.81	-1065.62	-803.34	-1716.65	-1496.53	-1120.96	-796.32
	MB	VaR	-477.91	-447.11	-427.29	-403.25	-462.49	-430.22	-409.88	-386.37
		CTE	-484.88	-452.92	-432.69	-407.87	-467.64	-435.68	-414.76	-390.42
	NB	VaR	-1104.70	-1035.53	-848.55	-622.88	-1155.29	-1048.03	-825.65	-613.86
		CTE	-2095.92	-1873.33	-1494.25	-1078.37	-2339.13	-2039.84	-1547.33	-1070.91
	NS	VaR	-392.81	-392.23	-401.30	-402.55	-382.13	-383.17	-389.70	-391.44
		CTE	-399.93	-401.49	-411.90	-414.31	-389.36	-391.47	-398.07	-401.50
	ON	VaR	-2071.66	-1985.32	-1829.45	-1528.52	-2052.41	-1950.61	-1798.12	-1494.03
		CTE	-2452.40	-2338.49	-2201.95	-1936.49	-2436.65	-2319.44	-2177.01	-1965.92
	QC	VaR	-456.68	-469.04	-492.10	-491.69	-444.52	-449.96	-471.69	-477.85
		CTE	-511.50	-519.18	-533.02	-537.39	-486.60	-494.38	-511.31	-527.18
	SK	VaR	-488.16	-466.49	-452.81	-422.15	-472.52	-445.42	-436.57	-411.36
		CTE	-534.49	-519.26	-505.52	-477.53	-514.05	-492.41	-485.36	-461.17
Global	3 Parts	VaR	-313.69	-305.87	-301.14	-295.42	-315.93	-304.47	-301.09	-293.58
		CTE	-332.26	-319.47	-317.36	-310.34	-331.02	-317.69	-317.05	-311.71
	3 Parts	VaR	-307.28	-295.57	-295.37	-288.31	-310.14	-297.19	-293.68	-286.56
		CTE	-317.18	-307.59	-310.71	-299.52	-321.53	-307.82	-306.50	-300.14
	1 Part	VaR	-192.09	-187.55	-187.87	-181.49	-189.56	-183.85	-183.08	-180.19
		CTE	-202.54	-197.15	-199.56	-195.62	-197.14	-191.56	-192.26	-190.09

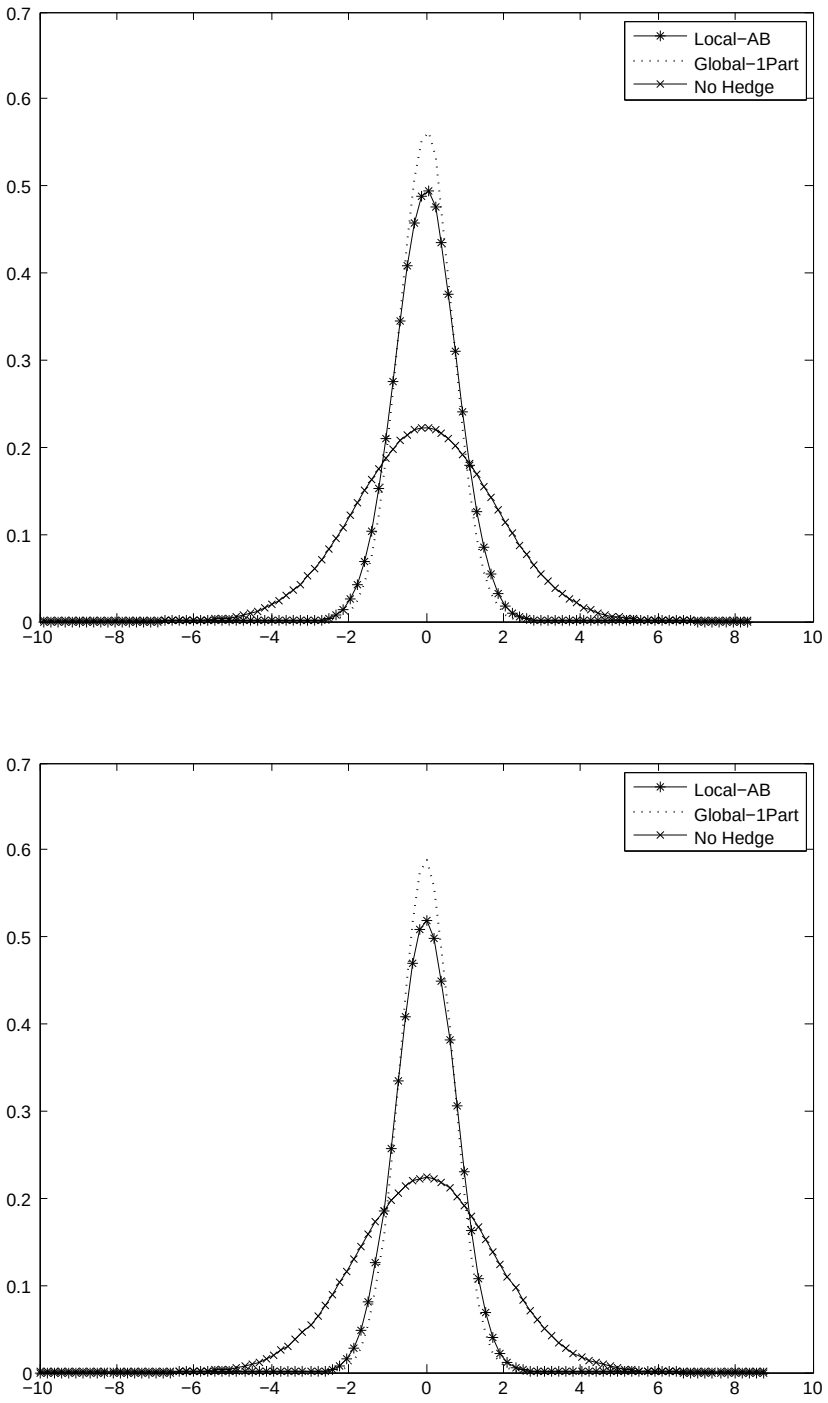


Figure 4.15: Simulated distributions of unexpected cash flows for different hedging strategies (MPR assumption: $\theta = 0$) (The first figure is for Independent dependence structure assumption and the second figure is for LSHAC copula assumption; Line with stars is for the best local hedging strategy, dotted line is for the best global hedging strategy, and line with crosses is for the original portfolio with no hedge.)

4.6 Conclusions

This chapter constructed a generalized multi-level LSHAC model and designed an estimation procedure that focused on a suitable grouping method to determine the hierarchical structure. We employed hierarchical clustering analysis with Euclidean metric, Kendall's τ metric, and τ-Euclidean metric to determine the grouping of variables. The simulation study showed that the newly proposed τ-Euclidean metric achieved a balance between distance and association measure, providing the better performance in identifying the true structure.

In the empirical analysis, the proposed estimation methodology was applied to the geographical dependence structure of the temperature processes in Canada. General LSHACs with the structure determined by τ-Euclidean metric produces better modelling performances than elliptical copulas and All-GM-HACs. In particular, we study the systemic weather risk in Canada and develop different weather risk hedging strategies for agricultural insurers and reinsurers. In order to reduce the basis risk and improve the efficiency of weather hedging, we refine the statistical framework of weather variables with a flexible marginal dynamic and a new copula model. Wavelet analysis is employed to study the detail characteristics of the weather data from both time and frequency scales, and the general hyperbolic (GH) family is used to capture the heavy-tail property of the marginal processes. This is the first time that the Lévy subordinated hierarchical Archimedean copula (LSHAC) model is proposed for the weather dependence modeling. The results lend support to the importance of capturing the appropriate dependence structure of weather risk. The LSHAC model reduces extreme weather downside risk by $ 3920.89 million, which is $ 321.61 more risk reduction compared to the independent model assumption, leading to more efficient hedging strategies. Moreover, the empirical hedging results support the hypothesis that higher levels of geographical aggregation achieve more efficient hedging strategies.

4A Appendix: Proofs

4A.1 Proof of Theorem 4.2.1

We prove Theorem 4.2.1 by induction.

For level $y=0$, since $\psi_{0,1}^{(0)}$ is a LT-AC generator with c.d.f $G(v_{0,1}^{(0)})$, we have

$$\psi_{0,1}^{(0)} = \int_0^\infty \exp\{ - v_{0,1}^{(0)} \cdot u\} dG_{0,1}^{(0)}(v_{0,1}^{(0)}), \tag{4A.56}$$

$$(F_{0,1}^{(0)}(u))^v = \exp\{ - v\psi_{0,1}^{(0)-1}(u)\}. \tag{4A.57}$$

According to the definition of LSHAC, the copulas emanated from $C_{0,1}^{(0)}$ with generator $\psi_{0,1}^{(0)}$ are

$$\left\{ C_{s_0,j_1}^{(1)} | s_0 = 1, j_1 = 1, \ldots, D_{s_0}^{(1)} \right\}. \tag{4A.58}$$

Consequently, we have

$$\begin{aligned} C_{0,1}^{(0)} &= C_{0,1}^{(0)}(C_{s_0,1}^{(1)}, \ldots, C_{s_0,D_{s_0}^{(1)}}^{(1)}) \\ &= \psi_{0,1}^{(0)} \left(\sum_{j_1=1}^{D_{s_0}^{(1)}} \psi_{0,1}^{(0)-1}(C_{s_0,j_1}^{(1)}) \right) \end{aligned} \tag{4A.59}$$

Then, using (4A.56) and (4A.57) yields

$$\begin{aligned} C(u_1, u_2, \cdots, u_d) &= \int_0^\infty \prod_{j_1=1}^{D_{s_0}^{(1)}} \exp\{ - v_{0,1}^{(0)} \psi_{0,1}^{(0)-1}(C_{s_0,j_1}^{(1)}) \} dG_{0,1}^{(0)}(v_{0,1}^{(0)}), \\ &= \int_0^\infty \prod_{j_1=1}^{D_{s_0}^{(1)}} \left(F_{0,1}^{(0)}(C_{s_0,j_1}^{(1)}) \right)^{v_{0,1}^{(0)}} dG_{0,1}^{(0)}(v_{0,1}^{(0)}), \\ &= \int_0^\infty \prod_{j_1=1}^{D_{s_0}^{(1)}} \left(F_{0,1}^{(0)}(C_{s_0,j_1}^{(1)}) \right)^{v_{0,1}^{(0)}} \left(dG \right)_{j_0}. \end{aligned} \tag{4A.60}$$

Similarly, for level $y=1$, the copulas emanated from $C_{s_0,j_1}^{(1)}$ with generator $\psi_{s_0,j_1}^{(1)}$ are

$$\left\{ C_{s_1,j_2}^{(2)} | s_1 = \left(\sum_{m=1}^{s_0-1} D_m^{(1)} \right) \mathbb{I}_{\{s_0>1\}} + j_1, j_2 = 1, \ldots, D_{s_1}^{(2)} \right\}. \tag{4A.61}$$

Since $s_0 = 0$, it is equivalent to

$$\left\{ C_{s_1,j_2}^{(2)} | s_1 = j_1, j_2 = 1, \ldots, D_{s_1}^{(2)} \right\}. \tag{4A.62}$$

Therefore, we have

$$C^{(1)}_{s_0,j_1} = C^{(1)}_{s_0,j_1}(\, C^{(2)}_{s_1,1},\ldots,C^{(2)}_{s_1,D^{(2)}_{s_1}}\,) = \psi^{(1)}_{s_0,j_1}(\sum_{j_2=1}^{D^{(2)}_{s_1}} \psi^{(1)-1}_{s_0,j_1}(C^{(2)}_{s_1,j_2})). \tag{4A.63}$$

Let

$$\tilde{\psi}^{(1)}_{s_0,j_1}(u;v^{(0)}_{0,1}) = \left(F^{(0)}_{0,1}(\psi^{(1)}_{s_0,j_1}(u))\right)^{v^{(0)}_{0,1}} = \exp\{-v^{(0)}_{0,1}\psi^{(0)-1}_{0,1}\circ\psi^{(1)}_{s_0,j_1}(u)\}, \tag{4A.64}$$

and

$$\widetilde{\Psi}^{(1)}_{s_0,j_1}(u) = \psi^{(0)-1}_{0,1}\circ\psi^{(1)}_{s_0,j_1} \tag{4A.65}$$

be the Laplace exponent of a Lévy subordinator, $X^{(1)}_{s_0,j_1}$, with c.d.f. $\widetilde{G}^{(1)}_{s_0,j_1}$. According to the property of Laplace exponent of a Lévy subordinator expressed in (4.2),

$$\begin{aligned}\tilde{\psi}^{(1)}_{s_0,j_1}(u;v^{(0)}_{0,1}) &= \exp\{-v^{(0)}_{0,1}\psi^{(0)-1}_{0,1}\circ\psi^{(1)}_{s_0,j_1}(u)\},\\ &= \exp\{-v^{(0)}_{0,1}\widetilde{\Psi}^{(1)}_{s_0,j_1}(u)\},\\ &= E(\exp\{-uX^{(1)}_{J_0,j_1}(v^{(0)}_{0,1})\}),\\ &= \int_0^\infty \exp\{-uv^{(1)}_{s_0,j_1}(v^{(0)}_{0,1})\}d\widetilde{G}^{(1)}_{s_0,j_1}(v^{(1)}_{s_0,j_1};v^{(0)}_{0,1}).\end{aligned} \tag{4A.66}$$

As proved in Theorem 2.1 of Hering et al. (2010), derivative of $\psi^{(0)-1}_{0,1}\circ\psi^{(1)}_{s_0,j_1}$ defined according to (4A.65) is c.m.. According to Joe (1997) and McNeil (2008), $\tilde{\psi}^{(1)}_{s_0,j_1}(u)$ is a LT-AC generator. As a result, we can rewrite (4A.60) according to (4A.63)-(4A.64),

$$\begin{aligned}&C(u_1,\cdots,u_d)\\ =&\int_0^\infty \prod_{j_1=1}^{D^{(1)}_{s_0}}\left(F^{(0)}_{0,1}(C^{(1)}_{s_0,j_1})\right)^{v^{(0)}_{0,1}} dG^{(0)}_{0,1}(v^{(0)}_{0,1}),\\ =&\int_0^\infty \prod_{j_1=1}^{D^{(1)}_{s_0}}\tilde{\psi}^{(1)}_{s_0,j_1}\left(\sum_{j_2=1}^{D^{(2)}_{s_1}}\psi^{(1)-1}_{s_0,j_1}(C^{(2)}_{s_1,j_2});v^{(0)}_{0,1}\right)dG^{(0)}_{0,1}(v^{(0)}_{0,1}),\\ =&\int_0^\infty \prod_{j_1=1}^{D^{(1)}_{s_0}}\int_0^\infty \exp\left(-v^{(1)}_{s_0,j_1}(\sum_{j_2=1}^{D^{(2)}_{s_1}}\psi^{(1)-1}_{s_0,j_1}(C^{(2)}_{s_1,j_2}))\right)d\widetilde{G}^{(1)}_{s_0,j_1}(v^{(1)}_{s_0,j_1};v^{(0)}_{0,1})dG^{(0)}_{0,1}(v^{(0)}_{0,1}),\end{aligned} \tag{4A.67}$$

Similarly, let $F^{(1)}_{s_0,j_1}(u)$ satisfy

$$\Big(F^{(1)}_{s_0,j_1}(u)\Big)^{v} \;=\; \exp\big(-v\psi^{(1)-1}_{s_0,j_1}(u)\big), \tag{4A.68}$$

then we have

$$\begin{aligned}
&C(u_1,\cdots,u_d)\\
=\;&\int_0^\infty\prod_{j_1=1}^{D^{(1)}_{s_0}}\int_0^\infty\prod_{j_2=1}^{D^{(2)}_{s_1}}\Big(F^{(1)}_{s_0,j_1}(C^{(2)}_{s_1,j_2})\Big)^{v^{(1)}_{s_0,j_1}}d\widetilde{G}^{(1)}_{s_0,j_1}(v^{(1)}_{s_0,j_1};v^{(0)}_{0,1})dG^{(0)}_{0,1}(v^{(0)}_{0,1}),\\
=\;&\int_0^\infty\prod_{j_1=1}^{D^{(1)}_{s_0}}\int_0^\infty\prod_{j_2=1}^{D^{(2)}_{s_1}}\Big(F^{(1)}_{s_0,j_1}(C^{(2)}_{s_1,j_2})\Big)^{v^{(1)}_{s_0,j_1}}\big(dG\big)^{(1)}_{j_1}.
\end{aligned} \tag{4A.69}$$

Therefore, (4.9) is satisfied at level $y=1$. Now let us assume at level $y:\ 0\le y\le l-2\,(l\ge 2)$, the following equation holds

$$\begin{aligned}
&C(u_1,\cdots,u_d)\\
=\;&\int_0^\infty\prod_{j_1=1}^{D^{(1)}_{s_0}}\int_0^\infty\prod_{j_2=1}^{D^{(2)}_{s_1}}\cdots\int_0^\infty\prod_{j_{y+1}=1}^{D^{(y+1)}_{s_y}}\Big(F^{(y)}_{s_{y-1},j_y}(C^{(l-1)}_{s_y,j_{y+1}})\Big)^{v^{(y)}_{s_{y-1},j_y}}\big(dG\big)^{(y)}_{j_y}.
\end{aligned} \tag{4A.70}$$

For notation consistency we let $s_{-1}=0, j_0=1$. Then at level $y+1$, the copulas emanated from $C^{(y+1)}_{s_y,j_{y+1}}$ with generator $\psi^{(y+1)}_{s_y,j_{y+1}}$ are

$$\Big\{C^{(y+2)}_{s_{y+1},j_{y+2}}|s_{y+1}=\big(\sum_{m=1}^{s_y-1}D^{(y+1)}_m\big)\mathbb{I}_{\{s_{y-1}>1\}}+j_{y+1},j_{y+2}=1,\ldots,D^{(y+2)}_{s_{y+1}}\Big\}. \tag{4A.71}$$

As a result, we have

$$C^{(y+1)}_{s_y,j_{y+1}}=\psi^{(y+1)}_{s_y,j_{y+1}}\Bigg(\sum_{j_{y+2}=1}^{D^{(y+2)}_{s_{y+1}}}\psi^{(y+1)-1}_{s_y,j_{y+1}}(C^{(y+2)}_{s_{y+1},j_{y+2}})\Bigg). \tag{4A.72}$$

Let

$$\begin{aligned}
\tilde{\psi}^{(y+1)}_{s_y,j_{y+1}}(u;v^{(y)}_{s_{y-1},j_y}) &= \big(F^{(y)}_{s_{y-1},j_y}(\psi^{(y+1)}_{s_y,j_{y+1}}(u))\big)^{v^{(y)}_{s_{y-1},j_y}} &(4A.73)\\
&= \exp\big\{-v^{(y)}_{s_{y-1},j_y}\psi^{(y)-1}_{s_{y-1},j_y}\circ\psi^{(y+1)}_{s_y,j_{y+1}}(u)\big\} &(4A.74)
\end{aligned}$$

and

$$\widetilde{\Psi}^{(y+1)}_{s_y,j_{y+1}}(u) = \psi^{(y)}_{s_{y-1},j_y} \circ \psi^{(y+1)}_{s_y,j_{y+1}}(u) \tag{4A.75}$$

be the Laplace exponent of a Lévy subordinator, denoted as $X^{(y+1)}_{s_y,j_{y+1}}$, with c.d.f. $\widetilde{G}s_y, j_{y+1}{}^{(y+1)}$. Consequently, substituting (4A.75) into (4A.74) yields

$$\begin{aligned}
\tilde{\psi}^{(y+1)}_{s_y,j_{y+1}}(u; v^{(y)}_{s_{y-1},j_y}) &= \exp\{ - v^{(y)}_{s_{y-1},j_y} \widetilde{\Psi}^{(y+1)}_{s_y,j_{y+1}}(u)\} \\
&= E\Big(\exp\{-uX^{(y+1)}_{s_y,j_{y+1}}(v^{(y)}_{s_{y-1},j_y})\}\Big) \\
&= \int_0^\infty \exp\{-uv^{(y+1)}_{s_y,j_{y+1}}\} d\widetilde{G}^{(y+1)}_{s_y,j_{y+1}}(v^{(y+1)}_{s_y,j_{y+1}}; v^{(y)}_{s_{y-1},j_y})
\end{aligned} \tag{4A.76}$$

Similarly, $\tilde{\psi}^{(y+1)}_{s_y,j_{y+1}}$ is a LT-AC generator. Therefore, (4A.70) can be rewritten as

$$C(u_1, \cdots, u_d) =$$

$$\int_0^\infty \prod_{j_1=1}^{D^{(1)}_{s_0}} \int_0^\infty \prod_{j_2=1}^{D^{(2)}_{s_1}} \cdots \int_0^\infty \prod_{j_{y+1}=1}^{D^{(y+1)}_{sy}} \tilde{\psi}^{(y+1)}_{s_y,j_{y+1}} \Bigg(\sum_{j_{y+1}=1}^{D^{(y+1)}_{sy}} \psi^{(y+1)-1}_{s_y,j_{y+1}}(C^{(y+2)}_{s_{y+1},j_{y+2}}; v^{(y)}_{s_{y-1},j_y})) \Bigg) \big(dG\big)^{(y)}_{j_y} \tag{4A.77}$$

which is equivalent to

$$\int_0^\infty \prod_{j_1=1}^{D^{(1)}_{s_0}} \int_0^\infty \prod_{j_2=1}^{D^{(2)}_{s_1}} \cdots \int_0^\infty \prod_{j_{y+1}=1}^{D^{(y+1)}_{sy}} \int_0^\infty \exp\Bigg(- v^{(y+1)-1}_{s_y,j_{y+1}} \sum_{j_{y+2}=1}^{D^{(y+2)}_{s_{y+1}}} \psi^{(y+1)-1}_{s_y,j_{y+1}}(C^{(y+2)}_{s_{y+1},j_{y+2}}))\Bigg) \tag{4A.78}$$

$$d\widetilde{G}_{s_y,j_{y+1}}(v^{(y+1)}_{s_y,j_{y+1}}; v^{(y)}_{s_{y-1},j_y}) \big(dG\big)^{(y)}_{j_y}. \tag{4A.79}$$

Let $F^{(y+1)}_{s_y,j_{y+1}}$ satisfy

$$(F^{(y+1)}_{s_y,j_{y+1}})^v = \exp\big(- v\psi^{(y+1)-1}_{s_y,j_{y+1}}(u)\big). \tag{4A.80}$$

Then we have

$$C(u_1, \cdots, u_d) =$$

$$\int_0^\infty \prod_{j_1=1}^{D^{(1)}_{s_0}} \int_0^\infty \prod_{j_2=1}^{D^{(2)}_{s_1}} \cdots \int_0^\infty \prod_{j_{y+1}=1}^{D^{(y+1)}_{sy}} \int_0^\infty \prod_{j_{y+2}=1}^{D^{(y+2)}_{s_{y+1}}} F^{(y+1)}_{s_y,j_{y+1}}(C^{(y+2)}_{s_{y+1},j_{y+2}})) \big(dG\big)^{(y+1)}_{j_{y+1}}, \tag{4A.81}$$

which means that (4.9) is satisfied at level $y+1$. Therefore, by mathematical induction, (4.9) holds for all $l = 1, \ldots, L-1$, and this completes the proof of Theorem 4.2.1.

Chapter 5

Weighted Distribution Premium Principle and Agricultural Reinsurance Pricing

5.1 Introduction

A scientific ratemaking methodology is fundamental and essential for producing a sustainable risk management solution for various stakeholders, including producers, insurers, reinsurers, and government. The absence of actuarially sound methods to achieve accurate fair premium rate limits the development of agricultural insurance and reinsurance program (Ozaki et al., 2008). Especially for heavily subsidized agricultural insurance programs, such as crop insurance programs in U.S. and Canada, a slight adjustment in premium rates may significantly change the subsidies and may have a substantial impact on taxpayers. Therefore, a strategic pricing framework will benefit the agricultural insurance industry and ensure its sustainability over the long term.

The pricing of agricultural insurance products is particularly challenging due to its unique features as compared to most other commercial lines of P & C insurance (Porth et al., 2014a). More specifically, agricultural insurance pricing suffers from the shortness of data with at most several decades historical (annual) loss observations. In addition, agricultural losses tend to be highly spatially correlated and at times in large magnitudes because of adverse natural hazards. Moreover, the underlying structural factors of the loss experiences could change overtime simply due to program changes, technological development, farming practice, etc.

The unique challenges pertaining to the agricultural insurance ratemaking suggest the importance of actuarially sound pricing methodologies. However, until recently, the agricultural insurance premium rates are based on simple average of historical loss-cost ratio (LCR) experience (Borman et al., 2013). There is limited literature that discuss how to scientifically weight historical losses, from different dimensions, to accommodate and adjust various risk factors. Woodard (2014) constructs a conditional Weibull distribution model that integrates the weather variables and technology evolutions into crop yields explicitly. Porth et al. (2014b) propose a "liability weighted" LCR to aggregate historical loss data and introduce a modified credibility model to weight the loss experiences from different geographical regions to enhance the reinsurance pricing. Borman et al. (2013) propose to incorporate weights into the data in order to reflect program changes and weather patterns.

This chapter discusses the ratemaking and risk management of agricultural (re)insurance from the premium principle perspective. To the best of our knowledge, this is the first work to formally introduce actuarial premium principles to agricultural ratemaking. Loosely speaking, a premium principle is a pricing rule that attaches a premium to insurance risks, and it is the core of actuarial insurance ratemaking (Wang, 1996; Young, 2004) in reflecting the underlying risk. In this chapter, we discuss some popular premium principles in actuarial science and compare the features of the different principles in agricultural ratemaking. In particular, we propose a new premium principle, which we denote as the multivariate weighted distribution premium principle, to facilitate the weighting of auxiliary variables into the pricing framework. This idea is stimulated by some empirical pricing results in agricultural ratemaking, and the Probability Proportional to size (PPS) sampling method widely used in statistical sampling. Based on the work by Bühlmann (1980), Furman and Zitikis (2008a), Patil et al. (1986), Rao (1965), and Wang (1995), we derive some useful properties of the premium based on multivariate weighted random variables. In addition, the economic premium principle discussed in Bühlmann (1980), and the *Esscher's principle* can be shown to be special cases of our proposed framework.

To test the advantages and usefulness of our proposed multivariate weighted premium principle relative to other well-known premium principles, an empirical study using reinsurance experience (from year 2001-2011) in Manitoba is conducted. Our empirical results highlight the importance of incorporating auxiliary information such as liabilities and other macroeconomic variables and that the proposed new premium principle is able to assign higher loading to more risky contract layers and achieve better sustainable long-run profits.

Main Contributions in This Chapter

This chapter contributes to the literature from the following perspectives. First, a new

premium principle based on multivariate distribution is proposed. This premium principle extends the work by Furman and Zitikis (Furman and Zitikis, 2008a;b). Some desirable properties of the proposed multivariate weighted premium principle are presented in this paper, including positive risk loading and no ripoff, no unjustified risk-loading, linearity, additivity, and stochastic dominance preserving. The multivariate weighted premium allocation is also additive among layers, making it very appealing to insurance layer pricing. It includes Wang's premium principle (Wang, 1995; 1996), univariate weighted premium principle (Furman and Zitikis, 2008a), Esshcher's premium principle, and economic premium principle (Bühlmann, 1980) as special cases. In particular, it is shown that the multivariate weighted premium principle has increasing relative risk loading with appropriately chosen auxiliary variables, while the univariate weighted premium principle has constant relative risk loading. This chapter, for the first time, introduces the concept of premium principle into agricultural ratemaking, and provides empirical evidence for the necessity of using auxiliary variables to enhance agricultural insurance ratemaking.

The rest of this chapter proceeds as follows. Section 5.2 gives a brief introduction of premium principles. Section 5.3 introduces our proposed multivariate weighted premium principle and derives some analytical results for a number of parametric models. Section 5.4 introduces some desirable properties of the multivariate weighted premium principle. Section 5.5 discusses the relationship of the multivariate weighted premium principle and Wang's premium principle. Section 5.6 provides some remarks about how to select appropriate auxiliary variables in applying multivariate weighted premium principle. Section 5.7 conducts an empirical analysis on the agricultural reinsurance ratemaking. Section 5.9 concludes the chapter and appendix collects the proofs.

5.2 Premium Principles

The premium principles are core to actuarial pricing. Let X be a non-negative loss random variable with the cumulative distribution function $F_X(x)$, decumulative distribution function (or survival function) $S_X(x)$ and density function $f_X(x)$. A premium principle is then defined as a functional Π assigned to the insurance risk X. We also denote the collection of all nonnegative random variables as a set $\mathfrak{X}$ on the probability space $(\Omega, \mathcal{F}, P)$. The quest for an appropriate premium principle has been an active area of research in actuarial science and it is also important for various applications including the agricultural ratemaking. An actuarially sound premium principle needs to satisfy some desirable properties. The list of these properties can be very long and we list a few standard and important ones below. For

an inventory of premium principles properties see Young (2004).

1. Positive risk loading: $\Pi(X) \geq \mathsf{E}(X)$ for all $X \in \mathcal{X}$. This property requires that the premium is no less than the expected payout of the risk (net premium or risk premium) in exchange for insuring the risk.

2. No unjustified risk-loading: If risk X is degenerated, namely, there exists constant c such that $\mathsf{P}(X = c) \equiv 1$, then $\Pi(X) = c$. Therefore, if a risk pays out a constant c for certain, the insurer should charge no risk loading and the premium should just be its certainty loss amount.

3. No ripoff: $\Pi(X) \leq \text{esssup}(X)$ for all $X \in \mathcal{X}$. This means that the insurer should not charge higher than the maximum value the risk may get.

4. Translation invariance: $\Pi(X + a) = \Pi(X) + a$ for all $X \in \mathcal{X}$ and $a \geq 0$. If a risk X is increased by a fixed number a, the premium for risk $X + a$ should be the original premium plus a.

5. Scale invariance: $\Pi(aX) = a\Pi(X)$ for all $X \in \mathcal{X}$ and $a \geq 0$. This property is also known as homogeneity of degree one in the economic literature to preclude arbitrage opportunities. For example, the premium for $2X$ should be equal to the premiums of two insurance policies for the risk X, otherwise, there is a chance for arbitrage.

   Combining Property 4 and Property 5 implies linearity.

6. Subadditivity: $\Pi(X + Y) \leq \Pi(X) + \Pi(Y)$ for all $X \in \mathcal{X}$ and $Y \in \mathcal{X}$.

7. First stochastic dominance (FSD) preserving: If $S_X(x) \leq S_Y(x)$ for all $x \geq 0$, then $\Pi(X) \leq \Pi(Y)$.

8. Stop-loss (SL) ordering preserving: If $\mathsf{E}\big[(X - d)_+\big] \leq \mathsf{E}\big[(Y - d)_+\big]$ for all $d > 0$, then $\Pi(X) \leq \Pi(Y)$.

Some most commonly used premium principles are listed as follows:

a. Expectation Premium Principle: $\Pi^e(X) = (1 + \theta)\mathsf{E}(X)$, where $\theta > 0$. Due to its simplicity, this is the most widely used premium principle in agricultural insurance ratemaking and in all other types of insurances.

b. Standard Deviation Premium Principle: $\Pi^{sd}(X) = \mathsf{E}(X) + \theta\sqrt{\mathsf{Var}(X)}$, where $\theta > 0$. This premium principle incorporates a risk loading that is proportional to the standard deviation of the risk. While widely use in general Property & Casualty (P&C) insurance, this premium principle has received little attention in agricultural insurance, except for the work by Porth et al. (2013), which analyzes the optimal

reinsurance contract structure that minimizes the total risk exposure of an insurer using private reinsurance loss experience for Manitoba and estimated that the optimal θ is 0.1861.

c. Esscher Premium Principle: $\Pi^{ess} = \dfrac{\mathsf{E}(Xe^{\theta X})}{\mathsf{E}(e^{\theta X})}$, where $\theta > 0$. This premium principle, which is based on the *Esscher Transform*, is widely used in option pricing especially in incomplete markets (See, for example Bühlmann et al., 1996; Gerber and Shiu, 1994; Hubalek and Neilsen, 2006). It is interesting to note that it is a special case of the *Equilibrium Premium Principle* proposed by Bühlmann (1980). Additionally, a more general form of Esscher premium is referred to as the *Exponential Tilting Premium Principle* (Heilmann, 1989; Kamps, 1998).

d. Distortion Premium Principle: For any increasing concave function $g : [0,1] \mapsto [0,1]$ with $g(0) = 0, g(1) = 1$, the premium is calculated as $\Pi^s(X) = \int_0^\infty g\big(S_X(u)\big)du$ (Wang, 1996; Wang et al., 1997). The function g is called *distortion* function and $g\big(S_X(u)\big)$ is called distorted probability. This premium principle is constructed based on a transformation of the decumulative distribution function $S_X(x)$. A popular special case of this premium class is called *Proportional Hazards Premium Principle* (Wang, 1995).

To facilitate reweighing historical losses with auxiliary variables, in this chapter, we first propose a new premium principle, discuss its desirable properties, and then apply it to agricultural reinsurance pricing and compare the pricing results with other premium principles.

5.3 Multivariate Weighted Premium (MWP)

5.3.1 Definitions

A general definition of multivariate weighted distribution can be defined according to Navarro et al. (2006).

Definition 5.3.1 (Navarro et al. (2006))**.** *The density of multivariate weighted distribution associated with random vector* $\boldsymbol{X} = (X_1, \ldots, X_p)$ *and weighting function* $w(\boldsymbol{x}) =$

$w(x_1, \ldots, x_p)$ *is*

$$\begin{aligned} f_{\boldsymbol{X}^{\boldsymbol{w}}}(\boldsymbol{x}) &= f_{\boldsymbol{X}^{\boldsymbol{w}}}(x_1, \ldots, x_p) \\ &= \frac{w(x_1, \ldots, x_p)}{E\big(w(X_1, \ldots, X_p)\big)} f_{\boldsymbol{X}}(x_1, \ldots, x_p) \\ &= \frac{w(\boldsymbol{x})}{E\big(w(\boldsymbol{X})\big)} f_{\boldsymbol{X}}(\boldsymbol{x}). \end{aligned} \tag{5.1}$$

For the rest of this chapter, we consider X as a non-negative random loss and $\boldsymbol{Y} = (Y_1, \ldots, Y_p)$ is some p-dimensional random vector with $Y_i \in \mathcal{X}$, where $i = 1, \ldots, p$. Then the multivariate weighted distribution with respect to $(X, \boldsymbol{Y})$ is defined as follows:

Definition 5.3.2. $(X^w, \boldsymbol{Y}^{\boldsymbol{w}})$ *is a vector of weighted random variables associated with (i.e., defined base on)* $(X, \boldsymbol{Y})$ *and* $w(\boldsymbol{y}) = w(y_1, \ldots, y_p)$*, with joint density*

$$\begin{aligned} f_{X^w, \boldsymbol{Y}^{\boldsymbol{w}}}(x, \boldsymbol{y}) &= \frac{w(y_1, \ldots, y_p)}{E[w(Y_1, \ldots, Y_p)]} f_{X, \boldsymbol{Y}}(x, y_1, \ldots, y_p) \\ &= \frac{w(\boldsymbol{y})}{E[w(\boldsymbol{Y})]} f_{X, \boldsymbol{Y}}(x, \boldsymbol{y}) \end{aligned} \tag{5.2}$$

According to these definitions, the density of the weighted random loss, X^w, can be expressed as (see also Kocherlakota, 1995; Mahfoud and Patil, 1982; Navarro et al., 2006)

$$\begin{aligned} f_{X^w}(x) &= \int_{\boldsymbol{y} \in \mathbb{R}_+^p} f_{X^w, \boldsymbol{Y}^{\boldsymbol{w}}}(x, \boldsymbol{y}) d\boldsymbol{y} \\ &= \frac{\displaystyle\int_{\boldsymbol{y} \in \mathbb{R}_+^p} w(\boldsymbol{y}) f_{\boldsymbol{Y}|X}(\boldsymbol{y}|x) f_X(x) d\boldsymbol{y}}{\mathsf{E}\big(w(\boldsymbol{Y})\big)} \\ &= \frac{\mathsf{E}\big(w(\boldsymbol{Y})|X = x\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)} f_X(x) \end{aligned} \tag{5.3}$$

We use $\psi(x)$ to denote the following ratio of two expectations:

$$\psi(x) = \frac{\mathsf{E}\big(w(\boldsymbol{Y})|X = x\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)}. \tag{5.4}$$

Remark 5.3.1. *It is of interest to note that* $\psi(x)$ *is the Radon-Nikodym derivative of the two random variables* X *and* X^w*. Given any loss random variable* X *with density* $f_X(x)$*, we can define another random variable* X^w *with density* $f_{X^w}(s)$ *according to equation* (5.3)*.*

Remark 5.3.2. *It should also be emphasized that if* $\psi(x)$ *is an increasing function with*

respect to x, the loss size (i.e., severity of the loss), more weight will be assigned to the adverse events, which satisfies the empirical pricing requirements.

By defining the mapping: $\mathfrak{T}^w : (X, \boldsymbol{Y}) \to X^w$ as the **weighted density transform**, where the superscript w refers to the weighing function w, we are now ready to define a new premium calculation principle that is based on the multivariate weighted distribution.

Definition 5.3.3. *For a risk X with density $f_X(x)$, the multivariate weighted premium associated with some random vector $\boldsymbol{Y}$ and positive weighting function w is defined as*

$$\Pi^w(X, \boldsymbol{Y}) \quad = \quad \mathsf{E}\big(\mathfrak{T}^w(X, \boldsymbol{Y})\big). \tag{5.5}$$

Proposition 5.3.1. *The multivariate weighted premium can be expressed as*

$$\Pi^w(X, \boldsymbol{Y}) \quad = \quad \frac{\mathsf{E}\big(w(\boldsymbol{Y})X\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)}. \tag{5.6}$$

Proof. According to Definition 5.3.3 and Equation (5.3), we have

$$\begin{aligned}
\Pi^w(X, \boldsymbol{Y}) &= \mathsf{E}\big(\mathfrak{T}^w(X, \boldsymbol{Y})\big) \\
&= \int_{x \in \mathbb{R}_+} x \frac{\mathsf{E}\big(w(\boldsymbol{Y})|X = x\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)} f_{\boldsymbol{X}}(x)dx \\
&= \frac{\mathsf{E}\big(X\mathsf{E}[w(\boldsymbol{Y})|X]\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)} \\
&= \frac{\mathsf{E}\big(w(\boldsymbol{Y})X\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)}.
\end{aligned}$$

□

Remark 5.3.3. *In empirical pricing, $\boldsymbol{Y}$ may be some auxiliary random vector variables of relevance to ratemaking for insurers and reinsurers. In addition, $\boldsymbol{Y}$ needs not be different from X. In fact, when $\boldsymbol{Y} = X$, the multivariate weighted premium principle in Definition 5.3.3 degenerates to the weighted premium discussed in Furman and Zitikis (2008a). The univariate case and the bivariate case will be further discussed in Section 5.3.2 with some examples of special weighting functions and distributions.*

5.3.2 Examples of Multivariate Weighted Premium

Univariate Weighting

1. **Power weighting ($w(u) = u^{\lambda}$)**

   In the case of power weighting, the univariate weighted premium can be written as

$$\Pi^w(X) = \frac{\mathsf{E}(X^{\lambda+1})}{\mathsf{E}(X^{\lambda})}. \tag{5.7}$$

   Example 5.3.1. If X has Gamma distribution with parameter $GAM(a, b)$ with the corresponding p.d.f. $f_X(x; a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}$, the power weighted density transform $\mathfrak{T}^w(X)$ also has a gamma distribution with parameters $GAM(a + \lambda, b)$. According to Equation (5.7), the univariate weighted premium $\Pi^w(X)$ is

$$\Pi^w(X) = \frac{a + \lambda}{b}.$$

   Example 5.3.2. If X has a log-normal distribution with parameter $LogN(\mu, \sigma^2)$ and the corresponding p.d.f. $f_X(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\frac{\log(x) - \mu}{\sigma})^2\right)$, the power weighted density transform $\mathfrak{T}^w(X)$ also has a log-normal distribution with parameters $LogN(\mu + \lambda\sigma^2, \sigma^2)$. The univariate weighted premium$\Pi^w(X)$ is

$$\Pi^w(X) = \exp\big(\mu + (\lambda + \frac{1}{2})\sigma^2\big).$$

2. **Quantile weighting ($w(u) = \mathbb{I}_{\{u > u_\lambda\}}$)**

   In the case of quantile weighting, the univariate weighted premium can be written as

$$\Pi^w(X) = \frac{\mathsf{E}(X\mathbb{I}_{\{X > x_\lambda\}})}{\mathsf{E}(\mathbb{I}_{\{X > x_\lambda\}})}, \tag{5.8}$$

   where $\mathbb{I}_{\{u \in A\}}$ is the indicator function which is 1 when event A happens, while 0 otherwise. We can see that if u_λ is chosen to be the λ-quantile of X, then the univariate weighted premium becomes the conditional tail expectation (CTE) at level λ.

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

$$
\begin{aligned}
f_I(s,t) &= \mathbb{I}_{\{s\le t\}}, && (5.9)\\
g_I(s,t) &= 1 - f_I(s,t), && (5.10)\\
\gamma(s,t) &= \int_0^t u^{s-1}e^{-u}du, && (5.11)
\end{aligned}
$$

where the function $\gamma(s,t)$ is called lower incomplete gamma function.

Example 5.3.3. If X has A Gamma distribution with parameter $GAM(a,b)$, we have

$$
\begin{aligned}
\mathsf{E}(\mathbb{I}_{\{X>x_\lambda\}}X) &= \mathsf{E}(g_I(X,x_\lambda)X)\\
&= \mathsf{E}(X) - \int_0^{x_\lambda} x\frac{b^a}{\Gamma(a)}x^{a-1}e^{-bx}dx\\
&\overset{u=bx}{=} \frac{a}{b}\left(1 - \frac{1}{\Gamma(a+1)}\int_0^{bx_\lambda} u^a e^{-u}du\right)\\
&= \frac{a}{b}\left(1 - \frac{\gamma(a+1,bx_\lambda)}{\Gamma(a+1)}\right), && (5.12)\\
\mathsf{E}(\mathbb{I}_{\{X>x_\lambda\}}) &= 1 - \frac{\gamma(a,bx_\lambda)}{\Gamma(a)}. && (5.13)
\end{aligned}
$$

Therefore, according to Equation (5.8), the univariate quantile weighted premium is expressed as:

$$
\Pi^w(X) = \frac{a}{b}\left(\frac{1 - \frac{\gamma(a+1,bx_\lambda)}{\Gamma(a+1)}}{1 - \frac{\gamma(a,bx_\lambda)}{\Gamma(a)}}\right).
$$

Example 5.3.4. If X has a log-normal distribution with parameter $LogN(\mu,\sigma^2)$, we have

$$
\begin{aligned}
\mathsf{E}(\mathbb{I}_{\{X>x_\lambda\}}X) &= \mathsf{E}(g_I(X,x_\lambda)X)\\
&= \int_{\log x_\lambda}^{\infty} e^x \frac{1}{\sqrt{2\pi}\sigma}\exp\big(-\frac{(x-\mu)^2}{2\sigma^2}\big)dx\\
&\overset{\substack{u=\frac{x-\mu}{\sigma}\\ \mu_\lambda=\frac{\log x_\lambda - u}{\sigma}}}{=} e^{\mu+\frac{1}{2}\sigma^2}\big(1-\Phi(\mu_\lambda-\sigma)\big), && (5.14)\\
\mathsf{E}(\mathbb{I}_{\{X>x_\lambda\}}) &= 1-\Phi(\mu_\lambda), && (5.15)
\end{aligned}
$$

where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution. Therefore, according to

Equation (5.8), the univariate quantile weighted premium is expressed as:

$$\Pi^w(X) \quad = \quad e^{\mu+\frac{1}{2}\sigma^2}\left(\frac{1-\Phi(\mu_\lambda-\sigma)}{1-\Phi(\mu_\lambda)}\right).$$

3. **Exponential weighting ($w(u) = e^{\lambda u}$)**

   In the case of exponential weighting, the univariate weighted premium can be written as

$$\Pi^w(X) \quad = \quad \frac{\mathsf{E}(Xe^{\lambda X})}{\mathsf{E}(e^{\lambda X})}, \tag{5.16}$$

   which is the well known Esscher's premium.

   Example 5.3.5. If X has A Gamma distribution with parameter $GAM(a,b)$, the exponential weighted density transform $\mathfrak{T}^w(X)$ also has a gamma distribution with parameters $GAM(a, b-\lambda)$. When $b > \lambda$, according to Equation (5.16), the univariate exponential weighted premium $\Pi^w(X)$ is

$$\Pi^w(X) \quad = \quad \frac{a}{b-\lambda}.$$

The univariate weighted premium examples are summarized in Table 5.1.

Table 5.1: Univariate Weighted Premium Examples

Weighting Function	$\boldsymbol{GAM(a,b)}$	$\boldsymbol{LogN(\mu,\sigma^2)}$
Power ($w(u) = u^\lambda$)	$\frac{a+\lambda}{b}$	$\exp\big(\mu+(\lambda+\frac{1}{2})\sigma^2\big)$
Quantile ($w(u) = \mathbb{I}_{\{u>u_\lambda\}}$)	$\frac{a}{b}\left(\frac{1-\frac{\gamma(a+1,bx_\lambda)}{\Gamma(a+1)}}{1-\frac{\gamma(a,bx_\lambda)}{\Gamma(a)}}\right)$	$e^{\mu+\frac{1}{2}\sigma^2}\left(\frac{1-\Phi(\mu_\lambda-\sigma)}{1-\Phi(\mu_\lambda)}\right)$ [a]
Exponential ($w(u) = e^{\lambda u}$)	$\frac{a}{b-\lambda}$	- [b]

[a] $\mu_\lambda = \frac{\sigma\log x_\lambda - x + \mu}{\sigma^2}$.

[b] Exponential weighted premium for log-normal distribution is not defined since the moment generating function (m.g.f.) of the log-normal distribution is not defined for any positive value of the argument.

Bivariate Weighting

1. **Power weighting ($w(u) = u^\lambda$)**

   In the case of power weighting, the bivariate weighted premium associated with Y can

be written as

$$\Pi^w(X,Y) \;=\; \frac{\mathsf{E}(XY^\lambda)}{\mathsf{E}(Y^\lambda)}. \tag{5.17}$$

Example 5.3.6. If Y has A Gamma distribution with parameter $GAM(\alpha,\beta)$ and conditional on $Y=y$, the random loss X has a Gamma distribution with parameter $GAM(\theta,y)$, it is easy to show that the marginal density of X is

$$f_X(x) \;=\; \frac{b^a\Gamma(\alpha+\theta)}{\Gamma(\theta)\Gamma(\alpha)}\frac{x^{\theta-1}}{(\beta+x)^{\alpha+\theta}},$$

which is generalized Pareto distribution with parameters $GDP(\alpha,\beta,\theta)$. We can show that the weighted density transform $\mathfrak{T}^w(X)$ also has a generalized Parato distribution with parameters $GDP(\alpha+\lambda,\beta,\theta)$. Therefore, using Equation (5.5), the bivariate power weighted premium $\Pi^w(X,Y)$ is

$$\Pi^w(X,Y) \;=\; \frac{\beta\theta}{\alpha+\lambda-1}.$$

Example 5.3.7. If the random loss X and another random variable have bivariate log-normal distribution with parameters $BLogN(\boldsymbol{\mu},\boldsymbol{\Sigma})$, where $\boldsymbol{\mu}=(\mu_x,\mu_y)'$, $\boldsymbol{\Sigma_{1,1}}=\sigma_x^2$, $\boldsymbol{\Sigma_{2,2}}=\sigma_y^2$, $\boldsymbol{\Sigma_{1,2}}=\boldsymbol{\Sigma_{2,1}}=\sigma_{xy}=\rho_{xy}\sigma_x\sigma_y$. To get the weighted premium, we let $X_1=\log(X), Y_1=\log(Y)$, which are bivariate normally distributed with covariance σ_{xy}. We also note that the conditional random variable $X_1|Y_1$ follows a normal distribution with parameters $N(\mu_x+\rho_{xy}\frac{\sigma_x}{\sigma_y}(Y_1-\mu_y),(1-\rho_{xy}^2)\sigma_x^2)$. Therefore,

$$\begin{aligned}
\mathsf{E}(XY^\lambda) &= \mathsf{E}(e^{X_1+\lambda Y_1}) \\
&= \mathsf{E}(e^{\lambda Y_1}\mathsf{E}(e^{X_1}|Y_1)) \\
&= \exp\Big\{\mu_x-\rho_{xy}\frac{\sigma_x}{\sigma_y}\mu_y+\frac{1}{2}\sigma_x^2(1-\rho_{xy}^2)\Big\}\mathsf{E}(e^{(\lambda+\rho_{xy}\frac{\sigma_x}{\sigma_y})Y_1}) \\
&\overset{\substack{\bar a=\exp\{\mu_x-\rho_{xy}\frac{\sigma_x}{\sigma_y}\mu_y+\frac{1}{2}\sigma_x^2(1-\rho_{xy}^2)\}\\ \bar b=\lambda+\rho_{xy}\frac{\sigma_x}{\sigma_y}}}{=} \bar a\exp\Big\{\mu_y\bar b+\frac{1}{2}\sigma_y^2\bar b^2\Big\}, \\
\mathsf{E}(Y^\lambda) &= \exp\Big\{\lambda\mu_y+\frac{1}{2}\sigma_y^2\lambda^2\Big\}.
\end{aligned} \tag{5.18}$$

Hence, according to Equation (5.17) the bivariate power weighted premium $\Pi^w(X,Y)$

is

$$\begin{aligned}\Pi^w(X,Y) &= \bar{a}\exp\Big\{\rho_{xy}\frac{\sigma_x}{\sigma_y}(\mu_y+\lambda\sigma_y^2+\frac{1}{2}\rho_{xy}\sigma_x\sigma_y)\Big\}\\ &= \exp\Big\{\mu_x+\frac{1}{2}\sigma_x^2+\lambda\rho_{xy}\sigma_x\sigma_y\Big\}.\end{aligned} \tag{5.19}$$

2. **Quantile weighting ($w(u)=\mathbb{I}_{\{u>u_\lambda\}}$)**

In the case of quantile weighting, the bivariate weighted premium can be written as

$$\Pi^w(X,Y) = \frac{\mathsf{E}(X\mathbb{I}_{\{Y>y_\lambda\}})}{\mathsf{E}(\mathbb{I}_{\{Y>y_\lambda\}})}. \tag{5.20}$$

Example 5.3.8. If random loss X and random variate Y have the Gamma distribution assumptions in Example 5.3.6, then we have

$$\begin{aligned}\mathsf{E}(\mathbb{I}_{\{Y>y_\lambda\}}X) &= \mathsf{E}(\mathbb{I}_{\{Y>y_\lambda\}}\mathsf{E}(X|Y))\\ &= \mathsf{E}(\frac{\theta}{Y})-\mathsf{E}(\frac{\theta}{Y}\mathbb{I}_{\{Y>y_\lambda\}})\\ &\overset{u=\beta y}{=} \frac{\theta\beta}{\alpha-1}-\frac{\theta\beta}{\alpha-1}\int_0^{\beta y_\lambda}\frac{u^{(\alpha-1)-1}}{\Gamma(\alpha-1)}e^{-u}du\\ &= \frac{\theta\beta}{\alpha-1}\bigg(1-\frac{\gamma(\alpha-1,\beta y_\lambda)}{\Gamma(\alpha-1)}\bigg),\end{aligned} \tag{5.21}$$

$$\mathsf{E}(\mathbb{I}_{\{Y>y_\lambda\}}) = 1-\frac{\gamma(\alpha,\beta y_\lambda)}{\Gamma(\alpha)}. \tag{5.22}$$

Therefore, according to Equation (5.20), the bivariate quantile weighted premium is expressed as:

$$\Pi^w(X) = \frac{\theta\beta}{\alpha-1}\bigg(\frac{1-\frac{\gamma(\alpha-1,\beta y_\lambda)}{\Gamma(\alpha-1)}}{1-\frac{\gamma(\alpha,\beta y_\lambda)}{\Gamma(\alpha)}}\bigg).$$

Example 5.3.9. If random loss X and random variate Y have the log-normal distribution assumptions in Example 5.3.7, then similarly by letting $X_1=\log(X), Y_1=\log(Y)$,

we have

$$\begin{aligned}
\mathsf{E}(\mathbb{I}_{\{Y>y_\lambda\}}X) &= \mathsf{E}(\mathbb{I}_{\{Y>y_\lambda\}}\mathsf{E}(X|Y)) \\
&= \exp\big\{\mu_x - \rho_{xy}\frac{\sigma_x}{\sigma_y}\mu_y + \frac{1}{2}\sigma_x^2(1-\rho_{xy}^2)\big\}\mathsf{E}(e^{\rho_{xy}\frac{\sigma_x}{\sigma_y}Y_1}\mathbb{I}_{\{Y>y_\lambda\}}) \\
&\overset{\substack{v=\frac{y_1-(\mu_y+\sigma_{xy})}{\sigma_y} \\ v_\lambda=\frac{\log y_\lambda-(\mu_y+\sigma_{xy})}{\sigma_y}}}{=} \exp\{\mu_x + \frac{1}{2}\sigma_x^2\}\int_{v_\lambda}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-v^2}dv \\
&= \exp\{\mu_x + \frac{1}{2}\sigma_x^2\}\big(1-\Phi(v_\lambda)\big), \qquad (5.23) \\
\mathsf{E}(\mathbb{I}_{\{Y>y_\lambda\}}) &= 1-\Phi(v_\lambda+\rho_{xy}\sigma_x). \qquad (5.24)
\end{aligned}$$

Therefore, the bivariate quantile weighted premium can be written as

$$\Pi^w(X,Y) = \exp\{\mu_x + \frac{1}{2}\sigma_x^2\}\left(\frac{1-\Phi(v_\lambda)}{1-\Phi(v_\lambda+\rho_{xy}\sigma_x)}\right).$$

3. **Exponential weighting ($w(u) = e^{\lambda u}$)**

   In the case of exponential weighting, the bivariate exponential weighted premium can be written as

$$\Pi^w(X,Y) = \frac{\mathsf{E}(Xe^{\lambda Y})}{\mathsf{E}(e^{\lambda Y})}. \qquad (5.25)$$

Example 5.3.10. If random loss X and random variate Y have the Gamma distribution assumptions in Example 5.3.6, we can show that the weighted density transform $\mathfrak{T}^w(X)$ also has a generalized Parato distribution with parameters $GDP(\alpha, \beta-\lambda, \theta)$. Therefore, using Equation (5.5), the bivariate power weighted premium $\Pi^w(X,Y)$ is

$$\Pi^w(X) = \frac{(\beta-\lambda)\theta}{\alpha-1}.$$

5.3.3 Calculating Multivariate Weighted Premium in More General Settings

To calculate the multivariate weighted premium according to Equation (5.6), the most challenging part is to compute $\mathsf{E}(Xw(\boldsymbol{Y}))$, which requires knowledge of the joint distribution of X and $\boldsymbol{Y}$. Furman and Zitikis (2008b) propose to split the covariance $\mathsf{Cov}\big(X, w(\boldsymbol{Y})\big)$. We note that in our multivariate premium setting, we can view $w(\boldsymbol{Y})$ as a new random variable $U = w(\boldsymbol{Y})$. In this way $\mathsf{E}(Xw(\boldsymbol{Y}))$ can be easily calculated with a bivariate copula ap-

proach. More specifically, according to Sklar's Theorem (Sklar, 1959), the joint distribution of X and U can be decomposed into two parts: the marginal distributions, denoted as $F_X(x)$ and $F_U(u)$, and the dependence structure, i.e., copula function, denoted as $C_{X,U}$. Therefore, we can express the joint c.d.f. and density of X and U as:

$$\begin{aligned} F_{X,U}(x,u) &= C_{X,U}(F_X(x), F_U(u)), \qquad (5.26) \\ f_{X,U}(x,u) &= \frac{\partial^2 F_{X,U}(x,u)}{\partial x \partial u} \\ &= c_{X,U}(F_X(x), F_U(u)) f_X(x) f_U(u), \qquad (5.27) \end{aligned}$$

where the function $c_{X,U}$ is the copula density. With Equations (5.26) and (5.27), we can easily calculate $\mathsf{E}(Xw(\boldsymbol{Y}))$ as follows:

$$\begin{aligned} \mathsf{E}(Xw(\boldsymbol{Y})) &= \mathsf{E}(XU) \\ &= \int_{\mathbb{R}^2} xu f_{X,U}(x,u) dx du \\ &= \int_{\mathbb{R}^2} xu c_{X,U}(F_X(x), F_U(u)) f_X(x) f_U(u) dx du. \qquad (5.28) \end{aligned}$$

Equation (5.28) is not difficult to compute with a known copula function. And this calculation is especially easy when the copula function is chosen from some copula families where the densities have closed forms, such as Gaussian copula and Archimedean copula family.

Table 5.2: Bivariate Weighted Premium Examples

Weighting Function	$\boldsymbol{Y \sim GAM(\alpha,\beta)}$, $\boldsymbol{X|Y \sim GAM(\theta,y)}$	$\boldsymbol{(X,Y) \sim BLogN(\mu,\Sigma)}$
Power $(\boldsymbol{w(u) = u^{\lambda}})$	$\frac{\beta\theta}{\alpha+\lambda-1}$	$\exp\{\mu_x + \frac{1}{2}\sigma_x^2 + \lambda\rho_{xy}\sigma_x\sigma_y\}$
Quantile $(\boldsymbol{w(u) = \mathbb{I}_{\{u>u_\lambda\}}})$	$\frac{\theta\beta}{\alpha-1}\left(\frac{1-\frac{\gamma(\alpha-1,\beta y_\lambda)}{\Gamma(\alpha-1)}}{1-\frac{\gamma(\alpha,\beta y_\lambda)}{\Gamma(\alpha)}}\right)$	$\exp\{\mu_x + \frac{1}{2}\sigma_x^2\}\left(\frac{1-\Phi(v_\lambda)}{1-\Phi(v_\lambda+\rho_{xy}\sigma_x)}\right)$ [a]
Exponential $(\boldsymbol{w(u) = e^{\lambda u}})$	$\frac{(\beta-\lambda)\theta}{\alpha-1}$	- [b]

[a] $v_\lambda = \frac{\log y_\lambda - (\mu_y + \sigma_{xy})}{\sigma_y}$.

[b] Exponential weighted premium for log-normal distribution is not defined since the moment generating function (m.g.f.) of the log-normal distribution is not defined for any positive value of the argument.

5.4 Properties of Multivariate Weighted Premium

As a function assigning the random loss to a real number, there are some properties to be satisfied to make certain rule a desirable premium calculation principle (Wang, 1995; Young, 2004). In this section, we will discuss these properties of the multivariate weighted premium principle.

5.4.1 Positive risk loading and no ripoff

Proposition 5.4.1. $\Pi^w(X, \boldsymbol{Y}) \geq \mathsf{E}(X)$ *if and only if* $\mathsf{Cov}\big(X, w(\boldsymbol{Y})\big) \geq 0$.

Proof. According to Definition 5.3.3 and Proposiiton 5.3.1, it is straightforward that

$$\begin{aligned} \Pi^w(X, \boldsymbol{Y}) - \mathsf{E}(X) &= \frac{\mathsf{E}\big(w(\boldsymbol{Y})X\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)} - \mathsf{E}(X) \\ &= \frac{\mathsf{E}\big(w(\boldsymbol{Y})X\big) - \mathsf{E}\big(w(\boldsymbol{Y})\big)\mathsf{E}(X)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)} \\ &= \frac{\mathsf{Cov}\big(X, w(\boldsymbol{Y})\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)}. \end{aligned} \tag{5.29}$$

Since $w(\boldsymbol{Y})$ is a positive function, we get

$$\Pi^w(X, \boldsymbol{Y}) \geq \mathsf{E}(X),$$

if and only if $\mathsf{Cov}\big(X, w(\boldsymbol{Y})\big) \geq 0$. □

In addition, note that

$$\frac{\mathsf{E}\big(w(\boldsymbol{Y})X\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)} \leq \text{esssup}(X),$$

therefore, we have

$$\mathsf{E}(X) \leq \Pi^w(X, \boldsymbol{Y}) \leq \text{esssup}(X).$$

5.4.2 No unjustified risk-loading

If risk X is degenerated, namely, there exists a constant c such that $P(X = c) = 1$, for any random vector $\boldsymbol{Y}$ we have $\mathsf{E}(X|\boldsymbol{Y}) = c$. Therefore,

$$\begin{aligned} \Pi^w(X, \boldsymbol{Y}) &= \frac{\mathsf{E}\big(w(\boldsymbol{Y})\mathsf{E}(X|\boldsymbol{Y})\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)} \\ &= \frac{\mathsf{E}\big(w(\boldsymbol{Y})c\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)} = c. \end{aligned} \tag{5.30}$$

5.4.3 Linearity

It is easy to see that for any constants a and b,

$$\begin{aligned} \Pi^w(aX + b, \boldsymbol{Y}) &= \frac{\mathsf{E}\big(w(\boldsymbol{Y})(aX + b)\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)} \\ &= \frac{\mathsf{E}\big(w(\boldsymbol{Y})X\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)} + b \\ &= a\Pi^w(X, \boldsymbol{Y}) + b. \end{aligned} \tag{5.31}$$

The linearity property indicates that the multivariate weighted premium is invariant under a scale change and also satisfies transitivity.

5.4.4 Additivity

For any two loss random variables X_1, X_2 (not necessarily independent), we have

$$\begin{aligned} \Pi^w(X_1 + X_2, \boldsymbol{Y}) &= \frac{\mathsf{E}\big(w(\boldsymbol{Y})(X_1 + X_2)\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)} \\ &= \frac{\mathsf{E}\big(w(\boldsymbol{Y})X_1\big) + \mathsf{E}\big(w(\boldsymbol{Y})X_2\big)\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)} \\ &= \Pi^w(X_1, \boldsymbol{Y}) + \Pi^w(X_2, \boldsymbol{Y}). \end{aligned} \tag{5.32}$$

5.4.5 First Stochastic Dominance Preserving

Definition 5.4.1 (Levy (1992))**.** *For two random variables $X_1, X_2 \in \mathcal{X}$, X_1 first order stochastically dominates (FSD) X_2, written $X_1 \succeq_{s.t.} X_2$, if*

$$S_{X_1}(x) \geq S_{X_2}(x) \qquad \forall x \in \mathbb{R}, \tag{5.33}$$

where $S_x(x) = 1 - F_X(x)$ is the survival function of the random variable X.

Proposition 5.4.2. *For any two random risks X_1, X_2, random vector $\mathbf{Y}$ and weighting function w, if $(X_1, w(\mathbf{Y}))$ and $(X_2, w(\mathbf{Y}))$ have the same dependence structure, then the multivariate weighted premium preserves the first ordering. In other words, if $(X_1, w(\mathbf{Y}))$ and $(X_2, w(\mathbf{Y}))$ have the same copula function $C(u, v) = \mathsf{P}(U \leq u, V \leq v)$, where U, V are uniform random variables, we have*

$$X_1 \succeq_{s.t.} X_2 \Rightarrow \Pi^w(X_1, \boldsymbol{Y}) \geq \Pi^w(X_2, \boldsymbol{Y}). \tag{5.34}$$

Proof. Let us start the proof by denoting $Z = w(\mathbf{Y})$ with c.d.f. $F_Z(z)$. The c.d.f. for X_1, X_2 are $F_{X_1}(x)$ and $F_{X_2}(x)$, respectively. We also define function $h(s, v) = s - C(s, v)$. It is easy to show that $h(s, v)$ is an increasing function with respect to s since

$$\begin{aligned} \frac{\partial h(s,v)}{\partial s} &= 1 - \frac{\partial}{\partial s} C(s,v) \\ &= 1 - \mathsf{P}(V \leq v | U = s) \geq 0. \end{aligned}$$

Also note that

$$X_1 \succeq_{s.t.} X_2 \Longleftrightarrow F_{X_1}(x) \leq F_{X_2}(x) \quad \text{for all} \quad x \geq 0,$$

hence,

$$h(F_{X_1}(x), v) \leq h(F_{X_2}(x), v), \quad \text{for all} \quad v \in [0, 1].$$

Let $v = F_Z(z)$,

$$\begin{aligned} & F_{X_1}(x) - C(F_{X_1}(x), F_Z(z)) \leq F_{X_2}(x) - C(F_{X_2}(x), F_Z(z)) \\ \Longleftrightarrow \quad & \mathsf{P}(X_1 \leq x) - \mathsf{P}(X_1 \leq x, Z \leq z) \leq \mathsf{P}(X_2 \leq x) - \mathsf{P}(X_2 \leq x, Z \leq z) \\ \Longleftrightarrow \quad & \mathsf{P}(X_1 \leq x, Z > z) \leq \mathsf{P}(X_2 \leq x, Z > z), \quad \text{for all} \quad x, z \geq 0 \\ \Longleftrightarrow \quad & \int_0^\infty \mathsf{P}(X_1 \leq x, Z > z) dz \leq \int_0^\infty \mathsf{P}(X_2 \leq x, Z > z) dz, \quad \text{for all} \quad x \geq 0. \end{aligned} \tag{5.35}$$

Also note that for any random loss X and positive random variable Z,

$$\begin{aligned}\int_0^\infty \mathsf{P}(X \le x, Z > z)dz &= \int_0^\infty \mathsf{E}\big(f_I(X,x)g_I(Z,z)\big)dz \\ &= \mathsf{E}\bigg(\int_0^\infty g_I(Z,z)dz \cdot f_I(X,x)\bigg) \\ &= \mathsf{E}\big(Zf_I(X,x)\big). \end{aligned} \tag{5.36}$$

Therefore, combining Equation (5.35) and Equation (5.36), we have

$$\mathsf{E}\big(Zf_I(X_1,x)\big) = \mathsf{E}\big(w(\mathbf{Y})f_I(X_1,x)\big) \;\le\; \mathsf{E}\big(Zf_I(X_2,x)\big) = \mathsf{E}\big(w(\mathbf{Y})f_I(X_2,x)\big). \tag{5.37}$$

Also note that the c.d.f. of X^w can be expressed as:

$$\begin{aligned} F_{X^w}(x) &= \int_0^x \frac{\mathsf{E}(w(\boldsymbol{Y})|u)f_X(u)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)}du \\ &= \int_0^\infty \frac{\mathsf{E}(w(\boldsymbol{Y})|u)f_I(X,x)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)}f_X(u)du \\ &= \frac{\mathsf{E}\big(f_I(X,x)\mathsf{E}(w(\boldsymbol{Y})|X)\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)} \\ &= \frac{\mathsf{E}\big(f_I(X,x)w(\boldsymbol{Y})\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)}. \end{aligned} \tag{5.38}$$

Therefore, combining inequality (5.37) and equation (5.38) we obtain

$$F_{X_1^w}(x) \le F_{X_2^w}(x) \Longleftrightarrow X_1^w \succeq_{s.t.} X_2^w.$$

Hence

$$\Pi^w(X_1, \boldsymbol{Y}) \ge \Pi^w(X_2, \boldsymbol{Y}).$$

□

5.4.6 Stop-loss Ordering Preserving

According to Proposition 5.4.2, it is easy to show that the MWP preserves stop-loss ordering, as formally asserted in the following corollary.

Corollary 5.4.1. *For any two random risks X_1, X_2 and random vector $\mathbf{Y}$, if $(X_1, w(\mathbf{Y}))$ and $(X_2, w(\mathbf{Y}))$ have the same dependence structure, then the multivariate weighted premium preserves the stop-loss ordering, that is, for any random pair $(X_1, w(\mathbf{Y}))$ and $(X_2, w(\mathbf{Y}))$*

with the same copula function $C(u,v) = \mathsf{P}(U \leq u, V \leq v)$*, where* U, V *are uniform random variables, and any deductible* d*,*

$$X_1 \succeq_{s.t.} X_2 \Rightarrow \Pi^w\big((X_1 - d)_+, \boldsymbol{Y}\big) \geq \Pi^w\big((X_2 - d)_+, \boldsymbol{Y}\big). \tag{5.39}$$

Proof. Since

$$X_1 \succeq_{s.t.} X_2 \Longrightarrow (X_1 - d)_+ \succeq_{s.t.} (X_2 - d)_+, \tag{5.40}$$

By denoting $Z = w(\mathbf{Y})$ with c.d.f function $F_Z(z)$, and also noting that for any random loss X,

$$\mathsf{P}\big((X - d)_+ \leq x, Z > z\big) \quad = \quad \begin{cases} 0, & x < 0, \\ \mathsf{P}(X \leq d + x, Z > z). \end{cases}$$

According to inequality (5.35), we have

$$\mathsf{P}\big((X_1 - d)_+ \leq x, Z > z\big) \leq \mathsf{P}\big((X_2 - d)_+ \leq x, Z > z\big), \quad \text{for all} \quad x, z \geq 0.$$

Similarly to the proof of Proposition 5.4.2, we can show that

$$(X_1 - d)_+^w \succeq_{s.t.} (X_2 - d)_+^w, \tag{5.41}$$

where $(X - d)_+^w$ denotes the weighted random variable of $(X - d)_+$. Therefore,

$$\Pi^w\big((X_1 - d)_+, \boldsymbol{Y}\big) \geq \Pi^w\big((X_2 - d)_+, \boldsymbol{Y}\big).$$

□

Proposition 5.4.2 shows that the weighted premium principle preserves the stochastic ordering. Corollary 5.4.1 shows that the introduction of deductibles does not modify the premiums. It is also interesting to have the following result.

Proposition 5.4.3. *If* $\mathsf{Cov}(X, w(\boldsymbol{Y})) \geq 0$*, the weighted loss random variable* X^w *is first order stochastically dominates the original random variable* X*:* $X^w \succeq_{s.t.} X$*.*

Proof. From Equation (5.38), we have

$$\begin{aligned} S_{X^w}(x) &= \frac{\mathsf{E}\big((1-f_I(X,x)w(\boldsymbol{Y})\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)} \\ &= \frac{\mathsf{E}\big(g_I(X,x)w(\boldsymbol{Y})\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)}, \end{aligned} \tag{5.42}$$

$$\begin{aligned} S_{X^w}(x) - S_X(x) &= \frac{\mathsf{E}\big(w(\boldsymbol{Y})\big)\mathsf{E}(f_I(X,x)) - \mathsf{E}\big(w(\boldsymbol{Y})f_I(X,x)\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)} \\ &= -\frac{\mathsf{Cov}(w(\boldsymbol{Y}), f_I(X,x))}{\mathsf{E}\big(w(\boldsymbol{Y})\big)}, \end{aligned} \tag{5.43}$$

where $\mathsf{Cov}(w(\boldsymbol{Y}), f_I(X,x))$ is the covariance of $w(\boldsymbol{Y})$ and $f_I(X,x)$, which can be expressed as:

$$\begin{aligned} \mathsf{Cov}(w(\boldsymbol{Y}), f_I(X,x)) &= \mathsf{E}\bigg([w(\boldsymbol{Y}) - \mathsf{E}\big(w(\boldsymbol{Y})\big)][f_I(X,x) - \mathsf{E}(f_I(X,x))]\bigg) \\ &\leq 0. \end{aligned} \tag{5.44}$$

The above inequality holds because $\mathsf{Cov}(X, w(\boldsymbol{Y})) \geq 0$ and $f_I(s,t)$ is a decreasing function with respect to s. Therefore, $S_{X^w}(x) - S_X(x) \geq 0$ for all $x \geq 0$, namely, $X^w \succeq_{s.t.} X$. □

Proposition 5.4.3 shows that X^w, the loss random variable weighted by another positively correlated random variable $w(\boldsymbol{Y})$, is distributed with heavier tails than the original loss X. This property has appealing empirical interpretations, since premiums calculated from the weighted loss X^w provides another way of loading to reflect the inherent risk. If we define the risk loading, denoted as Θ_π, of the premium on the loss X, $\Pi(X)$, as

$$\Theta_\Pi = \frac{\Pi(X)}{\mathsf{E}(X)} - 1, \tag{5.45}$$

$$\tag{5.46}$$

we can investigate the expressions of Θ_Π for each weighted premiums in Section 5.3.2. As some illustrative examples, we examine the risk loadings for power weighted gamma distribution in both univariate premium and bivariate weighted premium cases.

Example 5.4.1. If X has A Gamma distribution with parameter $GAM(a,b)$, the univariate power weighted ($w(u) = u^\lambda$) premium $\Pi^w(X)$ is $\Pi^w(X) = \dfrac{a+\lambda}{b}$. Therefore, the risk loading of this premium, Θ^w_Π, can be expressed as $\Theta^w_\Pi = \dfrac{\lambda}{a}$. We can see that if $\lambda > 0$, $\Pi^w(X) > \mathsf{E}(X)$

and $\Theta^w_\Pi > 0$. In other words, $\mathsf{Cov}(X, X^\lambda) > 0$ and the premium has positive risk loading.

Example 5.4.2. If random loss X and random variate Y have the Gamma distribution assumptions in Example 5.3.6, then the bivariate power weighted ($w(u) = u^\lambda$) premium $\Pi^w(X, Y)$ is $\Pi^w(X, Y) = \dfrac{\beta\theta}{\alpha + \lambda - 1}$. Therefore, if $\lambda < 0$, the risk loading, expressed as $\Theta^w_\Pi = \dfrac{-\lambda}{a + \lambda - 1}$, is greater than 0. We can also check that in the case of $\lambda < 0$, it satisfies the condition of $\mathsf{Cov}(X, Y^\lambda) > 0$.

Table 5.3: Risk loading Univariate Weighted Premium Examples. The last column shows the conditions for positive risk loadings.

Weighting Function	$\boldsymbol{GAM(a, b)}$ $(\lambda > 0)$	$\boldsymbol{LogN(\mu, \sigma^2)}$
Power $\boldsymbol{(w(u) = u^\lambda)}$	$\frac{\lambda}{b}$ $(\lambda > 0)$	$e^{\lambda\sigma^2} - 1$ $(\lambda > 0)$
Quantile $\boldsymbol{(w(u) = \mathbb{I}_{\{u > u_\lambda\}})}$	$\frac{\frac{\gamma(a, bx_\lambda)}{\Gamma(a)} - \frac{\gamma(a+1, bx_\lambda)}{\Gamma(a+1)}}{1 - \frac{\gamma(a, bx_\lambda)}{\Gamma(a)}}$ $(x_\lambda > 0)$	$\frac{\Phi(\mu_\lambda) - \Phi(\mu_\lambda - \sigma)}{1 - \Phi(\mu_\lambda)}$ $(x_\lambda \in \mathbb{R})$ [a]
Exponential $\boldsymbol{(w(u) = e^{\lambda u})}$	$\frac{\lambda}{b - \lambda}$ $(\lambda > 0)$	- [b]

[a] $u_\lambda = \frac{\sigma \log x_\lambda - x + \mu}{\sigma^2}$.

[b] Exponential weighted premium for log-normal distribution is not defined since the moment generating function (m.g.f.) of log-normal distribution is not defined for any positive value of the argument.

Table 5.4: Bivariate Weighted Premium Examples

Weighting Function	$\boldsymbol{Y \sim GAM(\alpha, \beta)}$, $\boldsymbol{X|Y \sim GAM(\theta, y)}$	$\boldsymbol{(X, Y) \sim BLogN(\mu, \Sigma)}$
Power $\boldsymbol{(w(u) = u^\lambda)}$	$\frac{-\lambda}{\alpha + \lambda - 1}$ $(\lambda < 0)$	$e^{\lambda\rho_{xy}\sigma_x\sigma_y} - 1$ $(\lambda\rho_{xy} < 0)$
Quantile $\boldsymbol{(w(u) = \mathbb{I}_{\{u > u_\lambda\}})}$	$\frac{\frac{\gamma(\alpha, \beta y_\lambda)}{\Gamma(\alpha)} - \frac{\gamma(\alpha-1, \beta y_\lambda)}{\Gamma(\alpha-1)}}{1 - \frac{\gamma(\alpha, \beta y_\lambda)}{\Gamma(\alpha)}}$ $(y_\lambda < 0)$	$\frac{\Phi(v_\lambda + \rho_{xy}\sigma_x) - \Phi(v_\lambda)}{1 - \Phi(v_\lambda + \rho_{xy}\sigma_x)}$ $(\rho_{xy} > 0, y_\lambda \in \mathbb{R})$ [a]
Exponential $\boldsymbol{(w(u) = e^{\lambda u})}$	$\frac{-\lambda}{\beta}$ $(\lambda < 0)$	- [b]

[a] $v_\lambda = \frac{\log y_\lambda - (\mu_y + \sigma_{xy})}{\sigma_y}$.

[b] Exponential weighted premium for log-normal distribution is not defined since the moment generating function (m.g.f.) of log-normal distribution is not defined for any positive value of the argument.

Table 5.5 provides a comparison of the various properties among some selected premium principles discussed in this chapter. A check mark ("✓") implies the premium principle satisfies the corresponding property, while a cross-mark ("✗") does not. Note that the distortion premium principle and our proposed MVP are the only two that satisfy all the properties.

Table 5.5: Properties of different premium principles. "Expectation" is for Expectation Premium Principle, "SD" stands for Standard Deviation Premium Principle, "Esscher" represents for Esscher Premium Principle, "Distortion" stands for Distortion Premium Principle, and "MWP" stands for Multivariate Weighted Premium Principle.

Property	a. Expectation	b. SD	c. Esscher	d. Distortion	e. MWP
1. Risk loading	✓	✓	✓	✓	✓[a]
2. No unjustified	✗	✓	✓	✓	✓
3. No ripoff	✗	✗	✓	✓	✓
4. Translation	✗	✓	✓	✓	✓
5. Scale	✓	✓	✗	✓	✓[b]
6. Subadditivity	✓	✗	✗	✓	✓[c]
7. FSD	✓	✗	✗	✓	✓[d]
8. SL	✓	✗	✗	✓	✓[e]

[a] $\mathsf{Cov}\big(X, w(\boldsymbol{Y})\big) \geq 0$.
[b] X and $\boldsymbol{Y}$ are different random variables.
[c] X and $\boldsymbol{Y}$ are different random variables.
[d] See Proposition 5.4.2.
[e] See Corollary 5.4.1.

5.4.7 Premium Allocation Among Layers

In this subsection we define the layer random loss variable and its corresponding absolute risk (AR) function and relative risk function (RR) based on Wang (1995).

Definition 5.4.2. *A layer $(a, b]$ of random loss X, denoted by $\mathbf{L}_{(a,b]}$, is defined as:*

$$\mathbf{L}_{(a,b]} = \begin{cases} 0, & 0 \leq X < a \\ X - a, & a \leq X < b \\ b - a, & b < X. \end{cases} \tag{5.47}$$

Definition 5.4.3. *1. A premium principle $\Pi(X)$ has decreasing absolute risk load if the Absolute Risk (AR) function, defined as,*

$$AR(x) \quad = \quad \Pi(\mathbf{L}_{(x,x+h]}), \quad h > 0, \tag{5.48}$$

is a decreasing function with respect to x.

2. A premium principle $\Pi(X)$ has increasing relative risk load if the Relative Risk (RR)

function, defined as,

$$RR(x) \quad = \quad \lim_{h\to 0} \frac{\Pi^w(\mathbf{L}_{(x,x+h]})}{E(\mathbf{L}_{(x,x+h]})}, \tag{5.49}$$

is an increasing function with respect to x.

Definition 5.4.4. *The hazard function of a random loss X is defined as*

$$\lambda_X(x) \quad = \quad \frac{f_X(x)}{S_X(x)}. \tag{5.50}$$

The hazard rate order is closely related to conditional stochastic ordering and the hazard function.

Definition 5.4.5 (Denuit et al. (2005))**.** *Given two random variables X_1 and X_2 , X_1 is said to precede X_2 in the hazard rate order, denoted as $X_1 \preceq_{h.r.} X_2$, if*

$$[X_1|X_1 > t] \preceq_{s.t.} [X_2|X_2 > t], \qquad \textit{for all} \quad t \in \mathbb{R}. \tag{5.51}$$

Section 5.4.5 shows that the MWP preserves stochastic order. A natural and follow-up question is the allocation of premium among layers. We are particularly interested in verifying whether the MWP premium is layer additive as well as preserving stochastic ordering among different layers.

Theorem 5.4.1. *Given a random loss X and a random vector $\boldsymbol{Y}$, we have:*

1. ***(Layer Additive)*** *multivariate weighted premium is "layer additive", i.e., given a partition of the domain of X, $\{(x_i, x_{i+1}], i = 0, 1, \ldots\}, 0 = x_0 < x_1 < x_2 < \ldots$, we have*

$$\Pi^w(X, \boldsymbol{Y}) = \sum_{i=0}^{\infty} \Pi^w(\mathbf{L}_{(x_i,x_{i+1}]}, \boldsymbol{Y}). \tag{5.52}$$

2. ***(Decreasing Absolute Risk Load)*** *the absolute risk function $AR(x)$ is decreasing with respect to x. In other words, for any constant $h > 0$, we have*

$$x < y \Rightarrow \Pi^w(\mathbf{L}_{(x,x+h]}, \boldsymbol{Y}) \geq \Pi^w(\mathbf{L}_{(y,y+h]}, \boldsymbol{Y}). \tag{5.53}$$

3. ***(Increasing Relative Risk Load)*** *the relative risk function $RR(x)$ is increasing with respect to x if, and only if, $X \preceq_{h.r.} X|\boldsymbol{Y}$.*

Proof. 1. Since

$$X = \sum_{i=0}^{\infty} \mathbf{L}_{(x_i, x_{i+1}]}, \tag{5.54}$$

we have

$$\begin{aligned} \mathsf{E}\big(Xw(\boldsymbol{Y})\big) &= \mathsf{E}\big(\sum_{i=0}^{\infty} \mathbf{L}_{(x_i,x_{i+1}]} w(\boldsymbol{Y})\big) \\ &= \sum_{i=0}^{\infty} \mathsf{E}\big(\mathbf{L}_{(x_i,x_{i+1}]} w(\boldsymbol{Y})\big), \\ \frac{\mathsf{E}\big(Xw(\boldsymbol{Y})\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)} &= \sum_{i=0}^{\infty} \frac{\mathsf{E}\big(\mathbf{L}_{(x_i,x_{i+1}]} w(\boldsymbol{Y})\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)}. \end{aligned} \tag{5.55}$$

Namely, $\Pi^w(X, \boldsymbol{Y}) = \sum_{i=0}^{\infty} \Pi^w(\mathbf{L}_{(x_i,x_{i+1}]}, \boldsymbol{Y})$.

2. The AR function and its derivative can be written as:

$$\begin{aligned} AR(x) &= \frac{\mathsf{E}\big(w(\boldsymbol{Y})\mathbf{L}_{(x,x+h]}\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)} \\ &= \frac{\mathsf{E}\big[w(\boldsymbol{Y})\mathsf{E}\big(\mathbf{L}_{(x,x+h]}|\boldsymbol{Y}\big)\big]}{\mathsf{E}\big(w(\boldsymbol{Y})\big)} \\ &= \frac{\mathsf{E}\big(w(\boldsymbol{Y})\int_x^{x+h} S_{X|\boldsymbol{Y}}(u|\boldsymbol{Y})du\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)} \end{aligned} \tag{5.56}$$

$$\begin{aligned} \frac{dAR(x)}{dx} &= \frac{\mathsf{E}\big[w(\boldsymbol{Y})(S_{X|\boldsymbol{Y}}(x+h|\boldsymbol{Y}) - S_{X|\boldsymbol{Y}}(x|\boldsymbol{Y}))\big]}{\mathsf{E}\big(w(\boldsymbol{Y})\big)} \\ &\leq 0. \end{aligned} \tag{5.57}$$

The inequality holds because $y = S_{X|\boldsymbol{Y}}(x|\boldsymbol{Y})$ is a decreasing function of x.

3. It is shown in Property 3.3.38 in Denuit et al. (2005) that

$$X \preceq_{h.r.} X|\boldsymbol{Y} \Leftrightarrow \lambda_X(x) \geq \lambda_{x|\boldsymbol{Y}}(x), \quad \text{for all } x \geq 0. \tag{5.58}$$

Also note that the RR function can be expressed as

$$RR(x) = \lim_{h\to 0} \frac{\mathsf{E}\big[w(\boldsymbol{Y})\int_x^{x+h} S_{X|\boldsymbol{Y}} du\big]}{\mathsf{E}\big(w(\boldsymbol{Y})\big)\int_x^{x+h} S_X(u)du}, \tag{5.59}$$

therefore, according to L'Hôpital's rule

$$\begin{aligned} RR(x) &= \lim_{h\to 0} \frac{\mathsf{E}\big(w(\boldsymbol{Y})[S_{X|\boldsymbol{Y}}(x+h|\boldsymbol{Y})]\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)S_X(x+h)} \\ &= \frac{\mathsf{E}\big(w(\boldsymbol{Y})S_{X|\boldsymbol{Y}}(x|\boldsymbol{Y})\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)S_X(x)}. \end{aligned} \tag{5.60}$$

Hence the derivative of the RR function is

$$\frac{dRR(x)}{dx} = \frac{\mathsf{E}\bigg(w(\boldsymbol{Y})\frac{S_{X|\boldsymbol{Y}}(x|\boldsymbol{Y})}{S_X(x)}(\lambda_X(x)-\lambda_{X|\boldsymbol{Y}}(x|\boldsymbol{Y}))\bigg)}{\mathsf{E}\big(w(\mathbf{Y})\big)}. \tag{5.61}$$

Combining Equation (5.61) and (5.58), we conclude that

$$X \preceq_{h.r.} X|\boldsymbol{Y} \Leftrightarrow \frac{dRR(x)}{dx} \geq 0.$$

□

In view of **Part 3** in Theorem 5.4.1, we can see that multivariate weighted premium principle is superior over the univariate version in the sense that univariate weighted premium has constant risk loading. We summarize this property in the following corollary.

Corollary 5.4.2. *The univariate weighted premium,* $\Pi^w(X) = \dfrac{E\big(w(X)X\big)}{E\big(w(X)\big)}$*, has a constant relative risk loading. More specifically,*

$$RR(x) = \lim_{h\to 0} \frac{\Pi^w(X)}{E(\mathbf{L}_{(x,x+h]})} = 1. \tag{5.62}$$

Proof. The RR function in the univariate weighted premium principle context can be expressed as

$$\begin{aligned} RR(x) &= \lim_{h\to 0} \frac{\mathsf{E}\big(w(\mathbf{L}_{(x,x+h]})\mathbf{L}_{(x,x+h]}\big)}{\mathsf{E}\big(w(\mathbf{L}_{(x,x+h]})\big)\mathsf{E}\big(\mathbf{L}_{(x,x+h]}\big)} \\ &= \lim_{h\to 0} \frac{\mathsf{E}\big(w(X-x)\mathbb{I}_{\{x\leq X<x+h\}}\big)+\mathsf{E}\big(w(h)h\mathbb{I}_{\{X\geq x+h\}}\big)}{[\mathsf{E}\big(w(0)\mathbb{I}_{\{X<x\}}\big)+\mathsf{E}\big(w(X-x)\mathbb{I}_{\{x\leq X<x+h\}}\big)+\mathsf{E}\big(w(h)\mathbb{I}_{\{X\geq x+h\}}\big)]\mathsf{E}\big(\mathbf{L}_{(x,x+h]}\big)}. \end{aligned}$$

For notation convenience, we define $E_u = \mathsf{E}\big(w(X-x)\mathbb{I}_{\{x\leq X<x+h\}}\big)+\mathsf{E}\big(w(h)h\mathbb{I}_{\{X\geq x+h\}}\big)$, $E_{d_1} = \mathsf{E}\big(w(0)\mathbb{I}_{\{X<x\}}\big)+\mathsf{E}\big(w(X-x)\mathbb{I}_{\{x\leq X<x+h\}}\big)+\mathsf{E}\big(w(h)\mathbb{I}_{\{X\geq x+h\}}\big)$ and $E_{d_2} = \mathsf{E}\big(\mathbf{L}_{(x,x+h]}\big)$.

Hence,

$$RR(x) = \lim_{h\to 0} \frac{E_u}{E_{d_1} E_{d_2}}.$$

Note that when $h \to 0$, $E_u \to 0$ and $E_{d_2} \to 0$, according to L'Hôpital's rule:

$$RR(x) = \lim_{h\to 0} \frac{\frac{d}{dx} E_u}{E_{d_1} \frac{d}{dx} E_{d_2} + E_{d_2} \frac{d}{dx} E_{d_1}}.$$

In addition,

$$\begin{aligned}
E_{d_1}|_{h\to 0} &= w(0)F_X(x) + w(0)S_X(x), \\
\frac{d}{dx} E_{d_2}|_{h\to 0} &= S_X(x+h)|_{h\to 0} = S_X(x), \\
\frac{d}{dx} E_u|_{h\to 0} &= \big(w'(h)h + w(h)\big) S_X(x+h)|_{h\to 0} = w(0)S_X(x), \\
E_{d_2} \frac{d}{dx} E_{d_1}|_{h\to 0} &= 0.
\end{aligned}$$

Thus

$$RR(x) = \frac{w(0)S_X(x)}{\big(w(0)F_X(y) + w(0)S_X(y)\big)S_X(x)} = 1. \tag{5.63}$$

□

5.5 Relationship with Wang's Premium Principle

Wang's premium premium is defined with respect to a distortion function, $g : [0,1] \to [0,1]$, which satisfies the following three conditions:

(i). $g'(x) \geq 0$;

(ii). $g(0) = 0, g(1) = 1$;

Then Wang's premium principle based on the function g, denoted as $H(X; g)$, is expressed as (Furman and Zitikis, 2008a; Wang, 1995)

$$H(X; g) = \int_{x\in\mathbb{R}_+} g\big(S_X(x)\big) dx. \tag{5.64}$$

The following proposition shows that for a particular MWP we can find a corresponding representation for the Wang's principle.

Theorem 5.5.1. *Given a random risk $X \in \mathfrak{X}$ such that $\mathsf{E}(|X|) < \infty, \mathsf{E}(|X|^2) < \infty$, a random vector $\boldsymbol{Y}$, which is X measurable, and a weighting function w, then there exists a function $g(x)$, satisfying conditions (i)-(iii).*

Proof. If $\boldsymbol{Y}$ is X measurable, we can find a non-negative function, $h(x)$, satisfying

$$h(X) = \frac{w(\boldsymbol{Y})}{\mathsf{E}\big(w(\boldsymbol{Y})\big)} \geq 0. \tag{5.65}$$

Also recall that

$$\begin{aligned} \Pi^w(X, \boldsymbol{Y}) &= \frac{\mathsf{E}\big(w(\boldsymbol{Y})X\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)} \\ &= \mathsf{E}(Xh(X)). \end{aligned} \tag{5.66}$$

Let us define the function $g(x)$ as $g(x) = \int_0^x h\big(S_X^{-1}(u)\big)du$, then we can verify that $g(x)$ satisfies the conditions (i)-(ii). More specifically,

(i). $g'(x) = h\big(S_X^{-1}(x)\big) \geq 0$;

(ii). $g(0) = 0$, $g(1) = \int_0^1 h\big(S_X^{-1}(u)\big)du = \int_0^\infty h(x)dF(x) = \mathsf{E}\big(h(X)\big) = 1$.

Additionally, the multivariate weighted premium principle based on the function g can be expressed as

$$\begin{aligned} \int_0^\infty g\big(S_X(x)\big)dx &= \mathsf{E}\big(Xg'(S_X(X))\big) \\ &= \mathsf{E}\big(Xh(X)\big) \\ &= \Pi^w(X, \boldsymbol{Y}). \end{aligned} \tag{5.67}$$

□

The risk-adjusted premium based on the proportional hazard (PH) transform is a special case of Wang's premium (Wang, 1995). According to the definition of hazard function in Equation (5.50), we can see that the PH transform adjusts the risk X by mapping the original loss variable X to another random variable X^{PH} with a hazard rate function

$$\lambda_{X^{\text{PH}}}(t) = \frac{1}{\rho}\lambda_X(t), \quad \rho > 0. \tag{5.68}$$

Considering the hazard rate function of the multivariate weighted random variable X^w, we find that

$$\lambda_{X^w}(t) = \frac{f_{X^w}(t)}{1 - F_{X^w}(t)} \tag{5.69}$$

$$= \frac{S_X(t)\mathsf{E}\big(w(\boldsymbol{Y})|X = t\big)}{S_{X^w}(t)\mathsf{E}\big(w(\boldsymbol{Y})\big)}\lambda_X(t). \tag{5.70}$$

In other words, instead of a constant ρ, multivariate weighted premium principle uses a function,

$$\rho(t) = \frac{S_{X^w}(t)\mathsf{E}\big(w(\boldsymbol{Y})\big)}{S_X(t)\mathsf{E}\big(w(\boldsymbol{Y})|X = t\big)}, \tag{5.71}$$

to deflect the hazard rate. Indeed, the weighting function of the risk-adjust premium can be expressed as

$$w(x) = c\big(S_X(x)\big)^{\frac{1}{\rho}-1}, \tag{5.72}$$

where c is a constant satisfying $c = \dfrac{\mathsf{E}(w(X))}{\rho}$.

5.6 Selecting the Auxiliary Variables

Despite many desirable properties of the multivariate premium principle, from an empirical application point of view, it is always important, yet challenging, to have a good idea of how to choose the auxiliary weighting random vector $\boldsymbol{Y}$. Bühlmann's 1980 paper provides a great example, in which an economic premium principle is derived under some equilibrium conditions. In particular, as a special case of this economic premium principle, when the exponential utility function $U(W) = e^{-\rho W}$ is considered, the economic premium principle is

$$\Pi(X, Y) = \frac{E\big(Xe^{\rho Y}\big)}{E\big(e^{\rho Y}\big)}, \tag{5.73}$$

where $\rho = 1/\sum_{i=1}^{p}\dfrac{1}{\rho_i}$, ρ_i is the risk aversion for each risk agency in the market and $Y = \sum_{i=1}^{p} X_i$ is the sum of all the risks in the market, $i = 1, \ldots, p$. It is interesting to note that if we define the weighting function $w(y)$ as $w(y) = e^{\rho y}$, and choose $Y = \sum_{i=1}^{p} X_i$

as the weighing random variable, then this economic premium can also be expressed as a multivariate weighted premium. More interestingly, the *Esscher principle* is a special case of this economic premium principle, where X and $Y - X$ are independent.

Now we give another interesting example to show the advantages of the proposed multivariate weighted premium principle. Generally speaking, an arbitrary premium principle, Π, can always be written as a summation of the risk premium (i.e., net premium) and a risk loading, namely,

$$\Pi = \mathsf{E}(X) + \Theta_{\Pi}(X), \tag{5.74}$$

where $\Theta_{\Pi}(X)$ denotes the risk loading. The expressions of risk loadings can be different according to the premium principle we choose. For example, the expectation premium principle uses $\theta\mathsf{E}(X)$ to present the risk loading, where θ is a constant (loading factor); while in the standard deviation premium principle, the risk loading is proportional to the standard deviation of the underlying risk, i.e., $\theta\sqrt{\mathsf{Var}(X)}$, where θ is a constant. In the case of weighted premium principle, the risk loading can be expressed as a function of the covariance of the random loss and the weighting variable. To be more specific, for the univariate weighted premium principle, the premium can be written as

$$\begin{aligned} \Pi^{w}(X) &= \frac{\mathsf{E}\big(w(X)X\big)}{\mathsf{E}\big(w(X)\big)} \\ &= \mathsf{E}(X) + \frac{\mathsf{Cov}\big(X, w(X)\big)}{\mathsf{E}\big(w(X)\big)}. \end{aligned} \tag{5.75}$$

Similarly, for multivariate weighted premium principle, the premium can be written as

$$\begin{aligned} \Pi^{w}(X, \boldsymbol{Y}) &= \frac{\mathsf{E}\big(w(\boldsymbol{Y})X\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)} \\ &= \mathsf{E}(X) + \frac{\mathsf{Cov}\big(X, w(\boldsymbol{Y})\big)}{\mathsf{E}\big(w(\boldsymbol{Y})\big)}. \end{aligned} \tag{5.76}$$

From an insurance ratemaking point of view, a good premium principle needs to provide the insurers sufficient and stable risk loading. Assume now that there are p risks in the market, $X_1, \ldots, X_p$, we choose the weighting variable to be $Y = (\sum_{i=1}^{p} X_i - d)_+ \wedge m$, and use the exponential weighting function $w(y) = e^{\eta y}$, then the corresponding univariate and

multivariate weighted premium for each risk, X_i, can be expressed as

$$\Pi^w(X_i) = \frac{\mathsf{E}\big(X_i\exp(X_i)\big)}{\mathsf{E}\big(\exp(X_i)\big)}, \tag{5.77}$$

$$\Pi^w(X_i,Y) = \frac{\mathsf{E}\big(X_i\exp\{\eta(\sum_{i=1}^p X_i-d)_+\wedge m\}\big)}{\mathsf{E}\big(\exp\{\eta(\sum_{i=1}^p X_i-d)_+\wedge m\}\big)}. \tag{5.78}$$

To compare, we set both premiums equal to an expected premium with loading factor θ, namely, we let $\Pi^w(X_i)=\Pi^w(X_i,Y)=(1+\theta)\mathsf{E}(X_i)$, and we compare the variations of the estimations of the two premiums. As an example, we further assume that all the risks follow independent gamma distributions with a common rate parameter, i.e., $X_i\sim GAM(a_i,b)$, therefore, $\sum_{i=1}^p X_i\sim GAM(\sum_{i=1}^p a_i,b)$. Denote $X=\sum_{i=1}^p X_i, a=\sum_{i=1}^p a_i, X_{-i}=\sum_{j\neq i}X_j$ and $a_{-i}=\sum_{j\neq i}a_j$, then $X\sim GAM(a,b)$ and $X_{-i}\sim GAM(a_{-i},b)$. Now, it is straightforward to write the denominator of equation (5.78) as

$$\begin{aligned}\mathsf{E}\big(\exp\{\eta(X-d)_+\wedge m\}\big) &= \mathsf{E}\big(\mathbb{I}_{\{X<d\}}\big)+\mathsf{E}\big(e^{\eta(X-d)}\mathbb{I}_{\{d\le X<m+d\}}\big)+\mathsf{E}\big(e^{\eta m}\mathbb{I}_{\{X\ge m+d\}}\big)\\ &= \frac{\gamma(a,bd)}{\Gamma(a)}+e^{-\eta d}\frac{b^a}{(b-\eta)^a}\frac{\gamma\big(a,b(d+m)\big)-\gamma\big(a,bd\big)}{\Gamma(a)}+e^{\eta m}\Big(1-\frac{\gamma\big(a,b(d+m)\big)}{\Gamma(a)}\Big).\end{aligned}$$

Meanwhile, the the numerator of equation (5.78) can be written as

$$\begin{aligned}\mathsf{E}\big(X_i\exp\{\eta(X-d)_+\wedge m\}\big) &= \mathsf{E}\big(X_i\mathsf{E}(\mathbb{I}_{\{X_{-i}<d-X_i\}}|X_i)\big)+\mathsf{E}\big(X_i\mathsf{E}(\mathbb{I}_{\{d-X_i\le X_{-i}<d+m-X_i\}}|X_i)\big)\\ &+ \mathsf{E}\big(X_i\mathsf{E}(\mathbb{I}_{\{X_{-i}\ge d+m-X_i\}}|X_i)\big).\\ \mathsf{E}\big(X_i\mathsf{E}(\mathbb{I}_{\{X_{-i}<d-X_i\}}|X_i)\big) &= \int_0^d x^{a_i}e^{-bx}\frac{b^{a_i}\gamma\big(a_{-i},b(d-x)\big)}{\Gamma(a_i)\Gamma(a_{-i})}dx\\ \mathsf{E}\big(X_i\mathsf{E}(\mathbb{I}_{\{d-X_i\le X_{-i}<d+m-X_i\}}|X_i)\big) &= \int_0^d x^{a_i}e^{(\eta-b)x-\eta d}b^{a_{-i}}\frac{\gamma\big(a_{-i},b(m+d-x)\big)-\gamma\big(a_{-i},b(d-x)\big)}{(b-\eta)^{a_{-i}}\Gamma(a_i)\Gamma(a_{-i})}dx\\ &+ \int_d^{d+m} x^{a_i}e^{(\eta-b)x-\eta d}b^{a_{-i}}\frac{\gamma\big(a_{-i},b(m+d-x)\big)}{(b-\eta)^{a_{-i}}\Gamma(a_i)\Gamma(a_{-i})}dx\\ \mathsf{E}\big(X_i\mathsf{E}(\mathbb{I}_{\{X_{-i}\ge d+m-X_i\}}|X_i)\big) &= \int_0^{d+m} x^{a_i}e^{-bx+\eta m}\frac{b^{a_i}}{\Gamma(a_i)}\Big(1-\frac{\gamma(a_{-i},b(d+m-x))}{\Gamma(a_{-i})}\Big)dx\\ &+ \int_{d+m}^{\infty} x^{a_i}e^{-bx+\eta m}\frac{b^{a_i}}{\Gamma(a_i)}dx.\end{aligned}$$

By setting $\Pi^w(X_i)=(1+\theta)\mathsf{E}(X_i), i=1,2,3$, we get $\eta_i^{\text{univ}}=\dfrac{b\theta}{1+\theta}$, where η_i^{univ} is the parameter under the univariate weighted premium for risk i. The corresponding estimator for the multivariate premium, η_i^{mult}, is calculated numerically by setting $\Pi^w(X_i,Y)=(1+\theta)\mathsf{E}(X_i)$. In the numerical example below, we assume that there are three risks, $a_1=1, a_2=$

$2, a_3 = 3$, and $b = 0.5$. We also assume that the loading factor under the expectation premium principle, θ, is 0.1. It is easy to calculate $\eta_i^{\text{univ}} = 0.0455, i = 1, 2, 3$, and $\eta_1^{\text{multi}} = 0.0269, \eta_2^{\text{multi}} = 0.0189, \eta_3^{\text{multi}} = 0.0145$. Under the condition that $\Pi^w(X_i) = \Pi^w(X_i, Y) = (1 + \theta)\mathsf{E}(X_i)$, we find the optimal $d^* = 4.90$ and $m^* = 49.60$, such that the variances of the multivariate weighted premium estimators are minimized. The standard errors of the estimated premiums under the univariate and multivariate premium principles are listed in Table 5.6. We can see that the variation of the multivariate premium for each risk is smaller than that of the univariate weighted premium.

Table 5.6: Variation of the premium estimations. "UWP" denotes univariate weighted premium and "MWP" denotes multivariate weighted premium.

Risk	Premium	Standard Errors of Estimation	
		UWP	MWP
X_1	2.2	0.4269	0.3805
X_2	4.4	0.6238	0.5027
X_3	6.6	0.6340	0.5254

5.7 Empirical Analysis

5.7.1 Data and Reinsurance Contract

The empirical study of this chapter employs a data set that covers private reinsurance in Manitoba, including actual indemnities and liabilities from 2001 through 2011. The private reinsurance program in Manitoba uses an excess of loss (XoL) reinsurance policy that is defined as follows. For the loss random variable $X \in \mathcal{X}$, the reinsurance company covers part of the loss between the layers a and b. In other words, for the XoL random variable, $\mathbf{L}_{[a,b]}$, defined as,

$$\mathbf{L}_{[a,b]} = \begin{cases} 0, & X < a \\ X - a, & a \le X < b \\ b - a, & X \ge b, \end{cases} \tag{5.79}$$

the reinsurance contract has the form, $\tau \mathbf{L}_{[a,b]}$, where $0 < \tau < 1$ is the coverage level. As mentioned before, it is critical to select the auxiliary variables $\boldsymbol{Y}$ to implement the multivariate premium principle. A natural choice of $\boldsymbol{Y}$ is the liability, since in agricultural insurance

and reinsurance, both indemnity and liability are closely related to the crop yield. In fact, in empirical pricing, the loss cost ratio (LCR), defined as the ratio of the indemnity over liability, is recommended for standardizing the trends of farm practice, technology improvement, etc. However, empirical pricing based on LCR is still weighting the liability with some ad hoc method that lacks scientific foundations. Using multivariate weighted premium principle, this chapter provides a more comprehensive and scientific methodology to integrate liability, as well as other auxiliary variables, into the reinsurance pricing. Another set of promising candidate for auxiliary variables is the economic variables, such as the Gross Domestic Product (GDP), Consumer Price Index(CPI), crop commodity prices, etc. The economic conditions have important impact on the agricultural reinsurance industry and of course integrating these variables into the reinsurance premium will improve the ratemaking framework.

In the empirical analysis of this chapter, we use both liability (denoted as Y_{Liab}) and Canadian GDP (denoted as Y_{GDP}) as the auxiliary variables. Note that our pricing framework is not necessarily restricted to two dimensions, and one can select an arbitrary k-dimensional auxiliary variables for the pricing, as we will show in the pricing formulas in Section 5.7.2 (Proposition 5.7.4 and Proposition 5.7.5). The descriptive statistical summary is displayed in Table 5.7. We observe that the data has skewness and kurtosis values that are very close to zero, leading us to consider multivariate normal to model their joint distribution. Therefore, we execute a variety of (univariate and multivariate) normality tests to confirm that multivariate normal distribution is an appropriate model for the data. The test results are listed in Table 5.8 and a "✓" to indicate the data has passed the test at 0.05 significant level. The univariate normality tests we consider in this study include Shapiro-Wilk's test, Cramer-von Mises's test, Kolmogorov-Smirnov's test, and Anderson-Darling's test, while the multivariate normality tests we consider are Mardia's test, Henze-Zirkler's test and Royston's test [1]. From Table 5.8, we find that the data pass the individual normality tests as well as the joint multivariate normality tests. Therefore, it is reasonable to define the distribution as follows:

$$\begin{aligned} X &= \mu_x + \epsilon_x, \quad \text{where} \quad \epsilon_x \sim N(0, \sigma_x^2), \\ Y_{\text{Liab}} &= \mu_{\text{Liab}} + \epsilon_{\text{Liab}}, \quad \text{where} \quad \epsilon_{\text{Liab}} \sim N(0, \sigma_{\text{Liab}}^2), \\ Y_{\text{GDP}} &= \mu_{\text{GDP}} + \epsilon_{\text{GDP}}, \quad \text{where} \quad \epsilon_{\text{GDP}} \sim N(0, \sigma_{\text{GDP}}^2), \end{aligned}$$

and $(X, Y_{\text{Liab}}, Y_{\text{GDP}})$ follows joint normal distribution.

[1] For more detailed introduction of these tests, refer to Anderson (1952; 1962); Henzea and Zirkler (1990); Kolmogorov (1933); Mardia (1970); Royston (1991); Shapiro and Wilk (1965); Smirnov (1948).

Table 5.7: Summary of descriptive statistics for the data including the loss experience and liability of the private reinsurance program in Manitoba and the GDP of Canada. The data period is from 2001 to 2011.

Statistical Summary			
	Loss	Liability	GDP
Mean	147.91	1427.62	1453.08
Standard Deviation	100.99	430.32	209.66
Median	99.49	1338.60	1486.92
Skewness	0.66	0.50	-0.15
Kurtosis	-1.18	-1.30	-1.57
Correlation Coefficients			
	Loss	Liability	GDP
Loss	1	0.3513	0.3864
Liability	0.3513	1	0.7873
GDP	0.3864	0.7873	1

5.7.2 Reinsurance Premiums

In this subsection, we consider the reinsurance contract as defined in Equation (5.79) and derive analytical expressions of the reinsurance premiums under different premium principles. Proofs of the propositions in this section are relegated to Appendix 5A.

Expectation Premium Principle

Under the expectation premium principle, namely $\Pi^e = (1+\theta)\mathsf{E}(\tau\mathbf{L}_{[a,b]})$, we can calculate the reinsurance premium according to Proposition 5.7.1.

Proposition 5.7.1. *Under the expectation premium principle, the reinsurance contract* $\tau\mathbf{L}_{[a,b]}$*, defined as Equaiton* (5.79)*, has the premium* Π^e *expressed as*

$$\begin{aligned}\Pi^e &= \tau(1+\theta)\Bigg\{(b-a)+(\mu_x-b)\Phi\left(\frac{b-\mu_x}{\sigma_x}\right)-\sigma_x\phi\left(\frac{b-\mu_x}{\sigma_x}\right)\\ &\quad -(\mu_x-a)\Phi\left(\frac{a-\mu_x}{\sigma_x}\right)+\sigma_x\phi\left(\frac{a-\mu_x}{\sigma_x}\right)\Bigg\}. \end{aligned} \tag{5.80}$$

Standard Deviation Premium Principle

When we consider the standard deviation premium principle, i.e., $\Pi^{sd} = \mathsf{E}(\tau\mathbf{L}_{[a,b]})+\theta\sqrt{\mathsf{Var}(\tau\mathbf{L}_{[a,b]})}$, the reinsurance premium is calculated according to Proposition 5.7.2.

Table 5.8: Normality test results. Both univariate normality tests and multivariate normality tests are performed for the data. Significant level of the tests is 0.05.

Univariate Normality Test				
Test Name	Variable	Statistic	p-value	Result
"Shapiro-Wilk's"	Loss	0.8679	0.0729	✓
	Liability	0.8954	0.1623	✓
	GDP	0.9501	0.6448	✓
"Cramer-von Mises's"	Loss	0.1006	0.0955	✓
	Liability	0.0848	0.1600	✓
	GDP	0.0377	0.6924	✓
"Kolmogorov-Smirnov's"	Loss	0.2296	0.1078	✓
	Liability	0.2084	0.2004	✓
	GDP	0.1593	0.6069	✓
"Anderson-Darling's"	Loss	0.6121	0.0824	✓
	Liability	0.4964	0.1666	✓
	GDP	0.2452	0.6896	✓
Multivariate Normality Test				
Test Name		Statistic	p-value	Result
"Mardia's"	Skewness	7.7241	0.6558	✓
	Kurtosis	-0.9971	0.3187	
"Henze-Zirkler's"	HZ	0.6455	0.1506	✓
"Royston's"	H	5.1092	0.1579	✓

Proposition 5.7.2. *Under the standard deviation premium principle, the reinsurance contract* $\tau\mathbf{L}_{[a,b]}$*, defined as Equation (5.79), has the premium* Π^{sd} *expressed as*

$$\Pi^{sd} = \tau L_1 + \tau\theta\sqrt{L_2 - (L_1)^2}, \tag{5.81}$$

where

$$\begin{aligned} L_1 &= \mathsf{E}(\mathbf{L}_{[a,b]}) = (b-a) + (\mu_x - b)\Phi\left(\frac{b-\mu_x}{\sigma_x}\right) - \sigma_x\phi\left(\frac{b-\mu_x}{\sigma_x}\right) \\ &\quad -(\mu_x - a)\Phi\left(\frac{a-\mu_x}{\sigma_x}\right) + \sigma_x\phi\left(\frac{a-\mu_x}{\sigma_x}\right). \end{aligned} \tag{5.82}$$

$$\begin{aligned} L_2 &= \mathsf{E}\big((\mathbf{L}_{[a,b]})^2\big) = \sigma_x(\mu_x - a)\phi\left(\frac{a-\mu_x}{\sigma_x}\right) - \big((\mu_x-a)^2 + \sigma_x^2\big)\Phi\left(\frac{a-\mu_x}{\sigma_x}\right) + (b-a)^2 \\ &\quad +\big((\mu_x-a)^2 + \sigma_x^2 + (b-a)^2\big)\Phi\left(\frac{b-\mu_x}{\sigma_x}\right) + \sigma_x(2a-b-\mu_x)\phi\left(\frac{b-\mu_x}{\sigma_x}\right). \end{aligned} \tag{5.83}$$

Esscher's Premium Principle

In the context of Esscher premium principle, namely, $\Pi^{ess} = \tau\dfrac{\mathsf{E}\big(\exp(\theta\tau\mathbf{L}_{[a,b]})\mathbf{L}_{[a,b]}\big)}{\mathsf{E}\big(\exp(\theta\tau\mathbf{L}_{[a,b]})\big)}$, the reinsurance premium can be written according to Proposition 5.7.3.

Proposition 5.7.3. *Under the expectation premium principle, the reinsurance contract* $\tau\mathbf{L}_{[a,b]}$*, defined as Equaiton (5.79), has the premium* Π^{ess} *expressed as*

$$\begin{aligned} \Pi^{ess} = \frac{\begin{array}{l} e^{\theta_\tau(\mu_x-a)+\frac{1}{2}\theta_\tau^2\sigma_x^2}\big\{\sigma_x\big[\phi\big(\frac{a-\mu_x-\theta_\tau\sigma_x^2}{\sigma_x}\big) - \phi\big(\frac{b-\mu_x-\theta_\tau\sigma_x^2}{\sigma_x}\big)\big] + \\ (\mu_x - a + \theta_\tau\sigma_x^2)\big[\Phi\big(\frac{b-\mu_x-\theta_\tau\sigma_x^2}{\sigma_x}\big) - \Phi\big(\frac{a-\mu_x-\theta_\tau\sigma_x^2}{\sigma_x}\big)\big]\big\} + (b-a)e^{\theta_\tau(b-a)}\big[1-\Phi\big(\frac{b-\mu_x}{\sigma_x}\big)\big] \end{array}}{\begin{array}{l} \Phi\big(\frac{a-\mu_x}{\sigma_x}\big) + e^{\theta_\tau(b-a)}\big[1-\Phi\big(\frac{b-\mu_x}{\sigma_x}\big)\big] \\ \quad + e^{\theta_\tau(b-a)+\frac{1}{2}\sigma_x^2\theta_\tau^2}\big[\Phi\big(\frac{b-\mu_x-\theta_\tau\sigma_x^2}{\sigma_x}\big) - \Phi\big(\frac{a-\mu_x-\theta_\tau\sigma_x^2}{\sigma_x}\big)\big] \end{array}} \end{aligned} \tag{5.84}$$

where $\theta_\tau = \tau\theta$*.*

Distortion Premium Principle

Recall that the distortion premium principle is defined as $\Pi^d = \int_0^\infty g\big(S_{\mathbf{L}_{[a,b]}}(u)\big)du$. Also note that the survival function of random variable, $\mathbf{L}_{[a,b]}$, can be written as

$$S_{\mathbf{L}_{[a,b]}}(u) = \begin{cases} S_x(u+a) & 0 \le u < b-a, \\ 0 & u \ge b-a, \end{cases} \tag{5.85}$$

therefore,

$$\Pi^d = \int_a^b g\big(S_X(u)\big)du. \tag{5.86}$$

In this study, we consider a class of distortion function called Beta transform, defined as

$$g_{a,b}(u) \;=\; \beta(a,b;u) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^u t^{a-1}(1-t)^{b-1}dt. \tag{5.87}$$

This distortion function has the form of incomplete beta function (Hogg and Klugman, 1984). Wirch and Hardy (1999) consider the beta distortion family in the context of distortion risk measures and discuss its advantage of utilising the whole loss distribution rather than focusing entirely on the tail as CTE, which is desirable for capital adequacy.

Moreover, the Beta distortion family includes two important distortion functions that are widely used in the insurance premium principles as its special cases: PH transform and Dual power transform. To be more specific, if we set $a = 1$, we get dual power transform with $\theta = b$:

$$g_\theta(u) \;=\; 1-(1-u)^\theta, \quad \theta \geq 1; \tag{5.88}$$

while if we set $b = 1$, we get PH transform with $\theta = 1/a$:

$$g_\theta(u) \;=\; u^{\frac{1}{\theta}}, \quad \theta \geq 1. \tag{5.89}$$

Multivariate Weighted Premium

We derive the reinsurance premium under a general multivariate normal distribution setting. Assume that our k-dimensional auxiliary vector follows a joint normal distribution, namely, $\boldsymbol{Y} = (Y_1, \ldots, Y_k)^T \sim N_k(\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y})$ and further assume that X and $\boldsymbol{Y}$ also follow multivariate normal distribution, namely,

$$(\boldsymbol{Y}, X) \sim N_{k+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

where

$$\begin{aligned}
\boldsymbol{\mu_Y} &= (\mu_1, \ldots, \mu_k)^T, \\
\boldsymbol{\mu} &= (\boldsymbol{\mu_Y}, \mu_x)^T = (\mu_1, \ldots, \mu_k, \mu_x)^T, \\
\boldsymbol{\Sigma} &= \begin{pmatrix} \boldsymbol{\Sigma_Y}, & \boldsymbol{\sigma_{YX}} \\ \boldsymbol{\sigma_{YX}^T} & \sigma_x^2 \end{pmatrix}.
\end{aligned}$$

let us now denote $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)^T, \boldsymbol{y} = (y_1, \ldots, y_k)^T$. In this chapter we are particularly interested in two types of weighting functions, linear weighting function and exponential weighting function.

- **Linear weighting function:**

$$w_l(\mathbf{y}) \quad = \quad \alpha_0 + \boldsymbol{\alpha}^T \boldsymbol{y}. \tag{5.90}$$

- **Exponential weighting function:**

$$w_e(\mathbf{y}) \quad = \quad e^{\alpha_0 + \boldsymbol{\alpha}^T \boldsymbol{y}}. \tag{5.91}$$

The k-dimensional weighted reinsurance premiums with the two types of weighting functions are expressed in Proposition 5.7.4 and Proposition 5.7.5.

Proposition 5.7.4. *Under the multivariate weighted premium principle, with the weighting random vector* $\mathbf{Y}$ *and the weighting function* $w_l(\boldsymbol{y})$ *in Equation (5.90), the reinsurance contract* $\tau \mathbf{L}_{[a,b]}$*, defined as Equaiton (5.79), has the premium* $\Pi^{w_l}(X, \mathbf{Y})$ *expressed as*

$$\begin{aligned} \Pi^{w_l}(X, \mathbf{Y}) \quad = \quad & (\mu_x - a + \frac{\sigma_x^2 \beta_1}{\alpha_0 + \boldsymbol{\alpha}^T \boldsymbol{\mu_Y}}) \left\{ \Phi\left(\frac{b - \mu_x}{\sigma_x} \right) - \Phi\left(\frac{a - \mu_x}{\sigma_x} \right) \right\} \\ & (b - a) \left\{ 1 - \Phi\left(\frac{b - \mu_x}{\sigma} \right) \right\} + \sigma_x \left\{ \phi\left(\frac{a - \mu_x}{\sigma_x} \right) - \phi\left(\frac{b - \mu_x}{\sigma_x} \right) \right\}, \end{aligned} \tag{5.92}$$

where

$$\beta_1 \quad = \quad \frac{\boldsymbol{\alpha}^T \boldsymbol{\sigma_{YX}}}{\sigma_x^2}.$$

Proposition 5.7.5. *Under the multivariate weighted premium principle, with the weighting random vector* $\mathbf{Y}$ *and the weighting function* $w_e(\boldsymbol{y})$ *in Equation (5.91), the reinsurance contract* $\tau \mathbf{L}_{[a,b]}$*, defined as Equaiton (5.79), has the premium* $\Pi^{w_e}(X, \mathbf{Y})$ *expressed as*

$$\begin{aligned} \Pi^{w_e}(X, Y) \quad = \quad & \sigma_x \big[\phi\left(\frac{a - \mu_x - \frac{\boldsymbol{\alpha}^T \boldsymbol{\sigma_{YX}}}{\sigma_x}}{\sigma_x} \right) - \phi\left(\frac{b - \mu_x - \frac{\boldsymbol{\alpha}^T \boldsymbol{\sigma_{YX}}}{\sigma_x}}{\sigma_x} \right) \big] \\ & + (\mu_x - a + \boldsymbol{\alpha}^T \boldsymbol{\sigma_{YX}}) \big[\Phi\left(\frac{b - \mu_x - \frac{\boldsymbol{\alpha}^T \boldsymbol{\sigma_{YX}}}{\sigma_x}}{\sigma_x} \right) - \Phi\left(\frac{a - \mu_x - \frac{\boldsymbol{\alpha}^T \boldsymbol{\sigma_{YX}}}{\sigma_x}}{\sigma_x} \right) \big] \\ & + (b - a) \big[1 - \Phi\left(\frac{b - \mu_x - \frac{\boldsymbol{\alpha}^T \boldsymbol{\sigma_{YX}}}{\sigma_x}}{\sigma_x} \right) \big]. \end{aligned} \tag{5.93}$$

5.8 Empirical Results

5.8.1 Parameter Estimation

In this section, we estimate the parameters for different premium principles presented in Section 5.7.2 using least square method. The estimated results are listed in Table 5.9. It is interesting to note from the estimated results that for the distortion premium principle, a is estimated to be 1, so that the Beta distortion function degenerates to the dual power distortion function. In fact, we can get exactly the same estimating result when we use the dual power distortion function. This result shows that compared to the PH transform, the dual power distortion function is more suitable to our data.

Table 5.9: Estimating results for different premium principles using least square method.[2].

Premium Principles	Parameter	Value
Expectation	θ	0.0722
Standard Deviation	θ	0.1753
Esscher's	θ	0.1963
Distortion	a	1.0000
	b	1.1492
Linear Weighted	α_1	45.1216
	α_2	7.3767
Exponential Weighted	α_1	3.6628
	α_2	5.7589

To reduce the number of parameters to estimate, we restrict $\alpha_0 = 0$ in the linear weighted premium and the exponential weighted premium.

5.8.2 Pricing Results

In this subsection, we study the features of different premium principles. Let us consider the XoL reinsurance contract, $\mathbf{L}_{[a,a+h)}$, defined in Equation (5.79), where b is set to be $a+h$. We increase the attachment level a from 0 up to 550 with 50 increments and fix $h = 50$. The very last contract has no limit, therefore the reinsurers will cover every loss greater than 550. Using the parameters estimated in Section 5.8.1, the pricing results are displayed in Table 5.10 and Figure 5.1.

One of the most important insights to our empirical pricing results is that by integrating information from the auxiliary variables, for example the liabilities and Canadian GDP as in this study, the resulting Weighted Premium Principle is able to adjust more risk loading at higher layers of reinsurance contracts. In contrast, the Expectation Premium Principle attaches too high loading at lower layers while not enough risk loading at higher layers. For example, for the first contract, $\mathbf{L}_{[0,50)}$, the premiums is 52.3616 under the Expectation Premium Principle and 50.0203 under the Standard Deviation Premium Principle. However, it might be unreasonable to charge higher than the capped risk. The Exponential Weighted Principle gives reasonable loading for lower layers while increases the risk loading for higher layers. Comparing these premium principles, we find the Esscher's Premium Principle charges lowest premiums and the Expectation Premium Principle charges the second lowest. Distortion, Linear Weighted and Standard Deviation Premium Principle have similar risk loading on the layer contracts, which are higher than Esscher's and Expectation Premium Principles and lower than the Exponential Weighted Premium Principle.

Table 5.10: Pricing results for different layer contracts under each premium principle. Each row shows the layer of the contract while the columns show the results for different premium principles. "Expectation" stands for Expectation Premium Principle, "SD" represents Standard Deviation Premium Principle, "Esscher" stands for Esscher's Premium Principle, "Distortion" is for Distortion Premium Principle, "L-Weighted" means Linear Weighted Premium Principle, and "E-Weighted" stands for Exponential Weighted Premium Principle.

Contract	Expectation	SD	Esscher	Distortion	L-Weighted	E-Weighted
$[0, 50)$	52.3616	50.0203	48.9200	49.3281	49.2954	49.5786
$[50, 100)$	50.0255	48.6513	46.9001	47.7516	47.7404	48.5630
$[100, 150)$	45.2246	45.1156	42.7122	44.0454	44.1845	46.0326
$[150, 200)$	37.4290	38.6477	35.7847	37.3487	37.8255	41.0522
$[200, 250)$	27.4260	29.6621	26.6419	28.0297	28.9159	33.3066
$[250, 300)$	17.2827	19.9288	17.0605	18.0145	19.1207	23.7875
$[300, 350)$	9.1545	11.5845	9.1485	9.6738	10.6619	14.5425
$[350, 400)$	4.0074	5.8421	4.0314	4.2689	4.9194	7.4469
$[400, 450)$	1.4321	2.6058	1.4431	1.5315	1.8528	3.1434
$[450, 500)$	0.4141	1.0629	0.4167	0.4436	0.5642	1.0812
$[500, 550)$	0.0963	0.4102	0.0966	0.1032	0.1379	0.3005
$[550, \infty)$	0.0209	0.1952	0.0216	0.0224	0.0317	0.0810

As a special section in Property & Casualty (re)insurance, agricultural reinsurance also pays special attention to managing catastrophic losses due to disasters leading to extremely high layers. In fact, agricultural insurers and reinsurers bear higher loss ratios than other

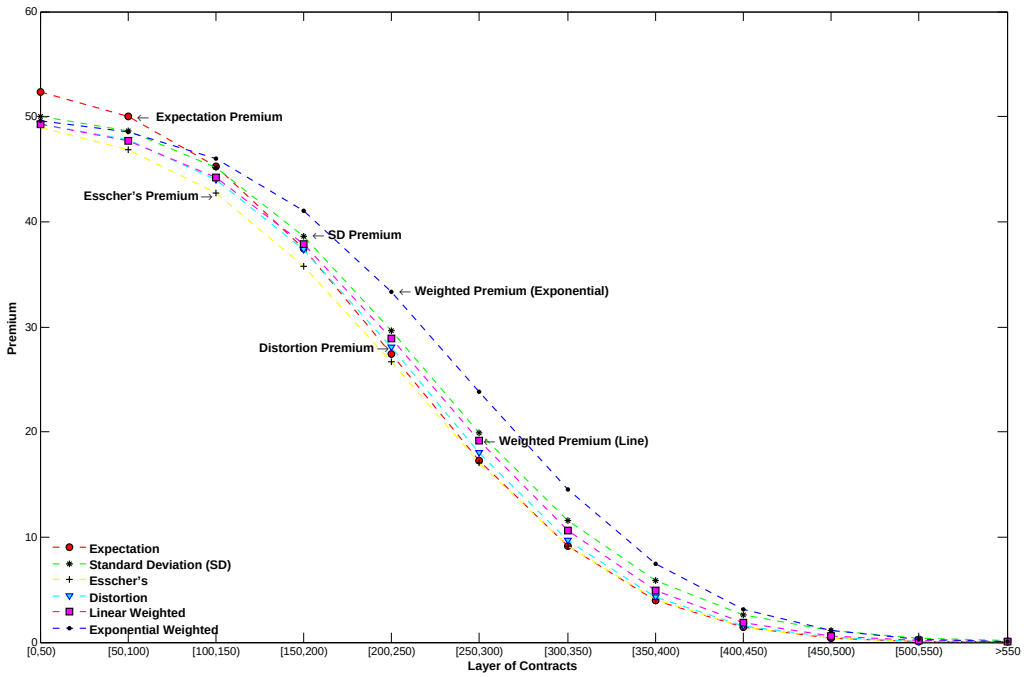


Figure 5.1: Premiums for different layer contracts under each premium principles.

lines of business in P & C sectors due to additional exposure to weather risks and spatially correlation geographical risks. Therefore, it is of great interest to study the relative risk loading (RRL) at higher layers of the contracts (Definition 5.4.3). As stated before, for any premium principle, the premium for the risk X, $\Pi(X)$, can always be written as

$$\Pi(X) \quad = \quad \mathsf{E}(X) + \theta(X), \tag{5.94}$$

where $\theta(X)$ is the risk loading of the premium $\Pi(X)$. Therefore, $\text{RRL} - 1 = \lim_{h \to 0} \dfrac{\Theta(\mathbf{L}_{(x,x+h]})}{\mathsf{E}(\mathbf{L}_{(x,x+h]})}$ compares the risk loading at each level of contract layer to the net premium. Higher RRL indicates more risk-adjusted loading to the layer relative to the net premium.

In our empirical analysis, we select a small value of $h = 10$ and display the results for the RRL in Figure 5.2. We can see that the Expectation Premium Principle has a constant RRL $= 1.0722$, which is the value of θ in Equation (5.80). This means that the Expectation Premium Principle fails to allow more relative loading at higher levels. The RRL of Esscher's Premium Principle is vary close to1 and stays almost the same for all the layers of contracts. Actually, as shown in Corollary 5.4.2, when $h \mapsto 0$, RRL of Esscher's transform is 1. It is disadvantageous of the Esscher's Premium Principle because essentially it does not charge

any extra loading to the risk in addition to the net premium for higher layers.

It is also interesting to compare the RRL values of the Standard Deviation Premium Principle (RRL_{SD}) and the Exponential Weighted Premium Principle ($\text{RRL}_{\text{E-Weighted}}$). We find that at lower layers $\text{RRL}_{\text{SD}} > \text{RRL}_{\text{E-Weighted}}$. However, at some point the two RRL lines come across and then $\text{RRL}_{\text{SD}} < \text{RRL}_{\text{E-Weighted}}$. Therefore, Exponential Weighted Premium Principle is able to assign higher Relative loading at higher layers compared to Standard Deviation Premium Principle.

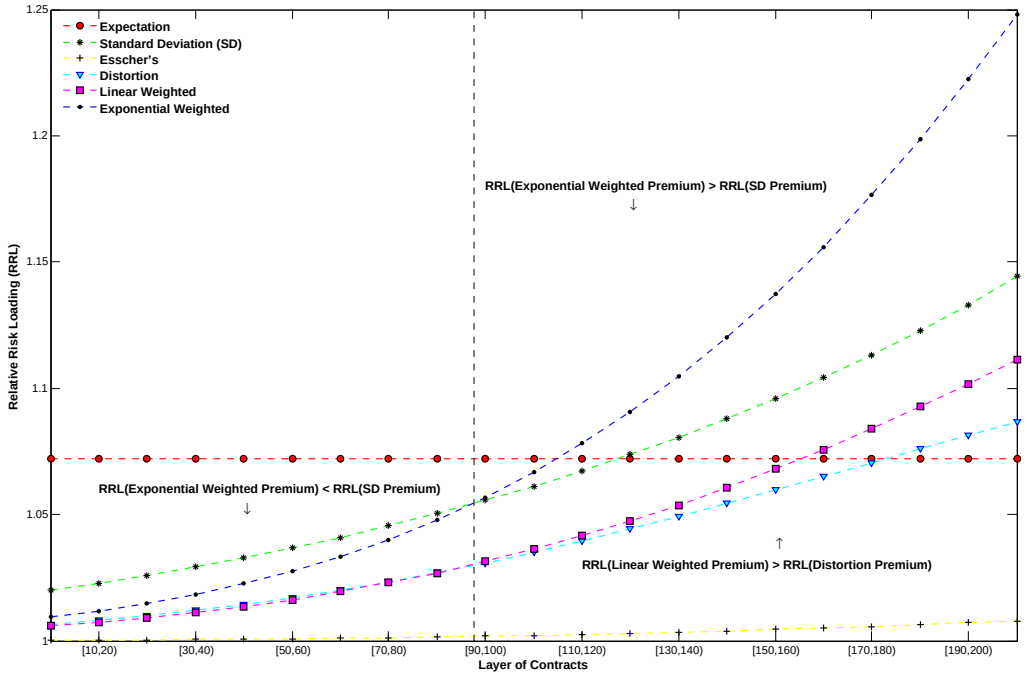


Figure 5.2: Relative risk loading for different layer contracts under each premium principles.

5.8.3 Profit and Loss Analysis

In this section, we perform a simple case study with a hypothetical long term profit & loss analysis for the reinsurance companies based on each premium principle. To proceed, we assume that in each of the future 10 years (i.e., 2012 - 2021), the reinsurers collect premiums, pay the indemnities, and invest the profits in stocks offering lognormal returns, with annual parameters $\mu_S = 0.05$ and $\sigma_S = 0.2$. We further assume that the risk free rate

of return is $r_f = 0.9\%$[3]. Figure 5.3 shows the simulated densities of the 10-year profits of the reinsurance companies under different premium principles, with negative values indicating loss and positive values indicating profits. Statistical summary and the histograms of the simulation results are summarized in Table 5.11 and Figure 5.4.

A major objective of agricultural reinsurance companies is to stabilize their cash-flows and long-term revenue in the presence of adverse weather events. The profit & loss analysis results highlight the advantage of the new proposed multivariate weighted premium principle for reinsurers to achieve this goal. As shown in Figure 5.3, the density curve of exponential weighted premium principle locates on the right of the the other densities. Table 5.11 shows that the exponential weighted premium principle is able to achieve the highest VaR and CTE compared to the other premium principles. This is mostly because, as discussed in Section 5.8.2, by integrating auxiliary variables into reinsurance pricing, multivariate weighted premium principle is able to assign higher risk loadings to more risky contracts, therefore, in general attain better profit to reinsurance companies in the long-run.

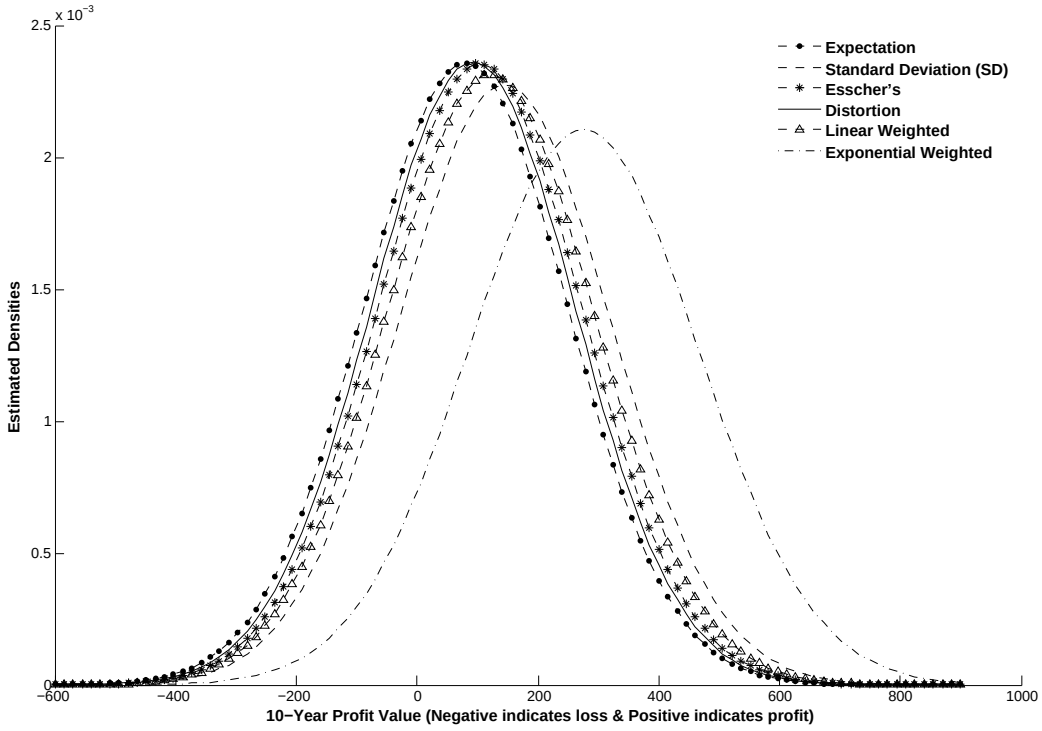


Figure 5.3: 10-year profit under each premium principles.

[3]The parameters are selected as a enlightening example. The risk free rate of return is based on the annual average yield of Treasury bill in Canada in 2014. Other appropriate discount factor can also be selected.

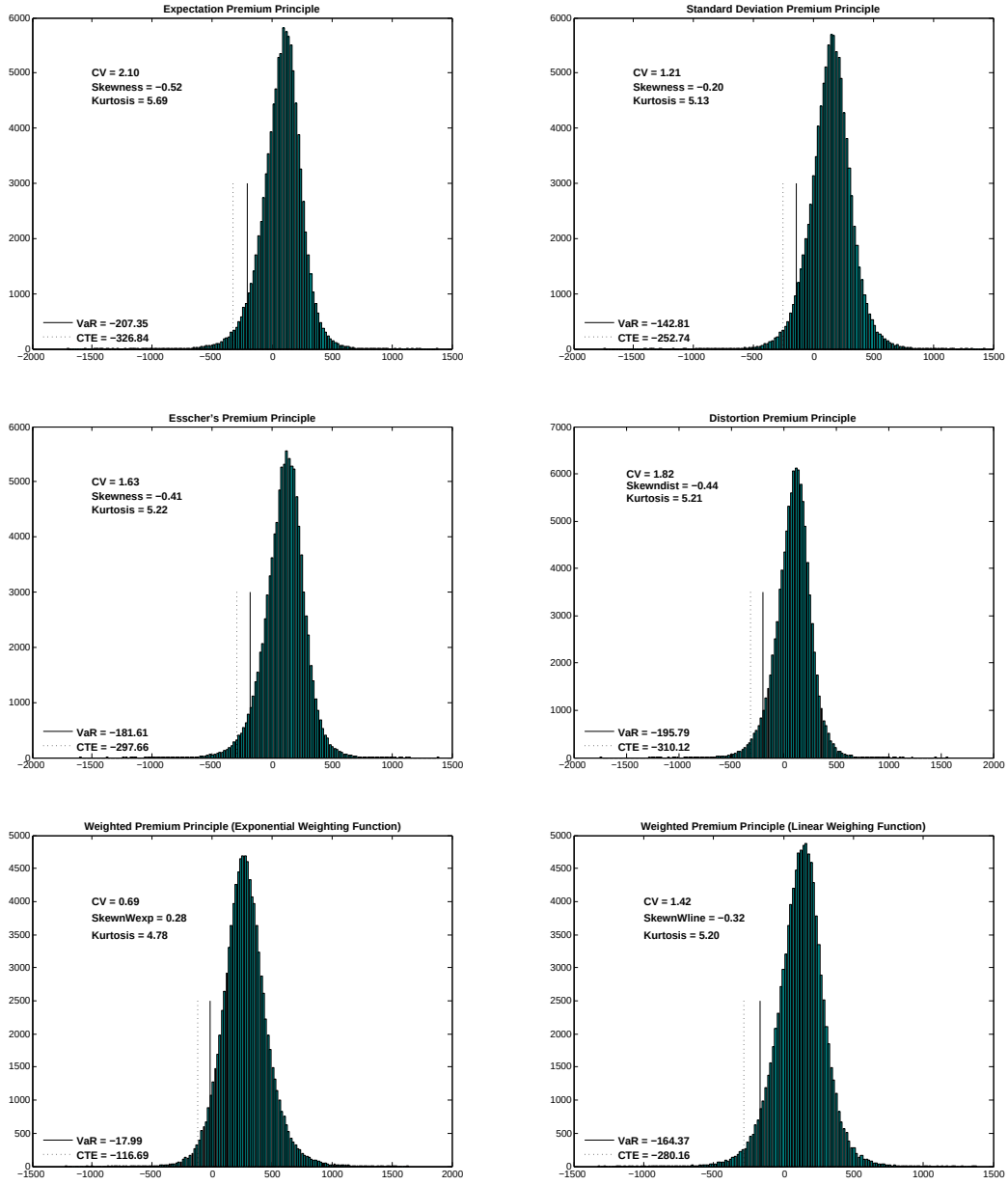


Figure 5.4: Histograms of the profits under each premium principles.

Table 5.11: Summary statistics of the profit simulation results (Unit: Million CAD$). Results are based on 100,000 times simulation.

	Mean	SD	CV	Skewness	Kurtosis	VaR	CTE
Expectation	80.52	169.19	2.10	-0.52	5.69	-207.35	-326.84
Standard Deviation	144.88	175.22	1.21	-0.20	5.13	-142.81	-252.74
Esscher	104.15	169.36	1.63	-0.41	5.22	-181.61	-297.66
Distortion	92.79	168.98	1.82	-0.44	5.21	-195.79	-310.12
Linear Weighted	121.29	172.36	1.42	-0.32	5.20	-164.37	-280.16
Exponential Weighted	275.52	189.55	0.69	0.28	4.78	-17.99	-116.69

5.9 Conclusion

Previous research indicates that it is necessary to integrate auxiliary variables into the ratemaking process to refine the pricing framework. This chapter proposes a new premium principle based on multivariate weighted distribution to provide a formal approach of weighting auxiliary variables in the historical loss experience. Some desirable properties of the premium based on multivariate weighted distributions are derived. In addition, the economic premium principle discussed in Bühlmann (1980) and the Esscher's premium are special cases of the framework proposed in this chapter. An application to the reinsurance experience in Manitoba from 2001 to 2011 reveals that integrating liability and economic variables into the pricing framework redistributes premium rates, assigning higher loadings to more risky layers of contracts and hence achieves more sustainable long-term profits.

5A Appendix: Proofs of Propositions in Section 5.7

5A.1 Proof of Proposition 5.7.1

First we define the indicator function $\mathbb{I}_A$ as

$$\mathbb{I}_A = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases} \tag{5A.95}$$

Note that $\mathsf{E}(\mathbf{L}_{[a,b]}) = \mathsf{E}\big((X-a)\mathbb{I}_{\{a\le X<b\}}\big)+\mathsf{E}\big((b-a)\mathbb{I}_{\{X\ge b\}}\big)$, and $X \sim N(\mu_x, \sigma_x^2)$, hence,

$$\begin{aligned}
\mathsf{E}\big((X-a)\mathbb{I}_{\{a\le X<b\}}\big) &= \int_a^b (x-a)\frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}}\,dx \\
&\overset{y=\frac{x-\mu_x}{\sigma_x}}{=} \int_{\frac{a-\mu_x}{\sigma_x}}^{\frac{b-\mu_x}{\sigma_x}} (\mu_x - a + \sigma_x y)\frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\,dy \\
&= \sigma_x[\phi\left(\frac{a-\mu_x}{\sigma_x}\right) - \phi\left(\frac{b-\mu_x}{\sigma_x}\right)] \\
&\quad +(\mu_x - a)[\Phi\left(\frac{b-\mu_x}{\sigma_x}\right) - \Phi\left(\frac{a-\mu_x}{\sigma_x}\right)]. \qquad (5A.96)
\end{aligned}$$

$$\mathsf{E}\big((b-a)\mathbb{I}_{\{X\ge b\}}\big) = (b-a)[1-\Phi\left(\frac{b-\mu_x}{\sigma_x}\right)]. \qquad (5A.97)$$

$$\begin{aligned}
\mathsf{E}(\mathbf{L}_{[a,b]}) &= (b-a) + (\mu_x - b)\Phi\left(\frac{b-\mu_x}{\sigma_x}\right) - \sigma_x\phi\left(\frac{b-\mu_x}{\sigma_x}\right) \\
&\quad -(\mu_x - a)\Phi\left(\frac{a-\mu_x}{\sigma_x}\right) + \sigma_x\phi\left(\frac{a-\mu_x}{\sigma_x}\right). \qquad (5A.98)
\end{aligned}$$

Therefore, the reinsurance premium under the expectation premium principle is

$$\begin{aligned}
\Pi^e &= \tau(1+\theta)\mathsf{E}(\mathbf{L}_{[a,b]}) \\
&= \tau(1+\theta)\Big\{(b-a) + (\mu_x - b)\Phi\left(\frac{b-\mu_x}{\sigma_x}\right) - \sigma_x\phi\left(\frac{b-\mu_x}{\sigma_x}\right) \\
&\quad -(\mu_x - a)\Phi\left(\frac{a-\mu_x}{\sigma_x}\right) + \sigma_x\phi\left(\frac{a-\mu_x}{\sigma_x}\right)\Big\}. \qquad (5A.99)
\end{aligned}$$

5A.2 Proof of Proposition 5.7.2

Note that $\mathsf{E}\big((\mathbf{L}_{[a,b]})^2\big) = \mathsf{E}\big((X-a)^2\mathbb{I}_{\{a\le X<b\}}\big) + \mathsf{E}\big((b-a)^2\mathbb{I}_{\{X\ge b\}}\big)$, hence,

$$\begin{aligned}
\mathsf{E}\big((X-a)^2\mathbb{I}_{\{a\le X<b\}}\big) &= \int_a^b x^2\frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}}\,dx - 2a\int_a^b x\frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}}\,dx \\
&\quad +a^2\int_a^b \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}}\,dx.
\end{aligned}$$

Specially,

$$\begin{aligned}\int_a^b x^2 \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} dx &= (\mu_x^2+\sigma_x^2)\left\{\Phi\left(\frac{b-\mu_x}{\sigma_x}\right)-\Phi\left(\frac{a-\mu_x}{\sigma_x}\right)\right\} \\ &\quad +\sigma_x(a+\mu_x)\phi\left(\frac{a-\mu_x}{\sigma_x}\right)-\sigma_x(b+\mu_x)\phi\left(\frac{b-\mu_x}{\sigma_x}\right) \quad (5A.100)\\ \int_a^b x \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} dx &= \mu_x\left\{\Phi\left(\frac{b-\mu_x}{\sigma_x}\right)-\Phi\left(\frac{a-\mu_x}{\sigma_x}\right)\right\} \\ &\quad -\sigma_x\left\{\phi\left(\frac{b-\mu_x}{\sigma_x}\right)-\phi\left(\frac{a-\mu_x}{\sigma_x}\right)\right\}. \quad (5A.101)\end{aligned}$$

Therefore,

$$\begin{aligned}\mathsf{E}\big((X-a)^2\mathbb{I}_{\{a\le X<b\}}\big) &= \big((\mu_x-a)^2+\sigma_x^2\big)\left\{\Phi\left(\frac{b-\mu_x}{\sigma_x}\right)-\Phi\left(\frac{a-\mu_x}{\sigma_x}\right)\right\} \\ &\quad +\phi\left(\frac{b-\mu_x}{\sigma_x}\right)(2a-b-\mu_x)\sigma_x+\phi\left(\frac{a-\mu_x}{\sigma_x}\right)(\mu_x-a)\sigma_x, \quad (5A.102)\end{aligned}$$

and hence,

$$\begin{aligned}\mathsf{E}\big((\mathbf{L}_{[a,b]})^2\big) &= \big((\mu_x-a)^2+\sigma_x^2\big)\Phi\left(\frac{b-\mu_x}{\sigma_x}\right)+\sigma_x(\mu_x-a)\phi\left(\frac{a-\mu_x}{\sigma_x}\right)+(b-a)^2 \\ &\quad -\big((\mu_x-a)^2+\sigma_x^2+(b-a)^2\big)\Phi\left(\frac{a-\mu_x}{\sigma_x}\right)+\sigma_x(2a-b-\mu_x)\phi\left(\frac{b-\mu_x}{\sigma_x}\right) \quad (5A.103)\end{aligned}$$

For notation simplicity, let us denote $L_1 = \mathsf{E}(\mathbf{L}_{[a,b]})$ and $L_2 = \mathsf{E}\big((\mathbf{L}_{[a,b]})^2\big)$, then $\mathsf{Var}(\mathbf{L}_{[a,b]}) = L_2 - (L_1)^2$, and combining Equation (5A.98) and (5A.103), we can see that Equation (5.81) holds.

5A.3 Proof of Proposition 5.7.3

Recall that

$$\Pi^{ess} = \tau \frac{\mathsf{E}\big(\exp(\theta_\tau \mathbf{L}_{[a,b]})\mathbf{L}_{[a,b]}\big)}{\mathsf{E}\big(\exp(\theta_\tau \mathbf{L}_{[a,b]})\big)}, \quad (5A.104)$$

where $\theta_\tau = \theta\tau$. We calculate the two parts in the fraction separately.

$$\begin{aligned}
\mathsf{E}\big(\exp(\theta_\tau \mathbf{L}_{[a,b]})\mathbf{L}_{[a,b]}\big) &= \int_a^b (x-a)e^{\theta_\tau(x-a)}\frac{1}{\sqrt{2\pi}\sigma_x}e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}}dx + (b-a)e^{\theta_\tau(b-a)}\mathsf{E}\big(\mathbb{I}_{\{X\geq b\}}\big),\\
\mathsf{E}\big(\exp(\theta_\tau \mathbf{L}_{[a,b]})\big) &= \mathsf{E}\big(\mathbb{I}_{\{X<a\}}\big) + \int_a^b e^{\theta_\tau(x-a)}\frac{1}{\sqrt{2\pi}\sigma_x}e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}}dx + e^{\theta_\tau(b-a)}\mathsf{E}\big(\mathbb{I}_{\{X\geq b\}}\big).
\end{aligned}$$

Specially, consider $\int_a^b (x-a)e^{\theta_\tau(x-a)}\frac{1}{\sqrt{2\pi}\sigma_x}e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}}dx =$

$$\begin{aligned}
&\int_a^b (x-\mu_x)e^{\theta_\tau(x-a)}\frac{1}{\sqrt{2\pi}\sigma_x}e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}}dx + \int_a^b (\mu_x-a)e^{\theta_\tau(x-a)}\frac{1}{\sqrt{2\pi}\sigma_x}e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}}dx\\
\overset{y=\frac{x-\mu_x}{\sigma_x}}{=}\ & e^{\theta_\tau(\mu_x-a)}\int_{\frac{a-\mu_x}{\sigma_x}}^{\frac{b-\mu_x}{\sigma_x}} \big(\sigma_x(y-\sigma_x\theta_\tau) + (\mu_x - a + \theta_\tau\sigma_x^2)\big)\frac{1}{\sqrt{2\pi}}e^{-\frac{(y-\theta_\tau\sigma_x)^2}{2}+\frac{1}{2}\theta_\tau^2\sigma_x^2}dy\\
=\ & \sigma_x e^{\theta_\tau(\mu_x-a)+\frac{1}{2}\theta_\tau^2\sigma_x^2}\int_{a_1}^{b_1} v\frac{1}{\sqrt{2\pi}}e^{-\frac{v^2}{2}}dv + (\mu_x-a+\theta_\tau\sigma_x^2)e^{\theta_\tau(\mu_x-a)+\frac{1}{2}\theta_\tau^2\sigma_x^2}\int_{a_1}^{b_1}\frac{1}{\sqrt{2\pi}}e^{-\frac{v^2}{2}}dv\\
& \text{where}\quad v = y-\theta_\tau\sigma_x,\ a_1 = \frac{a-\mu_x-\theta_\tau\sigma_x^2}{\sigma_x}\ \text{and}\ b_1 = \frac{b-\mu_x-\theta_\tau\sigma_x^2}{\sigma_x},\\
=\ & e^{\theta_\tau(\mu_x-a)+\frac{1}{2}\theta_\tau^2\sigma_x^2}\big[\sigma_x\big(\phi(a_1)-\phi(b_1)\big) + (\mu_x-a+\theta_\tau\sigma_x^2)\big(\Phi(b_1)-\Phi(a_1)\big)\big], \qquad \text{(5A.105)}
\end{aligned}$$

and similarly

$$\int_a^b e^{\theta_\tau(x-a)}\frac{1}{\sqrt{2\pi}\sigma_x}e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}}dx = e^{\theta_\tau(\mu_x-a)+\frac{1}{2}\theta_\tau^2\sigma_x^2}\big(\Phi(b_1)-\Phi(a_1)\big). \qquad \text{(5A.106)}$$

Therefore,

$$\begin{aligned}
\mathsf{E}\big(\exp(\theta_\tau \mathbf{L}_{[a,b]})\mathbf{L}_{[a,b]}\big) &= e^{\theta_\tau(\mu_x-a)+\frac{1}{2}\theta_\tau^2\sigma_x^2}\Bigg\{\sigma_x\Big[\phi\Big(\frac{a-\mu_x-\theta_\tau\sigma_x^2}{\sigma_x}\Big) - \phi\Big(\frac{b-\mu_x-\theta_\tau\sigma_x^2}{\sigma_x}\Big)\Big]\\
&\quad +(\mu_x-a+\theta_\tau\sigma_x^2)\Big[\Phi\Big(\frac{b-\mu_x-\theta_\tau\sigma_x^2}{\sigma_x}\Big) - \Phi\Big(\frac{a-\mu_x-\theta_\tau\sigma_x^2}{\sigma_x}\Big)\Big]\Bigg\}\\
&\quad +(b-a)e^{\theta_\tau(b-a)}\Big[1-\Phi\Big(\frac{b-\mu_x}{\sigma_x}\Big)\Big], \qquad \text{(5A.107)}\\
\mathsf{E}\big(\exp(\theta_\tau \mathbf{L}_{[a,b]})\big) &= e^{\theta_\tau(\mu_x-a)+\frac{1}{2}\theta_\tau^2\sigma_x^2}\Big[\Phi\Big(\frac{b-\mu_x-\theta_\tau\sigma_x^2}{\sigma_x}\Big) - \Phi\Big(\frac{a-\mu_x-\theta_\tau\sigma_x^2}{\sigma_x}\Big)\Big]\\
&\quad +\Phi\Big(\frac{a-\mu_x}{\sigma_x}\Big) + e^{\theta_\tau(b-a)}\Big[1-\Phi\Big(\frac{b-\mu_x}{\sigma_x}\Big)\Big]. \qquad \text{(5A.108)}
\end{aligned}$$

Combining Equaitions (5A.104), (5A.107) and (5A.108), we can get the expression of the premium as in Equation (5.84).

5A.4 Proof of Proposition 5.7.4

Since $Y \sim N(\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y})$, obviously,

$$\mathsf{E}(w_l(Y)) \;=\; \alpha_0 + \boldsymbol{\alpha}^T \boldsymbol{\mu_Y}. \tag{5A.109}$$

Also note that $\mathsf{E}(w_l(Y)\mathbf{L}_{[a,b]}) = \mathsf{E}(w_l(Y)(X-a)\mathbb{I}_{\{a\le X<b\}}) + \mathsf{E}(w_l(Y)(b-a)\mathbb{I}_{\{X\ge b\}})$, we consider the two parts separately. To proceed, let us first consider the distribution of Y conditional on $X = x$. It is not difficult to show that $Y|X = x \sim N(\breve{\mu}, \breve{\Sigma})$, where $\breve{\mu} = \boldsymbol{\mu_Y} + \frac{\boldsymbol{\sigma_{YX}}}{\sigma_x^2}(x-\mu_x)$ and $\breve{\Sigma} = \boldsymbol{\Sigma_Y} - \frac{\boldsymbol{\sigma_{YX}}\boldsymbol{\sigma_{YX}}^T}{\sigma_x^2}$. Therefore, $\alpha_0 + \boldsymbol{\alpha}^T Y|X = x \sim N(\widetilde{\mu}, \widetilde{\Sigma})$, where $\widetilde{\mu} = \alpha_0 + \boldsymbol{\alpha}^T\breve{\mu}, \widetilde{\Sigma} = \boldsymbol{\alpha}^T\breve{\Sigma}\boldsymbol{\alpha}$. Also note that $\alpha_0 + \boldsymbol{\alpha}^T\breve{\mu} = \alpha_0 + \boldsymbol{\alpha}^T\boldsymbol{\mu_Y} + \frac{\boldsymbol{\alpha}^T\boldsymbol{\sigma_{YX}}}{\sigma_x^2}(x-\mu_x)$, for notation simplicity, we denote $\beta_0 = \alpha_0 + \boldsymbol{\alpha}^T\boldsymbol{\mu_Y} - \frac{\boldsymbol{\alpha}^T\boldsymbol{\sigma_{YX}}}{\sigma_x^2}\mu_x$ and $\beta_1 = \frac{\boldsymbol{\alpha}^T\boldsymbol{\sigma_{YX}}}{\sigma_x^2}$, and hence $\widetilde{\mu} = \beta_0 + \beta_1 x$. Therefore,

$$\begin{aligned}
\mathsf{E}(w_l(Y)(X-a)\mathbb{I}_{\{a\le X<b\}}) &= \mathsf{E}\big((X-a)\mathbb{I}_{\{a\le X<b\}}\mathsf{E}(w_l(Y)|X))\big) \\
&= \mathsf{E}\big((\beta_0+\beta_1 X)(X-a)\mathbb{I}_{\{a\le X<b\}}\big) \\
&= (\beta_0 + a\beta_1)\mathsf{E}\big((X-a)\mathbb{I}_{\{a\le X<b\}}\big) + \beta_1\mathsf{E}\big((X-a)^2\mathbb{I}_{\{a\le X<b\}}\big).
\end{aligned}$$

Also note that the expressions of $\mathsf{E}\big((X-a)\mathbb{I}_{\{a\le X<b\}}\big)$ and $\mathsf{E}\big((X-a)^2\mathbb{I}_{\{a\le X<b\}}\big)$ can be obtained from Equations (5A.96) and (5A.102), hence,

$$\begin{aligned}
\mathsf{E}(w_l(Y)(X-a)\mathbb{I}_{\{a\le X<b\}}) &= \sigma_x(\beta_0+\beta_1\mu_x)\left\{\phi\left(\frac{a-\mu_x}{\sigma_x}\right) - \phi\left(\frac{b-\mu_x}{\sigma_x}\right)\right\} + \sigma_x\beta_1(a-b)\phi\left(\frac{b-\mu_x}{\sigma_x}\right) \\
&\quad + \big((\mu_x-a)(\beta_0+\beta_1\mu_x) + \sigma_x^2\beta_1\big)\big[\Phi\left(\frac{b-\mu_x}{\sigma_x}\right) - \Phi\left(\frac{b-\mu_x}{\sigma_x}\right)\big], \\
\mathsf{E}(w_l(Y)(b-a)\mathbb{I}_{\{X\ge b\}}) &= (b-a)(\beta_0+\beta_1\mu_x)\big[1-\Phi\left(\frac{b-\mu_x}{\sigma_x}\right)\big] + \beta_1(b-a)\sigma_x\phi\left(\frac{b-\mu_x}{\sigma_x}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathsf{E}(w_l(Y)\mathbf{L}_{[a,b]}) &= \sigma_x(\beta_0+\beta_1\mu_x)\left\{\phi\left(\frac{a-\mu_x}{\sigma_x}\right) - \phi\left(\frac{b-\mu_x}{\sigma_x}\right)\right\} \\
&\quad + \big((\mu_x-a)(\beta_0+\beta_1\mu_x) + \sigma_x^2\beta_1\big)\big[\Phi\left(\frac{b-\mu_x}{\sigma_x}\right) - \Phi\left(\frac{b-\mu_x}{\sigma_x}\right)\big] \\
&\quad + (b-a)(\beta_0+\beta_1\mu_x)\big[1-\Phi\left(\frac{b-\mu_x}{\sigma_x}\right)\big].
\end{aligned} \tag{5A.110}$$

Also note that $\beta_0 + \beta_1\mu_x = \alpha_0 + \boldsymbol{\alpha}^T\boldsymbol{\mu_Y}$, Combining Equations (5A.109) and (5A.110), we can show that

$$\begin{aligned}\Pi^{w_l}(X,Y) &= \sigma_x\left\{\phi\left(\frac{a-\mu_x}{\sigma_x}\right) - \phi\left(\frac{b-\mu_x}{\sigma_x}\right)\right\} + (\mu_x - a)\left\{\Phi\left(\frac{b-\mu_x}{\sigma_x}\right) - \Phi\left(\frac{a-\mu_x}{\sigma_x}\right)\right\} \\ &\quad (b-a)\left\{1 - \Phi\left(\frac{b-\mu_x}{\sigma}\right)\right\} + \frac{\sigma_x^2\beta_1}{\alpha_0 + \boldsymbol{\alpha}^T\boldsymbol{\mu_Y}}\left\{\Phi\left(\frac{b-\mu_x}{\sigma_x}\right) - \Phi\left(\frac{a-\mu_x}{\sigma_x}\right)\right\}, \quad (5A.111)\end{aligned}$$

and hence Equation (5.92) holds.

5A.5 Proof of Proposition 5.7.5

Since $\boldsymbol{\alpha}^T Y \sim N(\boldsymbol{\alpha}^T\boldsymbol{\mu_Y}, \boldsymbol{\alpha}^T\boldsymbol{\Sigma_Y}\boldsymbol{\alpha})$

$$\mathsf{E}\big(w_e(Y)\big) = \mathsf{E}\big(e^{\alpha_0 + \boldsymbol{\alpha}^T Y}\big) = \exp\big(\alpha_0 + \boldsymbol{\alpha}^T\boldsymbol{\mu_Y} + \frac{1}{2}\boldsymbol{\alpha}^T\boldsymbol{\Sigma_Y}\boldsymbol{\alpha}\big). \quad (5A.112)$$

Additionally, from the proof of Proposition 5.7.4, we have

$$\mathsf{E}\big(w_e(Y)|X\big) = \exp\big(\beta_0 + \frac{1}{2}\widetilde{\Sigma} + \beta_1 X\big) = \kappa e^{\beta_1 X},$$

where

$$\begin{aligned}\beta_0 &= \alpha_0 + \boldsymbol{\alpha}^T\boldsymbol{\mu_Y} - \frac{\boldsymbol{\alpha}^T\boldsymbol{\sigma_{YX}}}{\sigma_x^2}\mu_x \\ \beta_1 &= \frac{\boldsymbol{\alpha}^T\boldsymbol{\sigma_{YX}}}{\sigma_x^2} \\ \widetilde{\Sigma} &= \boldsymbol{\alpha}^T\left(\boldsymbol{\Sigma_Y} - \frac{\boldsymbol{\sigma_{YX}}\boldsymbol{\sigma_{YX}}^T}{\sigma_x^2}\right)\boldsymbol{\alpha} \\ \kappa &= \exp\big(\beta_0 + \frac{1}{2}\widetilde{\Sigma}\big).\end{aligned}$$

Also note that $\mathsf{E}\big(w_e(Y)\mathbf{L}_{[a,b]}\big) = \mathsf{E}\big(w_l(y)(X-a)\mathbb{I}_{\{a\le X,b\}}\big) + \mathsf{E}\big(w_l(y)(b-a)\mathbb{I}_{\{X\ge b\}}\big)$, combining Equation (5A.105), we have

$$\begin{aligned}\mathsf{E}\big(w_l(Y)(X-a)\mathbb{I}_{\{a\le X,b\}}\big) &= \kappa e^{\beta_1\mu_x + \frac{1}{2}\beta_1^2\sigma_x^2}\big[\sigma_x\big(\phi(\tilde{a}) - \phi(\tilde{b})\big) + (\mu_x - a + \beta_1\sigma_x^2)\big(\Phi(\tilde{b}) - \Phi(\tilde{a})\big)\big] \\ \mathsf{E}\big(w_l(Y)(b-a)\mathbb{I}_{\{X\ge b\}}\big) &= \kappa(b-a)e^{\beta_1\mu_x + \frac{1}{2}\beta_1^2\sigma_x^2}(1 - \Phi(\tilde{b})),\end{aligned}$$

where $\tilde{a} = \dfrac{a - \mu_x - \beta_1\sigma_x^2}{\sigma_x}, \tilde{b} = \dfrac{b - \mu_x - \beta_1\sigma_x^2}{\sigma_x}$, hence,

$$\begin{aligned}
\mathsf{E}\big(w_e(Y)\mathbf{L}_{[a,b]}\big) &= \kappa e^{\beta_1\mu_x + \frac{1}{2}\beta_1^2\sigma_x^2}\Bigg\{\sigma_x\big[\phi\bigg(\frac{a - \mu_x - \beta_1\sigma_x}{\sigma_x}\bigg) - \phi\bigg(\frac{b - \mu_x - \beta_1\sigma_x}{\sigma_x}\bigg)\big] \\
&\quad + (\mu_x - a + \beta_1\sigma_x^2)\big[\Phi\bigg(\frac{b - \mu_x - \beta_1\sigma_x}{\sigma_x}\bigg) - \Phi\bigg(\frac{a - \mu_x - \beta_1\sigma_x}{\sigma_x}\bigg)\big] \\
&\quad + (b - a)\big[1 - \Phi\bigg(\frac{b - \mu_x - \beta_1\sigma_x}{\sigma_x}\bigg)\big]\Bigg\}. \qquad \text{(5A.113)}
\end{aligned}$$

Also note that $\kappa e^{\beta_1\mu_x + \frac{1}{2}\beta_1^2\sigma_x^2} = \mathsf{E}\big(w_e(Y)\big)$, therefore,

$$\begin{aligned}
\frac{\mathsf{E}\big(w_e(Y)\mathbf{L}_{[a,b]}\big)}{\mathsf{E}\big(w_e(Y)\big)} &= \sigma_x\big[\phi\bigg(\frac{a - \mu_x - \beta_1\sigma_x}{\sigma_x}\bigg) - \phi\bigg(\frac{b - \mu_x - \beta_1\sigma_x}{\sigma_x}\bigg)\big] \\
&\quad + (\mu_x - a + \beta_1\sigma_x^2)\big[\Phi\bigg(\frac{b - \mu_x - \beta_1\sigma_x}{\sigma_x}\bigg) - \Phi\bigg(\frac{a - \mu_x - \beta_1\sigma_x}{\sigma_x}\bigg)\big] \\
&\quad + (b - a)\big[1 - \Phi\bigg(\frac{b - \mu_x - \beta_1\sigma_x}{\sigma_x}\bigg)\big],
\end{aligned}$$

and Equation (5.93) holds.

Chapter 6

Conclusions and Future Work Directions

6.1 Review and Conclusions

The main objective of this thesis is to propose actuarially sound ratemaking frameworks for pricing agricultural insurance. This is one of the key elements in ensuring the long-term sustainability of the agricultural insurance and reinsurance programs. In general, this thesis develops and evaluate three high dimensional approaches for agricultural insurance pricing and risk management, including two credibility approaches, a LSHAC approach, and a multivariate weighted distribution approach.

In order to enhance the crop (re)insurance pricing framework, we provide systematic discussion on the issues associated with small sample on yield distribution modeling and overcome this obstacle by extending the classic Bühlmann-Straub credibility model and proposing a credibility-based Erlang mixture model to improve the goodness-of-fit and reinsurance pricing. Following this work, we also propose a new credibility estimator, by expanding the traditional regression credibility model, to improve crop yield forecasting and crop reinsurance pricing. It is shown theoretically that this new credibility estimator has some appealing statistical properties including unbiasedness and smaller mean quadratic loss. Within this framework, high dimensional spatially correlated weather variables are integrated into the pricing system by developing comprehensive model selection algorithms that combine Cross-validation (CV) and principle component analysis (PCA) so that the in-sample and out-of-sample forecasting abilities are improved substantially.

Copula models are commonly used in high-dimensional modelling and have been studied extensively empirically in agricultural insurance modelling. However, most research has typically focused on linear correlation dependence involving the Gaussian copulas. In this thesis, we improve the dependence modeling of temperature processes in Canada with Lévy subordinated Archimedean copula (LSHAC) models. We develop a three-stage estimating procedure for LSHAC with special attention on the estimation of the hierarchical structure of the data. The empirical analysis shows that compared to traditional Gaussian copulas, LSHAC has better fitting ability and more flexibility in modeling the tail dependence. To capture the heavy tail property of the regional temperature data, non-Gaussian distributions (such as Variance Gamma (VG), Normal-Inverse Gaussian (NIG), etc.) are employed. This is the first time that LSHACs are used to model weather risk. Several hedging strategies are developed and compared. Some important findings of this research include: (1) optimal hedging strategies can be achieved by choosing an appropriate geographical level of aggregation; (2) using a basket of derivatives from diverse locations could lead to more efficient hedging strategies.

Previous research indicates that it is necessary to integrate auxiliary variables in to the ratemaking process to refine the pricing framework. This thesis proposes a new premium principle based on multivariate weighted distribution to provide a methodology of weighting other variables in the historical loss experience. We derived some desirable properties of the premium based on multivariate weighted distributions. In addition, the economic premium principle discussed in Bühlmann (1980), and the *Esscher principle* are special cases of the framework proposed in this thesis. An application to the reinsurance experience in Manitoba from 2001 to 2011 reveals that including liability and economic variables into the pricing framework will redistribute premium rates, assigning higher loadings to more risky layers of contracts and hence achieve more sustainable profits in the long term.

6.2 Areas of Future Work

6.2.1 Factor Models for Crop Yields Forecasting

A good crop yield forecasting model is essential for achieving an accurate loss predicting, hence an actuarially fair premium, therefore, it is of critical importance to construct a sustainable agricultural insurance and reinsurance programs. One important future work direction is to investigate efficient and accurate crop yield forecasting model using factor models.

High dimensional static and dynamic factor models have been widely utilized in macroeconomics, especially in the data rich environment where the data are over short period of time but have a large cross section. Therefore, factor models are promising for crop yield prediction since the agricultural data have many variables to be considered including crop types, soil, temperature, rainfall, etc (Borman et al., 2013). Yet there are very limited literature investigating these models in crop yield forecasting. Therefore, in order to improve the forecasting abilities, rich volume of weather data will be integrated into the model with high-dimensional static and dynamic factor models for crop yields modeling. In particular, the identification theory for high-dimensional static and dynamic factor models through linear and non-linear restrictions by Bai and Wang (2014) will be used to facilitate structural analysis of factor models with a large number of parameters.

6.2.2 Utility-based Credibility Pricing Model

The utility premium measures risks by imposing a corporate utility function in risk premium to specify the risk preference and rate levels. Let us assume that $U(w)$ is the utility function of an insurer with current wealth level w, and let X be the random variable representing the loss, then the utility premium π can be calculated according to the principle

$$U(w) \quad = \quad E[U(w + \pi - X)]. \tag{6.1}$$

Obviously, if we define a new wealth level, $\omega = w + \pi$, then equation (6.1) can be rewritten as

$$U(\omega - \pi) \quad = \quad E[U(\omega - X)]. \tag{6.2}$$

Given the risk aversion assumption which implies the monotonicity and convexity of the utility function, we have:

$$U(\omega - \pi) = E(U(\omega - X)) \leq U(E(\omega - X)) = U(\omega - E(X)), \tag{6.3}$$

which leads to

$$\pi \geq E(X). \tag{6.4}$$

Therefore, premiums constructed from the utility theory satisfy the risk loading property.

From the definition in Equation (6.2), we are interested in finding an estimator $\hat{\pi}$ such that $U(\omega - \hat{\pi})$ is as close to $U(\omega - \pi)$ as possible. Arising from this objective, we calculate the premium by estimating a vector of parameters $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_n)'$ from the following optimization problem

$$\underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \, E[U(w - \hat{\pi}) - U(w - \pi)]^2 \tag{6.5}$$

where

$$\hat{\pi} = \alpha_0 + \sum_{i=1}^{n} \alpha_i X_i.$$

We can see in this set-up, the traditional credibility theory is combined with the utility theory. We call the resulting premium *Utility-Credibility Premium*. There are several advantages in doing so. First, the risk preference of the insurer is considered in the ratemaking procedure. Since different insurers have diverse preference of misspecification of the insurance premiums (both over pricing or under pricing), there exist systemic pricing risk without taking this into consideration. Second, binding the utility function into rating system potentially providing more economic interpretation of the inherent loss random variables. For example, assume the individual utility function has an exponential form, i.e., $U(W) = e^{-\rho W}$, where ρ is defined as the risk aversion of the individual, then we can see that higher moments of the loss distribution are considered since basically we are considering the MGF of the loss distribution.

6.2.3 Basis Risk Decomposition for Index-based Insurance (IBI)

A lot of recent research has focused on weather index-based insurance (IBI) as an alternative to traditional indemnity-based insurance in order to avoid the moral hazard and adverse selection. However, a major difficulty for IBI is the basis risk, which is referred to the imperfect correlation between the risk exposures and instruments used to hedge the corresponding risks. Basis risk is cited as a primary concern in agricultural risk management (Brockett et al., 2005; Turvey et al., 2006). There are three key resources of basis risk in agricultural IBI.

- **Variable Basis Risk:** When weather variables used for hedging and the loss exposures are from the same geographic region, the basis risk exists because of the imperfect correlation between the hedging instrument and liability being hedged. Therefore, it is critical to come up with a good model that describe the correlation between crop yields and weather variables.

- **Contract Basis Risk:** In theory, the insurer needs to design an optimal index insurance product (we refer to this contract as "local contract") for every farmer/producer in the business, in order that every contract sold protects the actual loss of the client. However, in practice, it is unrealistic and costly, hence, insurers usually provide a uniform contract to all farmers within a geographical region (we refer to this contract as "global contract"). The difference contract designs between the local contract and global contract will result in contract basis risk.
- **Spatial Basis Risk:** Spatial basis risk arises when the underwriting risk exposures locate differently from where the weather indices are tabulated (Brockett et al., 2005). For example, the standardized weather indices listed at CME include 18 US cities and 6 Canadian cities, meaning that participants of these derivatives face basis risk when their exposures locate in other areas.

Using a comprehensive and detailed loss experience data set in Manitoba, Canada, it is possible to study the basis risk from a unique perspective by providing the basis risk decomposition. Assume that the insurer has a business that contains d farms. Let Y_i be crop yield of farm i and L_i be the loss of certain indemnity contract, where $i = 1, 2, \ldots, d$. An example of the indemnity contract can be Multiple Peril Crop Insurance (MPCI), a broad-based crop insurance program regulated by the U.S. Department of Agriculture (USDA) and subsidized by the Federal Crop Insurance Corporation (FCIC). Since the loss of an indemnity contract is a function of crop yield, we denote it as $L_i(Y_i)$. We also assume that the insurance company provides all the farmers the same IBI contract, which is claimed based on the records from the reference weather stations. To be more specific, the insurance company provides farm i a global contract $I(W^r; \theta^g)$, where W^r is the weather record from the reference weather station, and θ^g is the parameters of the global contract design. Therefore, the basis risk of the insurer can be expressed as

$$BR = \sum_{i=1}^{d} ||L_i(Y_i) - I(W^r; \theta^g||, \tag{6.6}$$

where BR is the total basis risk of the insurance company and $||\cdot||$ is the norm used to quantify the basis risk.

Suppose now that the insurance company has an optimal contract design for each farmer, $I_i(W_i^l; \theta_i^l)$, then the variable basis risk of the insurer can be expressed as

$$BR_{\text{Variable}} = \sum_{i=1}^{d} ||L_i(Y_i) - I_i(W_i^l; \theta_i^l||. \tag{6.7}$$

In addition, the contract basis risk is

$$BR_{\text{Contract}} = \sum_{i=1}^{d} ||I_i(W_i^l; \theta_i^l - I(W_i^l; \theta^g)||, \tag{6.8}$$

and the spatial basis risk is

$$BR_{\text{Spatial}} = \sum_{i=1}^{d} ||I(W_i^l; \theta^g - W^r; \theta^g||. \tag{6.9}$$

Therefore, we have the following basis risk decomposition:

$$\begin{aligned} BR &= BR_{\text{Variable}} + BR_{\text{Contract}} + BR_{\text{Spatial}} + \varepsilon, & (6.10)\\ &= \sum_{i=1}^{d} ||L_i(Y_i) - I_i(W_i^l; \theta_i^l|| + \sum_{i=1}^{d} ||I_i(W_i^l; \theta_i^l - I(W_i^l; \theta^g)|| \\ &\quad + \sum_{i=1}^{d} ||I(W_i^l; \theta^g - W^r; \theta^g|| + \varepsilon, & (6.11) \end{aligned}$$

where the error term ε represents other resources of basis risk not contained into this decomposition.

My Ph.D. research in actuarial ratemaking in agricultural insurance stimulates many other interesting future work directions. For example, it is appealing to investigate more efficient estimation procedure for the LHSAC model and apply it in the crop insurance pricing and risk management. In addition, based on the weather system and model selection algorithms in Chapter 3, it is promising to develop novel index-based insurance (IBI) products, which are able to reduce basis risk and moral hazards at the same time. It is also interesting to develop the optimal reinsurance allocation strategy and an effective public-private risk sharing partnership, in order to provide a scientific decision making baseline for both insurers and governments.

References

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198.

Acerbi, C. and Tasche, D. (2002). Expected shortfall: a natural coherent alternative to value at risk. *Economic Notes*, 31(2):379–388.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transaction on Automatic Control*, 19:716–723.

Alaton, P., Djehince, B., and Stillberg, D. (2002). On modelling and pricing weather derivatives. *Applied Mathematical Finance*, 9(1):20.

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

Anderson, T. W. (1952). Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. *The Annals of Mathematical Statistics*, 23(2):193–212.

Anderson, T. W. (1962). On the distribution of the two-sample cramér-von mises criterion. *The Annals of Mathematical Statistics*, 33(3):1148–1159.

Andrew Levin, Chien-Fu Lin, C.-S. J. C. (2002). Unit root tests in panel data: asymptotic and finite-sample properties. *Journal of Econometrics*, 108(1):1–24.

Bai, J. and Wang, P. (2014). Identification theory for high dimensional static and dynamic factor models. *Journal of Econometrics*, 178(2):794–804.

Barndorff-Nielsen, O. E. (1997). Normal inverse gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics*, 24(1):1–13.

Bellini, F. (2005). *The Weather Derivatives Market: Modelling and Pricing Temperature*. PhD thesis, University of Lugano, Lugano.

Benth, F. E. and Benth, J. S. (2005). Stochastic modelling of temperature variations with a view towards weather derivatives. *Appl Math Finance*, 12(1):53–85.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Borman, J. I., Goodwin, B. K., Coble, K. H., Knight, T. O., and Rejesus, R. (2013). Accounting for short samples and heterogeneous experience in rating crop insurance. *Agriculture Finance Review*, 73(1):88–101.

Brockett, P. L., Wang, M., and Yang, C. (2005). Weather derivatives and weather risk management. *Risk Management and Insurance Review*, 8(1):127–140.

Brown, D. W. (1969). Heat units for corn in southern ontario. *Ontario Department of Agriculture and Food*, pages 111–131.

Bühlmann, H. (1967). Experience rating and credibility. *ASTIN Bulletin*, 4(3):199–207.

Bühlmann, H. (1980). An economic premium principle. *ASTIN Bulletin*, 11:52–60.

Bühlmann, H. (1997). Credibility in the regression case revisited. *ASTIN Bulletin*, 27:83–98.

Bühlmann, H., Delbaen, F., Embrechts, P., and Shiryaev, A. N. (1996). No-arbitrage, change of measure and conditional esscher transforms. *CWI Quarterly*, 9(4):291–317.

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

Bühlmann, H., Shevchenko, P. V., and Wuthrich, M. V. (2007). A 'toy' model for operational risk quantification using credibility theory. *The Journal of Operational Risk*, 2(1):3–19.

Bühlmann, H. and Straub, E. (1970). Glaubwürdigkeit für Schadensätze. *Bulletin of Swiss Ass. of Act.*, 70:111–133.

Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference*. Springer, second edition edition.

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*, 45(4):719–737.

Campbell, S. D. and Diebold, F. X. (2005). Weather forecasting for weather derivatives. *Journal of the American Statistical Association*, 100(469):6–16.

Cassman, K. G. (1999). Ecological intensification of cereal production systems: yield potential, soil quality, and precision agriculture. *PNAS: Proceedings of the National Academy of Sciences of the United States of America*, 96(11):5952–5959.

Chambers, R. G. (1989). Insurability and moral hazard in agricultural insurance markets. *American Journal of Agriculture Economics*, 71:604–616.

Choroś-Tomczyk, B., Härdle, W. K., and Okhrin, O. (2013). Valuation of collateralized debt obligations with hierarchical archimedean copulae. *Journal of Empirical Finance*, 24:42–62.

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.

Conway, G. and Toenniessen, G. (1999). Feeding the world in the twenty-first century. *Nature*, 402(6761):55.

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

Daubechies, I. (1990). The wavelet transform time-frequency localization and signal anaysis. *The IEEE Transactions on Information Theory*, 36(5):961–1005.

Daubechies, I. (1992). *Ten lectures on wavelets lectures on wavelets*, volume 61. Society for Industrial and Applied Mathematics.

DeJong, D., Nankervis, J., Savin, N., and Whiteman, C. (1991). The power problems of unit root tests in time series with autoregressive errors. *Journal of Econometrics*, 53:323–343.

Denuit, M., Dhaene, J., Goovaerts, M., and Kaas, R. (2005). *Actuarial Theory for Dependent Risks: Measures, Orders and Models*. John Wiley & Sons, Ltd.

Dickey, D. A. (1976). *Estimation and Hypothesis Testing in Nonstationary Time Series*. PhD thesis, Iowa State University.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimation for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74:427–431.

Dischel, R. S. and Barrieu, P. (2002). *Financial weather contracts and their application in risk management. In: Climate Risk and the Weather Market: Financial Risk Management With Weather Hedges.* Risk Books, London, UK.

Doherty, N. A. and Dionne, G. (1993). Insurance with undiversifiable risk: contract structure and organization form of insurance firms. *Journal of Risk and Uncertainty*, 6(2):187–203.

Dong, W., Shah, H., and Wong, F. (1996). A rational approach to pricing of catastrophe. *Journal of Risk and Uncertainty*, 12(2-3):201–218.

Elliott, G., Rothenberg, T. J., and Stock, J. H. (1996). Efficient tests for and autoregressive unit root. *Econometrica*, 64(4):813–836.

Embrechts, P., Lindskog, F., and McNeil (2003). Modelling dependence with copulas and applications to risk management. *Handbook of heavy tailed distributions in finance*, 8(1):329–384.

Embrechts, P., McNeil, A., and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. In Dempster, M., editor, *Risk Management: Value at Risk and Beyond:*, pages 176–223. Cambridge University Press.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007.

Feller, W. (2008). *An Introduction to Probability Theory and Its Applications.* John Wiley & Sons, Inc.

Frittelli, M. (2000). The minimum entropy martingale measure and the valuation problem in incomplete markets. *Mathematical Finance*, 10(1):39–52.

Furman, E. and Zitikis, R. (2008a). Weighted premium calculation principles. *Insurance: Mathematics and Economics*, 42:459–465.

Furman, E. and Zitikis, R. (2008b). Weighted risk capital allocations. *Insurance: Mathematics and Economics*, 43(1):263–269.

Gallagher, P. (1986). U. S. corn yield capacity and probability: Estimation and forecasting with none-symmetric distributions. *North Central Journal of Agriculure Economics*, 8:109–122.

Gerber, H. and Shiu, S. (1994). Option pricing by Esscher transforms. *Transaction of Society of Actuaries*, 46:99–140.

Goodwin, B. K. (2001). Problems with market insurance in agriculture. *American Journal of Agriculture Economics*, 83(3):643–649.

Goodwin, B. K. and Hungerford, A. (2014). Copula-based models of systemic risk in U.S. agriculture: inplications for crop insurance and reinsurance contracts. *American Journal of Agricultural Economics*, pages 1–18.

Goodwin, B. K. and Ker, A. P. (1998). Nonparametric estimation of crop yield distributions: Implications for rating group-risk (GRP) crop insurance contracts. *American Journal of Agricultural Economics*, 80:139–153.

Hachemeister, C. A. and Kahn, P. M. (1975). *Credibility for regression models with application to trend*. Academic Press. New York.

Hadar, J. and Russell, W. (1969). Rules for ordering uncertain prospects. *American Economic Review*, 59(1):25–34.

Hamilton, J. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Hardy, M. R. and Panjer, H. H. (1998). A credibility approach to mortality risk. *ASTIN Bulletin*, 28(2):269–283.

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*, 31(1):163–182.

Heilmann, W. R. (1989). Decision theoretic foundations of credibility theory. *Insurance: Mathematics and Economics*, 8(1):77–95.

Hellmuth, M. E., Osgood, D. E., Hess, U., Moorhead, A., and Bhojwani, H. (2009). Index insurance and climate risk: Prospects for development and disaster management. Technical Report No. 2, Climate and Society, International Research Institute for Climate and Society, New York, USA.

Henzea, N. and Zirkler, B. (1990). A class of invariant consistent tests for multivariate normality. *Communications in Statistics - Theory and Methods*, 19(10):3595–3617.

Hering, C., Hofert, M., Mai, J.-F., and Scherer, M. (2010). Constructing hierarchical Archimedean copulas with Lévy subordinators. *Journal of Multivariate Analysis* , 101(6):1428–1433.

Hofert, M. (2008). Sampling Archimedean copulas. *Computational Statistics and Data Analysis*, 52(12):5163–5174.

Hofert, M. (2012). A stochastic representation and sampling algorithm for nested Archimedean copulas. *Journal of Statistical Computation and Simulation*, 82(9):1239–1255.

Hogg, R. and Klugman, S. (1984). *Loss Distributions*. Wiley, New York.

Hubalek, J. and Neilsen, B. (2006). On the Esscher transform and entropy for exponential lévy models. *Quantitative Finance*, 6(2):125–145.

Ingram, D. (2009). ERM and actuaries. *Casualty Actuarial Society E-Forum*.

Insurance Bureau of Canada (IBC) (2014). 2014 facts of the Property & Casualty insurance industry.

IPCC (2007). Fourth assessment report. Technical report, Intergovernmental Panel on Climate Change.

IPCC (2013). Climate change 2013: The physical science basis. Technical report, Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, UK: Cambridge University Press.

Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*. Taylor & Francis.

Jolliffe, I. T. (2002). *Principle Component Analysis*. Spinger series in statistics, 2nd edition edition.

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

Jr., J. H. W. (1963). Hierarchical grouping to optimize an objective function. *Journal of American Statistical Association*, 58(301):236–244.

Jung, A. R. and Ramezani, C. A. (1999). Valuing risk management tools as complex derivatives: An application to revenue insurance. *Journal of Financial Engineering*, 8:99–120.

Just, R. and Weninger, Q. (1999). Are crop yields normally distributed? *American Journal of Agricultural Economics*, 81:287–304.

Kamps, U. (1998). On a class of premium principles including the Esscher principle. *Scandinavian Actuarial Journal*, 1998(1):75–80.

Ker, A. and Goodwin, B. K. (2000). Nonparametric estimation of crop insurance rates revisited. *American Journal of Agricultural Economics*, 83(2):463–478.

Kimberling, C. H. (1974). A probabilistic interpretation of complete monotonicity. *Aequationes Mathematicae*, 10(2-3):152–164.

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*, 106(496):1291–1303.

Kleindorfer, P. R., Kunreuther, H., and Ou-Yang, C. (2012). Single-year and multi-year insurance policies in a competitive market. *Journal of Risk and Uncertainty*, 45(1):51–78.

Kocherlakota, S. (1995). Discrete bivariate weighted distributions under multiplicative weighted function. *Communications in Statistics - Theory and Methods*, 24:533–551.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Appears in the International Joint Conference on ArtifIcial Intelligence (IJCAI)*.

Kole, E., Koedijk, K., and Verbeek, M. (2007). Selecting copulas for risk management. *Journal of Banking & Finance*, 31(8):2405–2423.

Kolmogorov, A. N. (1933). Sulla determinazione empirica di una legge di distribuzione. *G. Ist. Ital. Attuari*, 4:83–91.

Kunreuther, H., Hogarth, R., and Meszaros, J. (1993). Insurer ambiguity and market failure. *Journal of Risk and Uncertainty*, 7(1):71–87.

Kurowicka, D. and Cooke, R. M. (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling*. John Wiley & Sons.

Kwiatkowski, D., Phillips, P., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54:159–178.

Lanoue, C., Sherrick, B. J., Woodard, J. D., and Paulson, N. D. (2010). Evaluating yield models for crop insurance rating. In *Selected Paper prepared for presentation at the Agricultural & Applied Economics Associations 2010 AAEA, CAES & WAEA Joint Annual Meeting*.

Lau, K. M. and Weng, H. Y. (1995). Climate signal detection using wavelet transform: How to make a time series sing. *Bulletin of the American Meteorological Society*, 76(12):2391–2402.

Lee, S. and Lin, S. (2010). Modeling and evaluating insurance losses via mixtres of Erlang distributions. *North American Actuarial Journal*, 14(1):107–130.

Levy, H. (1992). Stochastic dominance and expected utility: survey and analysis. *Management Sciences*, 38:555–593.

Li, J. S.-H. and Hardy, M. R. (2011). Measuring basis risk in longevity hedges. *North American Actuarial Journal*, 15(2):177–200.

Li, J. S.-H., Hardy, M. R., and Tan, K. S. (2010). On pricing and hedging the no-negative-equity guarantee in equity release mechanisms. *Journal of Risk and Insurance*, 77(2):499–522.

Lobell, D. B. and Asner, G. P. (2003). Climate and management contributions to recent trends in u.s. agricultural yields. *Science*, 299(5609):1032.

Lobell, D. B. and Burke, M. B. (2008). Why are agricultural impacts of climate change so uncertain? the imortance of temperature relative to precipitation. *Environmental Research Letters*, 3(1):1–8.

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*, 150(11):1443–1452.

Luttrell, C. and Gilbert, R. (1976). Crop yields: Random, cyclical, or bunchy? *American Journal of Agricultural Economics*, 58:521–31.

Maddala, G. S. and Wu, S. (1999). A comparative study of unit root tests with panel data and a new simple test. *Oxford Bulletin of Economics and Statistics*, 61(S1):631–652.

Mahfoud, M. and Patil, G. P. (1982). On weighted distribution. In *Statistics and Probability: Essays in Hornor of C. R. Rao*, pages 479–492. North Holland, Amsterdam.

Mahul, O. and Stutley, C. J. (2010). Government support to agricultural insurance: challenges and options for developing countries. *World Bank Publications*.

Mai, J.-F. and Scherer, M. (2012). H-extendible copulas. *Journal of Multivariate Analysis* , 110(0):151–160. Special Issue on Copula Modeling and Dependence.

Mallows, C. L. (1973). Some comments on c_p. *Technometrics*, 15:661–676.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3):519–530.

Marshall, A. W. and Olkin, I. (1988). Families of multivariate distributions. *Journal of the American Statistical Association*, 83(403):834–841.

Matheron, G. (1963). Principles of geostatistics. *Economic geology*, 58(8):1246–1266.

McCarl, B. A., Villavicencio, X., and Wu, X. (2008). Climate change and future analysis: is stationary dying? *American Journal of Agriculture Economics*, 90:1241–1247.

McNeil, A. J. (2008). Sampling nested Archimedean copulas. *Journal of Statistical Computation and Simulation*, 78(6):567–581.

Miranda, M. and Glauber, J. (1997). Systematic risk, reinsurance, and the failure of crop insurance markets. *American Journal of Agriculture Economics*, 79(1):206–215.

Moss, C. B. and Shonkwiler, J. S. (1993). Estimating yield distribution with a stochastic trend and nonnormal errors. *American Journal of Agriculture Economics*, 175(4):1056–1062.

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*, 70(1):137–164.

Navarro, J., Ruiz, J. M., and Aguila, Y. D. (2006). Multivariate weighted distributions: A review and some extensions. *Statistics: A Journal of Theoretical and Applied Statistics*, 40(1):51–64.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer, second edition edition.

Nelson, C. H. and Loehman, E. T. (1987). Further toward a theory of agricultural insurance. *AJAE*, 69(523-531).

Nelson, C. H. and Preckel, P. V. (1989). The conditional distribution as a stochastic production function. *American Journal of Agricultural Economics*, 71:370–378.

Okhrin, O., Odening, M., and Xu, W. (2013a). Systemic weather risk and crop insurance: The case of China. *Journal of Risk and Insurance*, 80(2):351–372.

Okhrin, O., Okhrin, Y., and Schmid, W. (2013b). On the structure and estimation of hierarchical Archimedean copulas. *Journal of Econometrics*, 173(2):189–204.

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*, 40(9):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*, 9999(999):1–17.

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*, 470:382–385.

Parodi, P. (2104). *Pricing in General Insurance*. CRC Press.

Patil, G. P., Rao, C. R., and Ratnaparkhi, M. V. (1986). On discrete weighted distributions and their use in model choice for observed data. *Communications in Statistics - Theory and Methods*, 15(3):907?918.

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, 47(2):527–556.

Patton, A. J. (2009). Copula-based models for financial time series. In Mikosch, T., Kreiß, J.-P., Davis, R. A., and Andersen, T. G., editors, *Handbook of Financial Time Series*, pages 767–785. Springer Berlin Heidelberg.

Perron, P. (1986). *Hypothesis Testing in Time Series Regression With a Unit Root*. PhD thesis, Yale University, Department of Economics. (unpublished).

Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics*, 22(2):265–312.

Phillips, P. C. B. and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75:335–346.

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

Porth, L., Pai, J., and Boyd, M. (2014a). A portfolio optimization approach using combinatorics with a genetic algorithm for developing a reinsurance model. *Journal of Risk and Insurance*.

Porth, L., Tan, K. S., and Weng, C. (2013). Optimal reinsurance analysis from a crop insurer's perspective. *Agriculture Finance Review*, 73(2):310 – 328.

Porth, L., Zhu, W., and Tan, K. S. (2014b). A credibility-based Erlang mixture model for pricing crop reinsurance. *Agricultural Finance Review*, 74(2):162–187.

Power, G. J. and Turvey, C. G. (2010). Long-range dependence in the volatility of commodity futures prices: wavelet-based evidence. *Physica A*, 389(1):79–90.

Priest, G. L. (1996). The government, the market, and the problem of catastrophic loss. *Journal of Risk and Uncertainty*, 12(2-3):219–237.

Quiggen, J., Karagiannis, G., and Stanton, J. (1994). Crop insurance and crop production: An empirical study of moral hazard and adverse selection. *Economics of Agricultural Crop Insurance*, pages 253–272.

Ramsey, F. L. and Schafer, D. W. (2013). *The Statistical Sleuth*. Brook/Cole Cengage Learning, 3rd edition edition.

Rao, C. R. (1965). On discrete distributions arising out of methods of ascertainment. *The Indian Journal of Statistics, Series A*, 27(2/4):311–324.

Royston, P. (1991). Estimating departure from normality. *Statistics in Medicine*, 10(8):1283–1293.

Salinger, M. (2005). Climate variability and change: past, present and future - an overview. *Climate Change*, 101(27):9971–9975.

Savu, C. and Trede, M. (2010). Hierarchies of Archimedean copulas. *Quantitative Finance*, 10(3):295–304.

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

Schneider, K. and Roth, M. (2013). Growing premium. *Insider Quarterly*.

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annual of Statistics*, 6.

Schwert, G. W. (1989). Tests for unit roots: a monte carlo investigation. *Journal of Business & Economic Statistics*, 7(2):147–59.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4):591–611.

Sherrick, B., Zanini, F., Schnitkey, G., and Irwin, S. (2004). Crop insurance valuation under alternative yield distributions. *American Journal of Agricultural Economics*, 86:406–419.

Shi, W., Tao, F., and Zhang, Z. (2013). A review on statistical models for identifying climate contributions to crop yields. *J. Geogr. Sci.*, 23(3):567–576.

Siu, T., Tong, H., and Yang, H. (2004). On pricing derivatives under GARCH models: a dynamic Gerber-Shiu approach. *North American Actuarial Journal*, 8:17–31.

Sklar, A. (1959). Fonctions dé repartition àn dimension et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231.

Smirnov, N. (1948). Table for estimating the goodness of fit of empirical distributions. *The Annals of Mathematical Statistics*, 19(2):279–281.

Stocks, J. R. (2000). A derivative security approach to setting crop revenue coverage insurance premiums. *Journal of Agricultural and Rsource Economics*, 25:159–176.

Székely, G. J. and Rizzo, M. L. (2005). Hierarchical clustering via joint between-within distance: extending Ward's minimum variance method. *Journal of Classification*, 22(2):151–183.

Tan, K. S., Weng, C., and Zhang, Y. (2011). Optimality of general reinsurance contracts under CTE risk measure. *IME*, 49:175–187.

Tankov, P. (2004). *Financial Modelling with Jump Processes*. CRC Press.

Tijms, H. C. (1994). *Stochastic Models: an Algorithmic Approach*. Wiley.

Tirupattur, V., Hauser, R. J., and Chaherli, N. M. (1996). Crop yield and price distributional effects on revenue hedging. In *OFOR Paper Number 96-05*.

Turvey, C. and Zhao, J. (1999). Parametric and nonparametric crop yield distributions and their effects on all-risk crop insurance premiums. In *Working Paper WP99/05*. Department of Agriculture Economics and Business, University of Guelph.

Turvey, C. G., Driver, H. C., and Baker, T. G. (1988). Systematic and nonsystematic risk in farm portfolio selection. *American Journal of Agricultural Economics*, 70(4):831–836.

Turvey, C. G., Weersink, A., and Chiang, S. H. C. (2006). Pricing weather insurance with random strike price: The Ontario ice-wine harvest. *American Journal of Agriculture Economics*, 88(1):696–709.

USDA (2014). World agricultural supply and demand estimates report (WASDE). Technical report, Office of the chief economist (OCE), United States Department of Agriculture.

Vylder, F. D. (1976a). Geometrical credibility. *Scand. Act. J.*, pages 121–149.

Vylder, F. D. (1976b). Optimal semilinear credibility. *Bulletin of Swiss Ass. of Act.*, pages 27–40.

Wang, S. (1995). Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance: Mathematics and Economics*, 17:43–54.

Wang, S. (1996). Premium calculation by transforming the layer premium density. *ASTIN Bulletin*, 26(1):71–92.

Wang, S., Young, V. R., and Panjer, H. H. (1997). Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics*, 21(2):173–183.

Weiß, G. N. F. and Scheffer, M. (2015). Mixture pair-copula-constructions. *Journal of Banking & Finance*, 54(0):175–191.

Wen, L. and Wu, X. (2011). The credibility estimator with general dependence structure over risks. *Communications in Statistic-Theory and Methods*, 40:1893–1910.

Whelan, N. (2004). Sampling from Archimedean copulas. *Quantitative Finance*, 4(3):339–352.

Wirch, J. L. and Hardy, M. R. (1999). A synthesis of risk measures for capital adequacy. *Insurance: Mathematics and Economics*, 25(3):337–347.

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*, 18(2):279–293.

Woodard, J. D. and Garcia, P. (2008a). Basis risk and weather hedging effectiveness. *Agricultural Finance Review*, 68(1):99–117.

Woodard, J. D. and Garcia, P. (2008b). Weather derivatives, spatial aggregation, and systemic risk: Implications for reinsurance hedging. *Journal of Agricultural and Rsource Economics*, 33(1):34–51.

Woodard, J. D., Schnitkey, G. D., Sherrick, B. J., Lozano-Gracia, N., and Anselin, L. (2012). A spatial econometric analysis of loss experience in the U.S. crop insurance program. *Journal of Risk and Insurance*, 79(1):261–285.

Woodard, J. D. and Sherrick, B. J. (2011a). Actuarial Impacts of Loss Cost Ratio Ratemaking in U.S. Crop Insurance Programs. *Journal of Agriculture and Resource Economics*, 36(1):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*, 93(4):968–982.

Yang, T. (2011). Measurement of yield distributions: time-varying mixture distribution models. In *Prepared for Agricultural & Applied Economics Association's 2011 Annual Meeting, Pittsburgh, Pennsylvania*.

Young, V. R. (2004). Premium principles. *Encyclopedia of Actuarial Science*, 3.

Zapata, H. O. and Rambaldi, A. H. (1989). Effects of data transformation on stochastic properties of economic data. In *Paper presented at the annual meetings of the American Agricultural Economics Association, Baton, Rouge, Louisiana*.

Zhang, W., Zhao, D., and Wang, X. (2013). Agglomerative clustering via maximum incremental path integral. *Pattern Recognition*, 46(11):3056–3065.