SPATIAL CROSS-SECTIONAL CREDIBILITY MODELS WITH GENERAL DEPENDENCE STRUCTURE AMONG RISKS

by

Jimmy Poon

B.Sc., Simon Fraser University, 2011

a Project submitted in partial fulfillment
of the requirements for the degree of

Master of Science

in the
Department of Statistics and Actuarial Science
Faculty of Science

APPROVAL

Name: Jimmy Poon

Degree: Master of Science

Title of Project: Spatial Cross-Sectional Credibility Models with General Dependence Structure Among Risks

Examining Committee:

Dr. Cary Tsai
Associate Professor
Chair

Dr. Yi Lu
Associate Professor
Senior Supervisor
Simon Fraser University

Dr. Gary Parker
Associate Professor
Co-Supervisor
Simon Fraser University

Dr. Tim Swartz
Professor
External Examiner
Simon Fraser University

Date Approved: ______________________

Abstract

Credibility models with general dependence structure among risks and conditional spatial cross-sectional dependence are studied in this project. Predictors of future losses for a Bühlmann-type credibility model under both types of dependence are derived by minimizing the quadratic loss function, and this is further extended to Bühlmann-Straub and regression credibility model formulations. Non-parametric estimators of structural parameters of various models under a spatial statistics context are also considered especially for the case of equal unconditional means. An example with crop insurance losses is studied to illustrate the use of predictors and estimators proposed in this project. Finally, the performance of the predictors and estimators are evaluated in a simulation study.

Keywords: Credibility premium; Spatial statistics; Dependence; Regression credibility model; Structural parameter estimation

To my family.

Acknowledgments

I would like to express my sincere appreciation to everyone who was involved in this project. The most important of all is my senior supervisor, Dr. Yi Lu. From my first research term with her as a research assistant to entering the Master's program at Simon Fraser University and finally the completion of this project, she has offered me constant encouragement and guidance throughout.

Next, I would like to offer my special thanks to my co-supervisor, Dr. Gary Parker. Not only has he given many precious inputs to this project, he has also provided substantial support to my many endeavours since my very first exposure to research in Actuarial Science.

Two other professors I would like to thank are Dr. Cary Tsai and Dr. Tim Swartz. They have both contributed in different ways from establishing my foundation in credibility theory to offering invaluable suggestions to this project for improvement.

Finally, I would like to acknowledge Natural Sciences and Engineering Research Council of Canada and the Simon Fraser University Senate Graduate Awards Adjudication Committee to have granted me a Canada Graduate Scholarship and a Graduate Fellowship which have been of great importance during my Master's studies.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Credibility theory is a widely used ratemaking technique that, in essence, predicts future values or estimates the true risk of entities given a history of their past claims and the risk classes to which they belong. Often, although risk classes are made as homogeneous as possible using a priori knowledge, entities in the risk class may still exhibit some inhomogeneity. In the most common form of credibility models, the minimization of the expected squared difference between the predicted values and the claim history leads to a prediction formula that interpolates between the experience of one entity and the common characteristics of all entities in the same risk class. This allows a very intuitive interpretation of the result of a credibility prediction formula: the experience of an entity is assigned a weight that represents how credible the experience of the specific entity is in comparison to the common characteristics of its risk class. If the experience of an entity is fully credible, then the experience of the entity can be used to predict its future loss. If the experience of an entity is not fully credible, then a mix of the experience of the entity and the common characteristics of its risk class should be used.

The desirable results and properties of credibility theory have led to its wide use in fields such as automobile insurance and health insurance. The literature on credibility theory is extensive. There are credibility models that have various forms of dependence through time and among entities. In this project, a credibility model that allows for dependence through risk parameters and conditional dependence in a distribution-free framework is considered, with a focus on estimation methods for applications with spatial dependence.

1.1 Background and Motivation

The beginning of modern credibility theory is attributed to Bühlmann (1967). The Bühlmann credibility model uses a Bayesian framework and assumes the risk parameters of each entity to be independent and to follow a common distribution. In addition, conditional on the risk parameter of an entity, the losses are assumed to be independent and identically distributed. The expected quadratic loss of a linear predictor is then minimized to produce the Bühlmann credibility premium. Much work has been done to extend Bühlmann's model, including Bühlmann and Straub (1970) who generalize the model in cases where volume is involved, and the Hachemeister regression credibility model (Hachemeister, 1975) which introduces covariates to the conditional mean of losses.

These models are common in one regard: losses are independent among entities. Risk parameters are assumed to be independent for different entities. Losses are also assumed to be independent conditional on risk parameters. As a result, losses are unconditionally independent among entities.

Extensions of credibility models to allow for dependence among entities through risk parameters are common in the literature. A prominent example is the Jewell's hierarchical credibility model (Jewell, 1975) which introduces dependence among entities with its hierarchical Bayesian model formulation. Another example is the common effects model, considered by Yeo and Valdez (2006) and Wen et al. (2009), which introduces dependence through an additional cross-entity latent variable. Even more generally, there are the crossed classification models by Dannenburg (1995) and Goulet (2001). Finally, Wen and Wu (2011) propose a general framework through dependence among risk parameters.

Another method to introduce dependence among losses is to alter the assumption of conditional independence in losses. Conditional temporal dependence has been studied extensively. Some examples are Frees et al. (1999) who study credibility predictors under the longitudinal data framework, Frees and Wang (2006) who use the elliptical copula to model temporal dependence, and Lo et al. (2006) and others who consider estimation of regression credibility models with temporal dependence. It is not as common to have conditional cross-sectional dependence in credibility theory. An exception is Schnapp et al. (2000) who propose a basic credibility model with conditional spatial cross-sectional dependence.

In this project, spatial dependence in credibility is a focus, as in Schnapp et al. (2000). Previously in the actuarial literature, parametric hierarchical spatial models have been used by, for example, Boskov and Verrall (1994), Fahrmeir et al. (2007), and Gschlöbl and Czado (2007). Spatial smoothing techniques are also investigated in Taylor (1989, 2001). However, the topic of spatial credibility models seems to have only been covered in Schnapp et al. (2000).

On the other hand, in the spatial statistics literature, kriging is a major field of interest with a long history, going back to Matheron (1963). Kriging enables optimal prediction in geostatistical spatial models via the method of minimization of squared prediction errors, which is equivalent to the method used in credibility theory for optimal prediction. In particular, in Bayesian kriging there are models (see, for example, Omre (1987) and Omre and Halvorsen (1989)) that are very similar to credibility models but with spatial dependence. In fact, credibility models with spatial dependence, like in Schnapp et al. (2000), can be considered as non-parametric spatio-temporal Bayesian Kriging models. However, in the spatial and spatio-temporal statistics literature, it appears that the preferred method of Bayesian kriging is parametric (see, for example, Cressie and Wilkie (2011) and Banerjee et al. (2003)).

The objective of this project is to extend the work of Wen and Wu (2011) to allow for dependence through both risk parameters and process covariance and further study estimation methods for models with spatial cross-sectional dependence as in Schnapp et al. (2000). Since these two works are of importance in the development of this project, in the next two subsections, an overview of these two publications is provided.

1.1.1 Ratemaking Considerations for Multiple Peril Crop Insurance

Schnapp et al. (2000) is a ratemaking discussion paper published in Casualty Actuarial Society Forum, a non-refereed journal. The paper describes the ratemaking process of Multiple Peril Crop Insurance (MPCI) and discusses its various problems and potential improvements.

MPCI is an insurance program that is offered to farm producers to protect them from financial losses due to low yields. The program is a collaboration between the government and the private insurance sector, where the Risk Management Agency (RMA) of the United States Department of Agriculture (USDA) sets MPCI rates and rules, subsidizes premiums and administrative costs, and administers reinsurance arrangements with participating insurers who in turn sell the insurance and pay the benefits. Coverage is offered to producers in almost all states.

In a step in the ratemaking process of MPCI, the concentric circle method is used to smooth pure premiums across counties to take into account large-scale geographic effects that induce correlations among loss experience in different counties. Smoothed pure premiums are computed by taking a weighted average of pure premiums of a county and its neighbours grouped by concentric circles. The circles are determined by RMA individually, while the weights are calculated based on the liability of each county.

Since this method to address spatial dependence is relatively simplistic, Schnapp et al. (2000) discuss three alternatives to the concentric circle method:

1. Create fixed rating territories of nearby counties with similar characteristics,
2. Use more sophisticated spatial smoothing approaches such as locally weighted regression smoothing, and
3. Extend the concept of credibility to consider spatial and intertemporal correlations between territories.

Although each alternative has its own advantages and disadvantages, the third alternative is of particular interest. In the appendix, Schnapp et al. (2000) propose the following model for the loss costs, X_{iu}, of county $i = 1, \dots, K$ and period $u = 1, \dots, n$:

$$X_{iu} = m + R_i + Q_{iu}, \tag{1.1}$$

where m denotes the mean loss over all counties, R_i is a random spatial effect on the mean loss on county i, and Q_{iu} represents a spatially correlated random fluctuation for county i in period u. The expectations of R_i and Q_{iu} are assumed to be 0 for all i and all u. The random time-independent spatial effects $\{R_i\}$ are assumed to be independent of the random fluctuations $\{Q_{iu}\}$. Furthermore, the vectors of random fluctuations across counties $(Q_{i1}, \dots, Q_{iK})$, for $i = 1, \dots, n$, are independent and identically distributed.

Additionally, Schnapp et al. (2000) assume that the spatial correlation among R_i's is strictly a function of distances between counties. The same is also assumed for the spatial dependence among the Q_{iu}'s. This represents a geostatistical approach to modelling lattice data with assumptions of second-order stationary and isotropy for R_i and Q_{iu} (see Chapter 2).

To obtain a linear credibility estimator of the loss for county i in a future year $n + 1$,

$X_{i,n+1}$, the coefficients a_0 and $\{a_{jv}\}_{\forall j,v}$ that minimize the mean square prediction error

$$\mathrm{E}\left[\left(X_{i,n+1} - \left(a_0 + \sum_j \sum_v a_{jv} X_{jv}\right)\right)^2\right] \tag{1.2}$$

are solved using a system of equations.

Some considerations for the structure of spatial dependence for R_i and Q_{iu} are discussed in the report. However, in applications, the parameters of spatial dependence are unknown, but estimation of these structural parameters is not discussed in the report.

1.1.2 Credibility Estimator with General Dependence Structure Over Risks

In Wen and Wu (2011), linear credibility estimators are generalized to allow for a general dependence structure over risks.

As remarked in the paper, there is only a limited amount of literature on credibility models with dependence amongst risks. To illustrate, in Bühlmann's credibility model (Bühlmann, 1967), the losses in different periods, $X_{i1}, \ldots, X_{in}$, for policy i are independent and identically distributed conditional on the unknown, random risk parameter Θ_i that represents the true risk characteristics of policy i. Furthermore, these risk parameters $\{\Theta_1, \ldots, \Theta_K\}$, are assumed to be independent and identically distributed among all K policies. As a result, losses of different policies have the property of unconditional independence. Many credibility models have retained similar specifications. For some applications, this may not be appropriate; there are cases where the losses of policies may well be correlated.

Allowing for a general dependence structure for the risk parameters $\{\Theta_1, \ldots, \Theta_K\}$, the authors of the paper generalize the Bühlmann credibility estimator. Denote the vector of losses of the K policies in the future period $n+1$ by $\boldsymbol{X}_{(n+1)} = (X_{1,n+1}, \ldots, X_{K,n+1})'$ and let $\overline{\boldsymbol{X}} = (\overline{X}_1, \ldots, \overline{X}_k)'$, where $\overline{X}_i = \sum_u X_{iu}$. Further, let $\boldsymbol{\mu} = \mathrm{E}[(X_{11}, \ldots, X_{K1})']$. Then, the following credibility estimator for $\boldsymbol{X}_{(n+1)}$ is obtained by minimizing the mean square prediction error in (1.2):

$$\widehat{\boldsymbol{X}}_{(n+1)} = \boldsymbol{Z}\overline{\boldsymbol{X}} + (\boldsymbol{I}_K - \boldsymbol{Z})\boldsymbol{\mu}, \tag{1.3}$$

where $\boldsymbol{Z} = \boldsymbol{F}[\boldsymbol{F} + \boldsymbol{\Sigma}]^{-1}$, $\boldsymbol{F} = \mathrm{Cov}[(\mathrm{E}[X_{11}|\Theta_1], \ldots, \mathrm{E}[X_{K1}|\Theta_K])']$, and $\boldsymbol{\Sigma} = \mathrm{E}[\mathrm{Cov}[(X_{11}, \ldots, X_{K1})'|\boldsymbol{\Theta}]]$.

Similar results are obtained for extensions to the Bühlmann-Straub model (Bühlmann and Straub, 1970) and regression credibility model (Hachemeister, 1975) to accommodate general dependence of risk parameters. The homogeneous credibility estimator (see Chapter 2) is also considered.

Note that there are no discussions about the estimation of structural parameters $\boldsymbol{\mu}$, $\boldsymbol{F}$, and $\boldsymbol{\Sigma}$ in their paper, as in Schnapp et al. (2000). This is presumably because it is difficult to do so with conventional estimation methods if no further assumptions on the structure of dependence among risks are made.

1.2 Contributions

Summarizing, in Schnapp et al. (2000), a credibility predictor of a Bühlmann-like model is derived in a setting that allows for both spatial dependence among risk parameters and spatial dependence among random fluctuations (i.e. conditional dependence among losses). In Wen and Wu (2011), only general dependence of risk parameters is considered but prediction is further done for Bühlmann's, Bühlmann-Straub and regression credibility models. However, neither papers discuss the estimation of structural parameters required to use the predictors in practice, as mentioned previously.

In this project, the following contributions are made:

1. Credibility predictors are derived for the Bühlmann's, Bühlmann-Straub and regression credibility models with both general dependence among risk parameters and conditional cross-sectional dependence among losses.

2. Non-parametric estimators of structural parameters required in credibility prediction are studied in a spatial statistics context.

This project is arranged in the following way. In the next chapter, the basic theory of credibility and some concepts from spatial statistics are covered. In Chapter 3, credibility predictors for a Bühlmann-type credibility model with general dependence structure among risk and conditional cross-sectional dependence are derived. An estimation method under a spatial statistics context is also discussed. Chapter 4 and Chapter 5 extend the predictors and estimation methods to the Bühlmann-Straub and regression credibility cases. In Chapter 6, an application of the Bühlmann-Straub credibility model proposed to multi-peril crop insurance data is explored. Then, a simulation study is performed to compare predictors and estimation methods in Chapter 7. Finally, Chapter 8 concludes this project.

Chapter 2

Preliminaries

In this chapter, well-known concepts and models in credibility and spatial statistics that are helpful to the development of this project are reviewed.

2.1 Credibility Theory

In this section, the basics of credibility theory in the Actuarial Science literature are reviewed. The development of the content in this section closely follows Wen and Wu (2011). For a more detailed development of credibility theory, see Bühlmann and Gisler (2005).

The main purpose of credibility models is to predict the loss (or any other quantities of interest) in a future time period for a certain entity (for example, policy or region) using previously observed data. Suppose that there are K entities and that for $i = 1, \ldots, K$, entity i has n_i periods of experience. Let X_{iu} denote the random previous losses in time period u for entity i, where $u = 1, \ldots, n_i$, $i = 1, \ldots, K$. Also, let $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{i,n_i})'$ be the vector of random losses for entity i and $\boldsymbol{X} = (\boldsymbol{X}'_1, \ldots, \boldsymbol{X}'_K)'$ be the vector of all previous losses across all K entities. Further, let $\boldsymbol{X}_{(n+1)} = (X_{1,n_1+1}, \ldots, X_{K,n_K+1})'$ be the vector losses in a future period. Then, mathematically, credibility models enable the prediction of $\boldsymbol{X}_{(n+1)}$ given $\boldsymbol{X}$.

There are many possible predictors for all or some elements of $\boldsymbol{X}_{(n+1)}$. Suppose that X_{i,n_i+1} is the quantity of interest. Given the class of predictor functions $G = \{g(\boldsymbol{X}) : g \text{ is a measurable function of } \boldsymbol{X}\}$, the optimality of predictors $g(\boldsymbol{X}) \in G$ of X_{i,n_i+1} can be evaluated using the expectation of a loss function. The commonly seen quadratic loss function is used in this project, which means that the optimal $g(\boldsymbol{X}) \in G$ can be determined by solving the following problem:

$$\min_{g \in G} \mathrm{E}\left[(X_{i,n_i+1} - g(\boldsymbol{X}))^2\right]. \tag{2.1}$$

The solution to the problem in (2.1) is commonly known as the Bayes premium, namely $\mathrm{E}[X_{i,n_i+1}|\boldsymbol{X}]$. However, the use of Bayes premium requires strong distributional assumptions on $\boldsymbol{X}$ and $\boldsymbol{X}_{(n+1)}$. Therefore, it is common to consider only the class of linear functions of past observations, which requires specification of the joint distribution up to second order moments only.

The two classes of linear functions that are used in this project are defined as follows. Let $L(\boldsymbol{X}, 1)$ denote the class of inhomogeneous linear functions

$$L(\boldsymbol{X}, 1) := \left\{ c_o + \sum_{j=1}^{K} \boldsymbol{c}_j' \boldsymbol{X}_j : c_0 \in \mathbb{R} \text{ and } \boldsymbol{c}_j \in \mathbb{R}^{n_j}, j = 1, \ldots, K \right\}, \tag{2.2}$$

and let $L_e(\boldsymbol{X})$ denote the class of homogeneous linear functions

$$L_e(\boldsymbol{X}) := \left\{ \sum_{j=1}^{K} \boldsymbol{c}_j' \boldsymbol{X}_j : \boldsymbol{c}_j \in \mathbb{R}^{n_i}, j = 1, \ldots, K \text{ and } \mathrm{E}\left[\sum_{j=1}^{K} \boldsymbol{c}_j' \boldsymbol{X}_j\right] = \mathrm{E}[X_{i,n_i+1}] \right\}. \tag{2.3}$$

The inhomogeneous and homogeneous credibility premiums of X_{i,n_i+1} are defined, respectively, as the linear functions $\widehat{X}_{i,n_i+1} \in L(\boldsymbol{X}, 1)$ and $\widehat{X}^{hom}_{i,n_i+1} \in L_e(\boldsymbol{X})$ that minimize the mean square error in (2.1).

In Wen et al. (2009), the following lemma is proved.

Lemma 2.1. Let $\boldsymbol{X}$ be a random vector in $\mathbb{R}^p$ with expectation $\boldsymbol{\mu}_X$ and let $\boldsymbol{Y}$ be a random vector in $\mathbb{R}^q$ with expectation $\boldsymbol{\mu}_Y$. Also, let the covariance matrix of $\boldsymbol{X}$ be $\boldsymbol{\Sigma}_{XX} = \mathrm{Cov}[\boldsymbol{X}]$ and the covariance matrix of $\boldsymbol{X}$ and $\boldsymbol{Y}$ be $\boldsymbol{\Sigma}_{YX} = \mathrm{Cov}[\boldsymbol{Y}, \boldsymbol{X}]$. Suppose that $\boldsymbol{\Sigma}_{XX}$ is invertible. Then,

1. $\mathrm{E}[(\boldsymbol{Y} - \boldsymbol{A} - \boldsymbol{B}\boldsymbol{X})(\boldsymbol{Y} - \boldsymbol{A} - \boldsymbol{B}\boldsymbol{X})']$, for $\boldsymbol{A} \in \mathbb{R}^q$ and $\boldsymbol{B} \in \mathbb{R}^q \times \mathbb{R}^p$, can be minimized in the Loewner partial order of matrices by

$$\boldsymbol{A} = \boldsymbol{\mu}_Y - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\mu}_X \quad \text{and} \quad \boldsymbol{B} = \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}.$$

2. Under the constraint $\boldsymbol{\mu}_Y = \boldsymbol{C}\boldsymbol{\mu}_X$, $\mathrm{E}[(\boldsymbol{Y} - \boldsymbol{C}\boldsymbol{X})(\boldsymbol{Y} - \boldsymbol{C}\boldsymbol{X})']$ can be minimized in the Loewner partial order of matrices by

$$\boldsymbol{C} = \left(\boldsymbol{\Sigma}_{YX} + \frac{\left(\boldsymbol{\mu}_Y - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\mu}_X\right)\boldsymbol{\mu}_X'}{\boldsymbol{\mu}_X'\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\mu}_X} \right) \boldsymbol{\Sigma}_{XX}^{-1}.$$

Lemma 2.1 states that the inhomogeneous linear credibility premium for $\boldsymbol{Y}$ that minimizes the expected square loss is the orthogonal projection (in the $\mathcal{L}^2$ Hilbert space) of $\boldsymbol{Y}$ on $L(\boldsymbol{X}, 1)$:

$$\text{Proj}(\boldsymbol{Y}|L(\boldsymbol{X},1)) = \boldsymbol{\mu}_Y + \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}_X), \tag{2.4}$$

and that the homogeneous linear credibility premium for $\boldsymbol{Y}$ is the orthogonal projection of $\boldsymbol{Y}$ on $L_e(\boldsymbol{X})$:

$$\text{Proj}(\boldsymbol{Y}|L_e(\boldsymbol{X})) = \left(\boldsymbol{\Sigma}_{YX} + \frac{\left(\boldsymbol{\mu}_Y - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\mu}_X\right)\boldsymbol{\mu}_X'}{\boldsymbol{\mu}_X'\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\mu}_X}\right)\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{X}. \tag{2.5}$$

Note that a property of the inhomogeneous linear credibility predictor is that it is unbiased. This is an implication of one of the normal equations that needs to be satisfied in an orthogonal projection (see Bühlmann and Gisler (2005)). For the homogeneous linear credibility predictor, its unbiasedness is obtained through the explicit constraint $\boldsymbol{\mu}_Y = \boldsymbol{C}\boldsymbol{\mu}_X$.

Define the mean square prediction error matrix for the estimator $g(\boldsymbol{X})$ as $\text{E}[(\boldsymbol{Y} - g(\boldsymbol{X}))(\boldsymbol{Y} - g(\boldsymbol{X}))']$. Often, one is interested in the prediction error of the proposed linear estimators. The following lemma can be used to determine the mean square prediction error matrix for the inhomogeneous and homogeneous credibility predictors.

Lemma 2.2. Continuing with Lemma 2.1, denote $\text{Cov}[\boldsymbol{Y}] = \boldsymbol{\Sigma}_{YY}$.

1. Let $\widehat{\boldsymbol{Y}}^{inhom} = \boldsymbol{A} + \boldsymbol{B}\boldsymbol{X}$. Then, the mean square prediction error matrix of the inhomogeneous predictor $\widehat{\boldsymbol{Y}}^{inhom}$ is given by

$$\text{E}\left[\left(\boldsymbol{Y} - \widehat{\boldsymbol{Y}}^{inhom}\right)\left(\boldsymbol{Y} - \widehat{\boldsymbol{Y}}^{inhom}\right)'\right] = \boldsymbol{\Sigma}_{YY} - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{YX}'. \tag{2.6}$$

2. Let $\widehat{\boldsymbol{Y}}^{hom} = \boldsymbol{C}\boldsymbol{X}$. Further assume that $\boldsymbol{\mu}_X = \mu\boldsymbol{1}_p$ and $\boldsymbol{\mu}_Y = \mu\boldsymbol{1}_q$, where $\boldsymbol{1}_r$ is a $r \times 1$ vector of ones. Then, the mean square prediction error matrix of the inhomogeneous predictor $\widehat{\boldsymbol{Y}}^{hom}$ is given by

$$\begin{aligned}\text{E}&\left[\left(\boldsymbol{Y} - \widehat{\boldsymbol{Y}}^{hom}\right)\left(\boldsymbol{Y} - \widehat{\boldsymbol{Y}}^{hom}\right)'\right] \\ =\ & \boldsymbol{\Sigma}_{YY} - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{YX}' + \frac{\left(\boldsymbol{1}_q - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{1}_p\right)\left(\boldsymbol{1}_q - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{1}_p\right)'}{\boldsymbol{1}_p'\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{1}_p}.\end{aligned} \tag{2.7}$$

Proof. First, we have

$$\begin{aligned}\mathrm{E}[(\boldsymbol{Y}-g(\boldsymbol{X}))(\boldsymbol{Y}-g(\boldsymbol{X}))'] &= \mathrm{Cov}[\boldsymbol{Y}-g(\boldsymbol{X})]+\mathrm{E}[\boldsymbol{Y}-g(\boldsymbol{X})]\,\mathrm{E}[(\boldsymbol{Y}-g(\boldsymbol{X}))'] \\ &= \mathrm{Cov}[\boldsymbol{Y}]+\mathrm{Cov}[g(\boldsymbol{X})]-\mathrm{Cov}[\boldsymbol{Y},g(\boldsymbol{X})]-\mathrm{Cov}[g(\boldsymbol{X}),\boldsymbol{Y}] \\ &\quad +\mathrm{E}[\boldsymbol{Y}-g(\boldsymbol{X})]\,\mathrm{E}[(\boldsymbol{Y}-g(\boldsymbol{X}))'].\end{aligned}$$

It follows that

$$\begin{aligned}&\mathrm{E}\left[\left(\boldsymbol{Y}-\widehat{\boldsymbol{Y}}^{inhom}\right)\left(\boldsymbol{Y}-\widehat{\boldsymbol{Y}}^{inhom}\right)'\right] \\ &= \mathrm{Cov}[\boldsymbol{Y}]+\mathrm{Cov}\left[\widehat{\boldsymbol{Y}}^{inhom}\right]-\mathrm{Cov}\left[\boldsymbol{Y},\widehat{\boldsymbol{Y}}^{inhom}\right]-\mathrm{Cov}\left[\widehat{\boldsymbol{Y}}^{inhom},\boldsymbol{Y}\right] \\ &\quad +\mathrm{E}\left[\boldsymbol{Y}-\widehat{\boldsymbol{Y}}^{inhom}\right]\mathrm{E}\left[(\boldsymbol{Y}-\widehat{\boldsymbol{Y}}^{inhom})'\right] \\ &= \boldsymbol{\Sigma}_{YY}+\boldsymbol{B}\boldsymbol{\Sigma}_{XX}\boldsymbol{B}'-\boldsymbol{\Sigma}_{YX}\boldsymbol{B}'-\boldsymbol{B}\boldsymbol{\Sigma}_{YX}'+0 \\ &= \boldsymbol{\Sigma}_{YY}+\left(\boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\right)\boldsymbol{\Sigma}_{XX}\left(\boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\right)' \\ &\quad -\boldsymbol{\Sigma}_{YX}\left(\boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\right)'-\left(\boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\right)\boldsymbol{\Sigma}_{YX}' \\ &= \boldsymbol{\Sigma}_{YY}+\boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{YX}'-\boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{YX}'-\boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{YX}' \\ &= \boldsymbol{\Sigma}_{YY}-\boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{YX}',\end{aligned}$$

since $\widehat{\boldsymbol{Y}}^{inhom}$ is unbiased for $\boldsymbol{Y}$, which proves (2.6). Similarly,

$$\begin{aligned}&\mathrm{E}\left[\left(\boldsymbol{Y}-\widehat{\boldsymbol{Y}}^{hom}\right)\left(\boldsymbol{Y}-\widehat{\boldsymbol{Y}}^{hom}\right)'\right] \\ &= \mathrm{Cov}\left[\boldsymbol{Y}\right]+\mathrm{Cov}\left[\widehat{\boldsymbol{Y}}^{hom}\right]-\mathrm{Cov}\left[\boldsymbol{Y},\widehat{\boldsymbol{Y}}^{hom}\right]-\mathrm{Cov}\left[\widehat{\boldsymbol{Y}}^{hom},\boldsymbol{Y}\right] \\ &\quad +\mathrm{E}\left[\boldsymbol{Y}-\widehat{\boldsymbol{Y}}^{hom}\right]\mathrm{E}\left[\left(\boldsymbol{Y}-\widehat{\boldsymbol{Y}}^{hom}\right)'\right] \\ &= \boldsymbol{\Sigma}_{YY}+\boldsymbol{C}\boldsymbol{\Sigma}_{XX}\boldsymbol{C}'-\boldsymbol{\Sigma}_{YX}\boldsymbol{C}'-\boldsymbol{C}\boldsymbol{\Sigma}_{YX}+0 \\ &= \boldsymbol{\Sigma}_{YY} \\ &\quad +\left(\boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}+\frac{\left(\mathbf{1}_q-\boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p\right)\mathbf{1}_p'}{\mathbf{1}_p'\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p}\boldsymbol{\Sigma}_{XX}^{-1}\right)\boldsymbol{\Sigma}_{XX} \\ &\quad \times\left(\boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}+\frac{\left(\mathbf{1}_q-\boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p\right)\mathbf{1}_p'}{\mathbf{1}_p'\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p}\boldsymbol{\Sigma}_{XX}^{-1}\right)' \\ &\quad -\boldsymbol{\Sigma}_{YX}\left(\boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}+\frac{\left(\mathbf{1}_q-\boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p\right)\mathbf{1}_p'}{\mathbf{1}_p'\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p}\boldsymbol{\Sigma}_{XX}^{-1}\right)' \\ &\quad -\left(\boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}+\frac{\left(\mathbf{1}_q-\boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p\right)\mathbf{1}_p'}{\mathbf{1}_p'\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p}\boldsymbol{\Sigma}_{XX}^{-1}\right)\boldsymbol{\Sigma}_{YX}'\end{aligned}$$

$$
\begin{aligned}
&= \boldsymbol{\Sigma}_{YY} + \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{YX}' + \frac{\left(\mathbf{1}_q - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p\right)\mathbf{1}_p'}{\mathbf{1}_p'\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{YX}' \\
&\quad + \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\frac{\left(\left(\mathbf{1}_q - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p\right)\mathbf{1}_p'\right)'}{\mathbf{1}_p'\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p} \\
&\quad + \frac{\left(\mathbf{1}_q - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p\right)\mathbf{1}_p'}{\mathbf{1}_p'\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p}\boldsymbol{\Sigma}_{XX}^{-1}\frac{\left(\left(\mathbf{1}_q - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p\right)\mathbf{1}_p'\right)'}{\mathbf{1}_p'\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p} \\
&\quad - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{YX}' - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\frac{\left(\left(\mathbf{1}_q - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p\right)\mathbf{1}_p'\right)'}{\mathbf{1}_p'\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p} \\
&\quad - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{YX}' - \frac{\left(\mathbf{1}_q - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p\right)\mathbf{1}_p'}{\mathbf{1}_p'\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{YX}' \\
&= \boldsymbol{\Sigma}_{YY} - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{YX}' + \frac{\left(\mathbf{1}_q - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p\right)\left(\mathbf{1}_q - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p\right)'}{\mathbf{1}_p'\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p},
\end{aligned}
$$

since

$$
\begin{aligned}
\boldsymbol{C} &= \left(\boldsymbol{\Sigma}_{YX} + \frac{\left(\boldsymbol{\mu}_Y - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\mu}_X\right)\boldsymbol{\mu}_X'}{\boldsymbol{\mu}_X'\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\mu}_X}\right)\boldsymbol{\Sigma}_{XX}^{-1} \\
&= \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1} + \frac{\mu^2\left(\mathbf{1}_q - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p\right)\mathbf{1}_p'}{\mu^2\mathbf{1}_p'\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p}\boldsymbol{\Sigma}_{XX}^{-1} \\
&= \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1} + \frac{\left(\mathbf{1}_q - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p\right)\mathbf{1}_p'}{\mathbf{1}_p'\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p}\boldsymbol{\Sigma}_{XX}^{-1},
\end{aligned}
$$

and $\widehat{\boldsymbol{Y}}^{hom}$ is unbiased for $\boldsymbol{Y}$. This completes the proof of (2.7) □

It can be seen that the term $\frac{\left(\mathbf{1}_q - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p\right)\left(\mathbf{1}_q - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p\right)'}{\mathbf{1}_p'\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{1}_p}$ in (2.7) represents the additional mean square prediction error that results from using the homogeneous credibility premium over the inhomogeneous credibility premium. However, in most applications, the unconditional mean μ is unknown and has to be estimated in order to use the inhomogeneous credibility premium, while the homogeneous credibility premium does not require this. Therefore, the actual mean square prediction error of the inhomogeneous credibility premium, if μ is unknown and estimated with some linear combination of $\boldsymbol{X}$, is larger than or equal to that of the homogeneous credibility premium.

Finally, although this assumption is implicitly made in the previous discussion of orthogonal projections, it should be made clear that all random variables in this project are assumed to be in the $\mathcal{L}^2$ Hilbert space, where

$$
\mathcal{L}^2 := \left\{X : X \text{ is a random variable with } \mathrm{E}[X^2] < \infty\right\}.
$$

A well-known and convenient result with orthogonal projections in $\mathcal{L}^2$, the iterativity of projections, is as follows.

Lemma 2.3. Let M^* and M be two closed subspaces of $\mathcal{L}^2$ with $M^* \subset M$ and Y be an element of $\mathcal{L}^2$. Then,

$$\text{Proj}(Y|M^*) = \text{Proj}(\text{Proj}(Y|M)|M^*).$$

This property of orthogonal projections allows projections to be derived in two separate steps.

2.2 Spatio-Temporal Statistics

In this section, the basics of spatial-temporal statistics are reviewed. For more details, see Cressie and Wilkie (2011).

Spatio-temporal covariance functions are a fundamental part of spatio-temporal statistics. In short, spatio-temporal covariance functions describe the degree of linear dependence within random processes.

Important definitions from Cressie and Wilkie (2011) are presented here. First, a covariance function must be a *nonnegative-definite* function.

Definition 2.1. A function $f(\boldsymbol{u}, \boldsymbol{v}) : \boldsymbol{u}, \boldsymbol{v} \in D$ defined on $D \times D$ is said to be *nonnegative-definite*, if for any complex numbers $\{a_i : i = 1, \ldots, m\}$, any $\{\boldsymbol{u}_i : i = 1, \ldots, m\}$ in D, and any integer m, we have

$$\sum_{i=1}^{m} \sum_{j=1}^{m} a_i \bar{a}_j f(\boldsymbol{u}_i, \boldsymbol{u}_j) \geq 0$$

.

Usually, spatio-temporal covariance functions are written as $f((\boldsymbol{s}_i; t_i), (\boldsymbol{s}_j; t_j))$, where $\boldsymbol{s}_i$ denotes spatial coordinates or indices and t_i denotes time index.

A *stationary spatio-temporal covariance function* is a spatio-temporal covariance function that depends only on the difference between the spatial locations and time indices. To be specific, we have the following definition.

Definition 2.2. A function f is a *stationary spatio-temporal covariance function* on $\mathbb{R}^d \times \mathbb{R}$, if it satisfies Definition 2.1 and can be written as

$$f((\boldsymbol{s}_i; t_i), (\boldsymbol{s}_j; t_j)) = C(\boldsymbol{s}_i - \boldsymbol{s}_j; t_i - t_j), \qquad \boldsymbol{s}_i, \boldsymbol{s}_j \in \mathbb{R}^d, t_i, t_j \in \mathbb{R}, i, j = 1, \ldots, m,$$

for any integer m.

If, further, a random process $Y(\boldsymbol{s}, t)$ has a constant expectation for all $\boldsymbol{s}$ and t and a stationary covariance function, then the process is said to be *second-order stationary.* For some applications, the implication that covariances and expectations cannot vary with location but only distance could be too restrictive, but unbiased non-parametric estimation may not be possible without assuming stationarity.

Spatial isotropy is a property of spatio-temporal processes when their covariance functions only depend on the distance between two spatial locations, but not directions.

Definition 2.3. A spatio-temporal process $Y(\cdot, \cdot)$ is said to exhibit *spatial isotropy* if its covariance function, $\mathrm{Cov}[Y(\boldsymbol{s}_i; t_i), Y(\boldsymbol{s}_j; t_j)]$, can be written in the following way:

$$\mathrm{Cov}[Y(\boldsymbol{s}_i; t_i), Y(\boldsymbol{s}_j; t_j)] \equiv C(||\boldsymbol{s}_i - \boldsymbol{s}_j||; t_i, t_j), \qquad \boldsymbol{s}_i, \boldsymbol{s}_j \in \mathbb{R}^d, t_i, t_j \in \mathbb{R}, i, j = 1, \ldots, m,$$

where $||\cdot||$ is the vector norm, for any integer m.

The *variogram* is a concept widely used in spatial statistics. Suppose that the following is satisfied:

$$\mathrm{V}[Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j)] = 2\gamma(\boldsymbol{s}_i - \boldsymbol{s}_j), \qquad \text{for all } \boldsymbol{s}_i, \boldsymbol{s}_j \in D,$$

where $Y(\cdot)$ is a spatial process on D. Then, the quantity $2\gamma(\cdot)$ is called the variogram and $\gamma(\cdot)$ is called the semivariogram. If $C(\cdot)$ is the corresponding *covariogram* (i.e., a stationary spatial covariance function), then

$$\gamma(\boldsymbol{h}) = C(\boldsymbol{0}) - C(\boldsymbol{h}).$$

Although the covariogram is more easily interpreted than the variogram, the variogram exists under a more general class of processes (*intrinsically stationary processes*) and there are advantages with unbiasedness and asymptotic properties with estimating the variogram rather than the covariogram in a spatial context. Note that the variogram has to be *conditionally negative-definite* (see Cressie (1993)). Parametric variograms that are used in this project and their interpretation can be found in Appendix .

Chapter 3

Bühlmann's Credibility Model with General Dependence Structure and Conditional Cross-Sectional Dependence

In this chapter, Bühlmann's credibility model with general dependence structure in Wen and Wu (2011) is first generalized to allow for conditional cross-sectional dependence among observations. Then, non-parametric estimation methods in the spatial context are proposed to estimate the structural parameters of the generalized model.

3.1 The Credibility Estimator

In Section 1.1.2, the Bühlmann's credibility model with general dependence structure is described. In this section, we assume conditional dependence among observations of different risks in the same time period. Precisely, the model assumptions are as follows.

Assumption 3.1. Conditional on the vector of risk parameters $\mathbf{\Theta} = (\Theta_1, \ldots, \Theta_K)'$, the cross-sectional vectors of claims in a single period, $\boldsymbol{X}_{(u)} = (X_{1u}, \ldots, X_{Ku})'$, are independent and identically distributed for $u = 1, \ldots, n+1$ with conditional moments $\mathrm{E}[\boldsymbol{X}_{(u)}|\mathbf{\Theta}] = \boldsymbol{\mu}(\mathbf{\Theta}) = (\mu(\Theta_1), \ldots, \mu(\Theta_K))'$ and $\mathrm{Cov}[\boldsymbol{X}_{(u)}|\mathbf{\Theta}] = \mathbf{\Sigma}(\mathbf{\Theta}) = [\sigma^2(\Theta_i, \Theta_j)]_{i,j=1,\ldots,K}$.

Assumption 3.2. The distribution of $\boldsymbol{\Theta}$ is such that $\mathrm{E}[\mu(\boldsymbol{\Theta})] = \boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)'$, $\mathrm{E}[\boldsymbol{\Sigma}(\boldsymbol{\Theta})] = \boldsymbol{V} = [v_{ij}]_{i,j=1,\ldots,K}$, and $\mathrm{Cov}[\mu(\boldsymbol{\Theta})] = \boldsymbol{A} = [a_{ij}]_{i,j=1,\ldots,K}$.

Using credibility theory terminology, $\boldsymbol{\mu}$ is the expectation of the hypothetical means, $\boldsymbol{V}$ is the expected process covariance matrix, and $\boldsymbol{A}$ is the covariance matrix of the hypothetical means.

As mentioned previously, the goal is to predict the future claims $\boldsymbol{X}_{(n+1)}$ for all risks, given the knowledge of previous claims $\boldsymbol{X} = (\boldsymbol{X}_1', \ldots, \boldsymbol{X}_K')'$, where $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{in_i})'$. Note that it is assumed that the number of periods available are equal for all risks: $n_i = n$ for all $i = 1, \ldots, K$. This simplifies derivations and formulas when conditional cross-sectional dependence is present.

The following lemma states results that can be obtained about the dependence features of the model described.

Lemma 3.1. Under Assumptions 3.1 and 3.2 and with the notation in this section,

1. The expectations of $\boldsymbol{X}$ and $\boldsymbol{X}_{(n+1)}$ are

$$\boldsymbol{\mu}_X = E[\boldsymbol{X}] = \boldsymbol{\mu} \otimes \mathbf{1}_n \quad \text{and} \quad \boldsymbol{\mu}_{X_{(n+1)}} = E[\boldsymbol{X}_{(n+1)}] = \boldsymbol{\mu},$$

   where $\otimes$ denotes the Kronecker product operator and $\mathbf{1}_n$ is a length-n column vector with all of its n elements equal to 1.

2. The covariance matrix of $\boldsymbol{X}$ is

$$\boldsymbol{\Sigma}_{XX} = \mathrm{Cov}[\boldsymbol{X}] = \boldsymbol{U}\boldsymbol{A}\boldsymbol{U}' + \boldsymbol{V} \otimes \boldsymbol{I}_n,$$

   where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix and

$$\boldsymbol{U}_{Kn\times K} = \begin{bmatrix} \mathbf{1}_n & 0 & \cdots & \cdots & 0 \\ 0 & \mathbf{1}_n & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & 0 & \mathbf{1}_n \end{bmatrix}.$$

3. The covariance matrix of $\boldsymbol{X}_{(n+1)}$ and $\boldsymbol{X}$ is

$$\boldsymbol{\Sigma}_{X_{(n+1)}X} = \mathrm{Cov}[\boldsymbol{X}_{(n+1)}, \boldsymbol{X}] = \boldsymbol{A}\boldsymbol{U}'.$$

4. The inverse of the covariance matrix $\mathbf{\Sigma}_{XX}$ is

$$\mathbf{\Sigma}_{XX}^{-1} = (\boldsymbol{V}^{-1} \otimes \boldsymbol{I}_n) - (\boldsymbol{V}^{-1} \otimes \mathbf{1}_n)\boldsymbol{U}(\boldsymbol{A}^{-1} + n\boldsymbol{V}^{-1})^{-1}\boldsymbol{U}'(\boldsymbol{V}^{-1} \otimes \mathbf{1}_n').$$

Proof. 1. Immediately follows from definitions and notations.

2. Since $\mathrm{Cov}[\mathrm{E}[\boldsymbol{X}_i|\mathbf{\Theta}], \mathrm{E}[\boldsymbol{X}_i|\mathbf{\Theta}]] = a_{ij}\mathbf{1}_n\mathbf{1}_n'$ and $\mathrm{E}[\mathrm{Cov}[\boldsymbol{X}_i, \boldsymbol{X}_j|\mathbf{\Theta}]] = v_{ij}\boldsymbol{I}_n$ for $i, j = 1, \ldots, K$, we have

$$\begin{aligned}\mathbf{\Sigma}_{XX} &= \mathrm{Cov}[\boldsymbol{X}] \\ &= \mathrm{Cov}[\mathrm{E}[\boldsymbol{X}|\mathbf{\Theta}]] + \mathrm{E}[\mathrm{Cov}[\boldsymbol{X}|\mathbf{\Theta}]] \\ &= \left(\boldsymbol{A} \otimes \mathbf{1}_n\mathbf{1}_n'\right) + (\boldsymbol{V} \otimes \boldsymbol{I}_n) \\ &= \boldsymbol{U}\boldsymbol{A}\boldsymbol{U}' + \boldsymbol{V} \otimes \boldsymbol{I}_n.\end{aligned}$$

3. Since $\mathrm{E}[\mathrm{Cov}[\boldsymbol{X}_{(n+1)}, \boldsymbol{X}|\mathbf{\Theta}]] = 0$ by model assumptions,

$$\begin{aligned}\mathbf{\Sigma}_{X_{(n+1)}X} &= \mathrm{Cov}[\boldsymbol{X}_{(n+1)}, \boldsymbol{X}] \\ &= \mathrm{Cov}[\mathrm{E}[\boldsymbol{X}_{(n+1)}|\mathbf{\Theta}], \mathrm{E}[\boldsymbol{X}|\mathbf{\Theta}]] + \mathrm{E}[\mathrm{Cov}[\boldsymbol{X}_{(n+1)}, \boldsymbol{X}|\mathbf{\Theta}]] \\ &= \boldsymbol{A}\boldsymbol{U}'.\end{aligned}$$

4. We obtain

$$\begin{aligned}\mathbf{\Sigma}_{XX}^{-1} &= \left(\boldsymbol{V} \otimes \boldsymbol{I}_n + \boldsymbol{U}\boldsymbol{A}\boldsymbol{U}'\right)^{-1} \\ &= (\boldsymbol{V} \otimes \boldsymbol{I}_n)^{-1} - (\boldsymbol{V} \otimes \boldsymbol{I}_n)^{-1}\boldsymbol{U}(\boldsymbol{A}^{-1} + \boldsymbol{U}'(\boldsymbol{V} \otimes \boldsymbol{I}_n)^{-1}\boldsymbol{U})^{-1}\boldsymbol{U}'(\boldsymbol{V} \otimes \boldsymbol{I}_n)^{-1} \\ &= (\boldsymbol{V}^{-1} \otimes \boldsymbol{I}_n) - (\boldsymbol{V}^{-1} \otimes \mathbf{1}_n)(\boldsymbol{A}^{-1} + n\boldsymbol{V}^{-1})^{-1}(\boldsymbol{V}^{-1} \otimes \mathbf{1}_n')\end{aligned}$$

by setting $\boldsymbol{E} = \boldsymbol{V} \otimes \boldsymbol{I}_n, \boldsymbol{F} = \boldsymbol{U}, \boldsymbol{G} = \boldsymbol{A}$, and $\boldsymbol{H} = \boldsymbol{U}'$ in the relation $(\boldsymbol{E} + \boldsymbol{F}\boldsymbol{G}\boldsymbol{H})^{-1} = \boldsymbol{E}^{-1} - \boldsymbol{E}^{-1}\boldsymbol{F}(\boldsymbol{G}^{-1} + \boldsymbol{H}\boldsymbol{E}^{-1}\boldsymbol{F})^{-1}\boldsymbol{H}\boldsymbol{E}^{-1}$.

□

With the necessary quantities of interest laid out in Lemma 3.1, the inhomogeneous credibility premium for the future claims $\boldsymbol{X}_{(n+1)}$ can be obtained.

Theorem 3.2. Under Assumptions 3.1 and 3.2 and the notations in this section, the inhomogeneous credibility premium of $\boldsymbol{X}_{(n+1)}$ that is obtained by minimizing (2.1) in the class of inhomogeneous linear functions $L(\boldsymbol{X}, 1)$ is given by

$$\widehat{\boldsymbol{X}}_{(n+1)} = \boldsymbol{Z}\overline{\boldsymbol{X}} + (\boldsymbol{I}_K - \boldsymbol{Z})\boldsymbol{\mu}, \tag{3.1}$$

where $\boldsymbol{Z} = \boldsymbol{A}\left(\frac{1}{n}\boldsymbol{V} + \boldsymbol{A}\right)^{-1}$, $\overline{\boldsymbol{X}} = (\overline{X}_1, \ldots, \overline{X}_K)'$, and $\overline{X}_i = \frac{1}{n}\sum_{u=1}^{n} X_{iu}$. The associated mean square prediction error matrix is

$$\mathrm{E}[(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}_{(n+1)})(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}_{(n+1)})'] = (\boldsymbol{I}_K - \boldsymbol{Z})\boldsymbol{A} + \boldsymbol{V}. \tag{3.2}$$

Proof. First, by Lemma 3.1

$$\begin{aligned}
&\boldsymbol{\Sigma}_{X_{(n+1)}X}\boldsymbol{\Sigma}_{XX}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}_X) \\
&= \boldsymbol{A}\boldsymbol{U}'\left\{(\boldsymbol{V}^{-1} \otimes \boldsymbol{I}_n) - (\boldsymbol{V}^{-1} \otimes \mathbf{1}_n)(\boldsymbol{A}^{-1} + n\boldsymbol{V}^{-1})^{-1}(\boldsymbol{V}^{-1} \otimes \mathbf{1}_n')\right\}(\boldsymbol{X} - \boldsymbol{\mu}_X) \\
&= \left\{\boldsymbol{A}(\boldsymbol{V}^{-1} \otimes \mathbf{1}_\mathbf{n}') - \boldsymbol{A}\boldsymbol{U}'(\boldsymbol{V}^{-\mathbf{1}} \otimes \mathbf{1}_\mathbf{n})(\boldsymbol{A}^{-\mathbf{1}} + \mathbf{n}\boldsymbol{V}^{-\mathbf{1}})^{-\mathbf{1}}(\boldsymbol{V}^{-\mathbf{1}} \otimes \mathbf{1}_\mathbf{n}')\right\}(\boldsymbol{X} - \boldsymbol{\mu}_X) \\
&= \left\{\boldsymbol{A} - \boldsymbol{A}n\boldsymbol{V}^{-1}(\boldsymbol{A}^{-1} + n\boldsymbol{V}^{-1})^{-1}\right\}(\boldsymbol{V}^{-1} \otimes \mathbf{1}_n')(\boldsymbol{X} - \boldsymbol{\mu}_X) \\
&= \left\{\boldsymbol{A} - \boldsymbol{A}n\boldsymbol{V}^{-1}(\boldsymbol{A}^{-1} + n\boldsymbol{V}^{-1})^{-1}\right\}n\boldsymbol{V}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}).
\end{aligned}$$

Using the relation $(\boldsymbol{E} + \boldsymbol{F})^{-1} = \boldsymbol{E}^{-1} - \boldsymbol{E}^{-1}\boldsymbol{F}(\boldsymbol{E} + \boldsymbol{F})^{-1}$, we have

$$\boldsymbol{\Sigma}_{X_{(n+1)}X}\boldsymbol{\Sigma}_{XX}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}_X) = (\boldsymbol{A}^{-1} + n\boldsymbol{V}^{-1})^{-1}n\boldsymbol{V}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})$$

Now apply $(\boldsymbol{E} + \boldsymbol{F})^{-1}\boldsymbol{F} = \boldsymbol{E}^{-1}(\boldsymbol{F}^{-1} + \boldsymbol{E}^{-1})^{-1}$,

$$\begin{aligned}
\boldsymbol{\Sigma}_{X_{(n+1)}X}\boldsymbol{\Sigma}_{XX}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}_X) &= \boldsymbol{A}\left(\frac{1}{n}\boldsymbol{V} + \boldsymbol{A}\right)^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}) \\
&= \boldsymbol{Z}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}).
\end{aligned}$$

Finally, with Lemma 2.1 we obtain

$$\begin{aligned}
\widehat{\boldsymbol{X}}_{(n+1)} &= \boldsymbol{\mu} + \boldsymbol{\Sigma}_{X_{(n+1)}X}\boldsymbol{\Sigma}_{XX}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}_X) \\
&= \boldsymbol{\mu} + \boldsymbol{Z}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}) \\
&= \boldsymbol{Z}\overline{\boldsymbol{X}} + (\boldsymbol{I}_K - \boldsymbol{Z})\boldsymbol{\mu},
\end{aligned}$$

which concludes the proof for (3.1). For the mean square prediction error (3.2), since

$$\begin{aligned}
\mathrm{Cov}[X_{i,n+1}, \overline{X}_j] &= \frac{1}{n}\sum_{v=1}^{n}\mathrm{Cov}[X_{i,n+1}, X_{jv}] \\
&= a_{ij}
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Cov}[\overline{X}_i, \overline{X}_j] &= \frac{1}{n^2}\sum_{u=1}^{n}\sum_{v=1}^{n}\mathrm{Cov}[X_{i,u}, X_{jv}] \\
&= \frac{1}{n^2}\sum_{u=1}^{n}\sum_{v=1}^{n}(a_{ij} + \delta_{uv}v_{ij}) \\
&= a_{ij} + \frac{v_{ij}}{n},
\end{aligned}$$

where δ_{uv} is the Kronecker delta function, for $i, j = 1, \ldots, K$, we have

$$\begin{aligned}
&\mathrm{E}[(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}_{(n+1)})(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}_{(n+1)})'] \\
&= \mathrm{Cov}[\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}_{(n+1)}] + \mathrm{E}[\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}_{(n+1)}]\,\mathrm{E}[(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}_{(n+1)})'] \\
&= \mathrm{Cov}[\boldsymbol{X}_{(n+1)}] + \mathrm{Cov}[\widehat{\boldsymbol{X}}_{(n+1)}] - \mathrm{Cov}[\boldsymbol{X}_{(n+1)}, \widehat{\boldsymbol{X}}_{(n+1)}] - \mathrm{Cov}[\widehat{\boldsymbol{X}}_{(n+1)}, \boldsymbol{X}_{(n+1)}] \\
&\quad + \mathrm{E}[\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}_{(n+1)}]\,\mathrm{E}[(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}_{(n+1)})'] \\
&= (\boldsymbol{A} + \boldsymbol{V}) + \boldsymbol{Z}\left(\boldsymbol{A} + \frac{1}{n}\boldsymbol{V}\right)\boldsymbol{Z}' - \boldsymbol{A}\boldsymbol{Z}' - \boldsymbol{Z}\boldsymbol{A} + 0 \\
&= (\boldsymbol{A} + \boldsymbol{V}) + \boldsymbol{A}\boldsymbol{Z}' - \boldsymbol{A}\boldsymbol{Z}' - \boldsymbol{Z}\boldsymbol{A} \\
&= (\boldsymbol{I}_K - \boldsymbol{Z})\boldsymbol{A} + \boldsymbol{V}.
\end{aligned}$$

□

Remark 3.1. It can be seen that as $n \to \infty$, $\boldsymbol{Z} \to \boldsymbol{I}_K$. This means that more credibility will be assigned to the sample mean of the individual regions as sample size becomes larger. For fixed $\boldsymbol{\Theta}$, also note that when $n \to \infty$, $\overline{\boldsymbol{X}} \to \boldsymbol{\mu}(\boldsymbol{\Theta})$ by the central limit theorem, where $\boldsymbol{\mu}(\boldsymbol{\Theta})$ is the vector of individual risk premiums of the K risks. Therefore,

$$\widehat{\boldsymbol{X}}_{(n+1)} = \boldsymbol{Z}\overline{\boldsymbol{X}} + (\boldsymbol{I}_K - \boldsymbol{Z})\boldsymbol{\mu} \to \boldsymbol{\mu}(\boldsymbol{\Theta}), \qquad \text{as } n \to \infty,$$

which means that the inhomogeneous linear credibility estimator $\widehat{\boldsymbol{X}}_{(n+1)}$ is consistent for the individual risk premiums $\boldsymbol{\mu}(\boldsymbol{\Theta})$.

Remark 3.2. If it is further assumed that for $u = 1, \ldots, n$, losses across entities $X_{1u}, \ldots, X_{Ku}|\boldsymbol{\Theta}$ are conditionally independent, it can be seen that $\boldsymbol{V} = diag(v_{11}, v_{22}, \ldots, v_{KK})$. Then,

$$\widehat{\boldsymbol{X}}_{(n+1)} = \boldsymbol{Z}\overline{\boldsymbol{X}} + (\boldsymbol{I}_K - \boldsymbol{Z})\boldsymbol{\mu},$$

where $\boldsymbol{Z} = \boldsymbol{A}(diag(v_{11}/n, \ldots, v_{KK}/n) + \boldsymbol{A})^{-1}$. This is the Bühlmann's credibility model with general dependence structure in Wen and Wu (2011), but with $n_i = n$ for $i = 1, \ldots, K$.

Remark 3.3. Consider the model in (1.1):

$$X_{iu} = m + R_i + Q_{iu},$$

along with the assumptions outlined in Section 1.1.1. Then, the results of Lemma 3.1 and Theorem 3.2 are valid with

$$\begin{aligned}
\boldsymbol{\mu} &= m\mathbf{1}_K \\
\boldsymbol{A} &= \mathrm{Cov}[(R_1, \ldots, R_K)'] \\
\boldsymbol{V} &= \mathrm{Cov}[(Q_{1u}, \ldots, Q_{Ku})'].
\end{aligned}$$

In fact, model (1.1) is a linear representation of the model defined in this section, but with $\mu_i = m$ and $\sigma^2(\cdot,\cdot) = \sigma_{ij}^2$, where σ_{ij}^2 is some non-negative constant, for $i, j = 1, \ldots, K$. This can be seen from the following correspondences:

$$\begin{aligned} \mu(\Theta_i) &= m + R_i \\ X_{iu} - \mu(\Theta_i) &= Q_{iu}, \end{aligned}$$

for $i = 1, \ldots, K$ and $u = 1, \ldots, n$.

For the homogeneous credibility premium, it is necessary to assume equal unconditional means to avoid the need for an estimate of $\boldsymbol{\mu}$.

Assumption 3.3. For the homogeneous credibility premium, assume that for $i = 1, \ldots, K$, $\mu_i = \mu$.

In the following theorem, the homogeneous credibility premium is stated.

Theorem 3.3. Under Assumptions 3.1, 3.2, and 3.3 and the notations in this section, the homogeneous credibility premium of $\boldsymbol{X}_{(n+1)}$ that is obtained by minimizing (2.1) in the class of homogeneous linear functions $L_e(\boldsymbol{X})$ is given by

$$\widehat{\boldsymbol{X}}_{(n+1)}^{hom} = \boldsymbol{Z}\overline{\boldsymbol{X}} + (\boldsymbol{I}_K - \boldsymbol{Z})\widehat{\mu}^{hom}\mathbf{1}_K, \tag{3.3}$$

where $\boldsymbol{Z} = \boldsymbol{A}\left(\frac{1}{n}\boldsymbol{V} + \boldsymbol{A}\right)^{-1}$ and

$$\widehat{\mu}^{hom} = \frac{\mathbf{1}_K'(\frac{1}{n}\boldsymbol{V} + \boldsymbol{A})^{-1}\overline{\boldsymbol{X}}}{\mathbf{1}_K'(\frac{1}{n}\boldsymbol{V} + \boldsymbol{A})^{-1}\mathbf{1}_K}.$$

The associated mean square prediction error matrix is

$$\begin{aligned} &\mathrm{E}\left[\left(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}_{(n+1)}^{hom}\right)\left(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}_{(n+1)}^{hom}\right)'\right] \\ &= (\boldsymbol{I}_K - \boldsymbol{Z})\boldsymbol{A} + \boldsymbol{V} + \frac{(\boldsymbol{I}_K - \boldsymbol{Z})\mathbf{1}_K\mathbf{1}_K'(\boldsymbol{I}_K - \boldsymbol{Z})'}{\mathbf{1}_K'(\frac{1}{n}\boldsymbol{V} + \boldsymbol{A})^{-1}\mathbf{1}_K}. \end{aligned} \tag{3.4}$$

Proof. Noting that $\boldsymbol{U}'\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{X} = (\frac{1}{n}\boldsymbol{V} + \boldsymbol{A})^{-1}\overline{\boldsymbol{X}}$ and $\boldsymbol{U}'\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\mu}_X = (\frac{1}{n}\boldsymbol{V} + \boldsymbol{A})^{-1}\boldsymbol{\mu}$ from the proof of Theorem 3.2, by Lemma 3.1,

$$\widehat{\boldsymbol{X}}_{(n+1)}^{hom} = \left(\boldsymbol{\Sigma}_{X_{(n+1)}X} + \frac{\left(\boldsymbol{\mu}_{X_{(n+1)}} - \boldsymbol{\Sigma}_{X_{(n+1)}X}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\mu}_X\right)\boldsymbol{\mu}_X'}{\boldsymbol{\mu}_X'\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\mu}_X}\right)\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{X}$$

$$
\begin{aligned}
&= \boldsymbol{\Sigma}_{X_{(n+1)}X}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{X} + \frac{\left(\boldsymbol{\mu} - \boldsymbol{\Sigma}_{X_{(n+1)}X}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\mu}_X\right)\boldsymbol{\mu}'\boldsymbol{U}'}{\boldsymbol{\mu}'\boldsymbol{U}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\mu}_X}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{X} \\
&= \boldsymbol{Z}\overline{\boldsymbol{X}} + \frac{(\boldsymbol{\mu} - \boldsymbol{Z}\boldsymbol{\mu})\boldsymbol{\mu}'\boldsymbol{U}'\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{X}}{\boldsymbol{\mu}'\boldsymbol{U}'\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\mu}_X} \\
&= \boldsymbol{Z}\overline{\boldsymbol{X}} + \frac{(\boldsymbol{I}_K - \boldsymbol{Z})\boldsymbol{\mu}\boldsymbol{\mu}'(\frac{1}{n}\boldsymbol{V} + \boldsymbol{A})^{-1}\overline{\boldsymbol{X}}}{\boldsymbol{\mu}'(\frac{1}{n}\boldsymbol{V} + \boldsymbol{A})^{-1}\boldsymbol{\mu}} \\
&= \boldsymbol{Z}\overline{\boldsymbol{X}} + (\boldsymbol{I}_K - \boldsymbol{Z})\boldsymbol{1}_K\frac{\mu^2\boldsymbol{1}_K'(\frac{1}{n}\boldsymbol{V} + \boldsymbol{A})^{-1}\overline{\boldsymbol{X}}}{\mu^2\boldsymbol{1}_K'(\frac{1}{n}\boldsymbol{V} + \boldsymbol{A})^{-1}\boldsymbol{1}_K},
\end{aligned}
$$

which leads to (3.3). For the mean square prediction error (3.4), the first part of the proof of (3.2) applies, but with $\boldsymbol{Z}$ replaced by the appropriate matrix $\boldsymbol{Z}^{hom}$:

$$
\boldsymbol{Z}^{hom} = \boldsymbol{Z} + (\boldsymbol{I}_K - \boldsymbol{Z})\frac{\boldsymbol{1}_K\boldsymbol{1}_K'(\frac{1}{n}\boldsymbol{V} + \boldsymbol{A})^{-1}}{\boldsymbol{1}_K'(\frac{1}{n}\boldsymbol{V} + \boldsymbol{A})^{-1}\boldsymbol{1}_K}.
$$

That is,

$$
\begin{aligned}
&\mathrm{E}\left[\left(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}_{(n+1)}^{hom}\right)\left(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}_{(n+1)}^{hom}\right)'\right] \\
&\quad= (\boldsymbol{A} + \boldsymbol{V}) + \boldsymbol{Z}^{hom}\left(\boldsymbol{A} + \frac{1}{n}\boldsymbol{V}\right)(\boldsymbol{Z}^{hom})' - \boldsymbol{A}(\boldsymbol{Z}^{hom})' - \boldsymbol{Z}^{hom}\boldsymbol{A}.
\end{aligned}
$$

Now, noting that $\left(\boldsymbol{A} + \frac{1}{n}\boldsymbol{V}\right)^{-1}$ is symmetric, expand the terms:

$$
\begin{aligned}
&\mathrm{E}\left[\left(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}_{(n+1)}^{hom}\right)\left(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}_{(n+1)}^{hom}\right)'\right] \\
&= \boldsymbol{A} + \boldsymbol{V} \\
&\quad + \boldsymbol{Z}\left(\boldsymbol{A} + \frac{1}{n}\boldsymbol{V}\right)\boldsymbol{Z}' \\
&\quad + \left((\boldsymbol{I}_K - \boldsymbol{Z})\frac{\boldsymbol{1}_K\boldsymbol{1}_K'\boldsymbol{Z}'}{\boldsymbol{1}_K'(\frac{1}{n}\boldsymbol{V} + \boldsymbol{A})^{-1}\boldsymbol{1}_K}\right) \\
&\quad + \left(\frac{\boldsymbol{Z}\boldsymbol{1}_K\boldsymbol{1}_K'}{\boldsymbol{1}_K'(\frac{1}{n}\boldsymbol{V} + \boldsymbol{A})^{-1}\boldsymbol{1}_K}(\boldsymbol{I}_K - \boldsymbol{Z})'\right) \\
&\quad + \left((\boldsymbol{I}_K - \boldsymbol{Z})\frac{\boldsymbol{1}_K\boldsymbol{1}_K'\left(\boldsymbol{A} + \frac{1}{n}\boldsymbol{V}\right)^{-1}\boldsymbol{1}_K\boldsymbol{1}_K'}{\left(\boldsymbol{1}_K'(\frac{1}{n}\boldsymbol{V} + \boldsymbol{A})^{-1}\boldsymbol{1}_K\right)^2}(\boldsymbol{I}_K - \boldsymbol{Z})\right) \\
&\quad - \boldsymbol{A}\boldsymbol{Z}' - \left(\frac{\boldsymbol{Z}\boldsymbol{1}_K\boldsymbol{1}_K'}{\boldsymbol{1}_K'(\frac{1}{n}\boldsymbol{V} + \boldsymbol{A})^{-1}\boldsymbol{1}_K}(\boldsymbol{I}_K - \boldsymbol{Z})'\right) \\
&\quad - \boldsymbol{Z}\boldsymbol{A} - \left((\boldsymbol{I}_K - \boldsymbol{Z})\frac{\boldsymbol{1}_K\boldsymbol{1}_K'\boldsymbol{Z}'}{\boldsymbol{1}_K'(\frac{1}{n}\boldsymbol{V} + \boldsymbol{A})^{-1}\boldsymbol{1}_K}\right).
\end{aligned}
$$

Continuing on, we obtain

$$\begin{aligned}
&\mathrm{E}\left[\left(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}^{hom}_{(n+1)}\right)\left(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}^{hom}_{(n+1)}\right)'\right] \\
&= \boldsymbol{A} + \boldsymbol{V} + \boldsymbol{Z}\left(\boldsymbol{A} + \frac{1}{n}\boldsymbol{V}\right)\boldsymbol{Z}' - \boldsymbol{A}\boldsymbol{Z}' - \boldsymbol{Z}\boldsymbol{A} + \frac{(\boldsymbol{I}_K - \boldsymbol{Z})\mathbf{1}_K\mathbf{1}_K'(\boldsymbol{I}_K - \boldsymbol{Z})'}{\mathbf{1}_K'(\frac{1}{n}\boldsymbol{V} + \boldsymbol{A})^{-1}\mathbf{1}_K} \\
&= (\boldsymbol{I}_K - \boldsymbol{Z})\boldsymbol{A} + \boldsymbol{V} + \frac{(\boldsymbol{I}_K - \boldsymbol{Z})\mathbf{1}_K\mathbf{1}_K'(\boldsymbol{I}_K - \boldsymbol{Z})'}{\mathbf{1}_K'(\frac{1}{n}\boldsymbol{V} + \boldsymbol{A})^{-1}\mathbf{1}_K}.
\end{aligned}$$

□

Remark 3.4. As in Remark 3.2, when $n_i = n, i = 1, \dots, K$, simply let $\boldsymbol{V}$ be a diagonal matrix to obtain the homogeneous credibility predictor for the Bühlmann model with general dependence structure among risks in Wen and Wu (2011).

3.2 Estimation of Structural Parameters in a Spatial Context

To use the credibility estimator in Theorem 3.2, it is necessary to have the knowledge of the structural parameters of the underlying model. In this case, the structural parameters needed are

$$\begin{aligned}
\mathrm{E}[\mu(\boldsymbol{\Theta})] &= \boldsymbol{\mu} = (\mu_1, \dots, \mu_K)' \\
\mathrm{E}[\boldsymbol{\Sigma}(\boldsymbol{\Theta})] &= \boldsymbol{V} = [v_{ij}]_{i,j=1,\dots,K} \\
\mathrm{Cov}[\mu(\boldsymbol{\Theta})] &= \boldsymbol{A} = [a_{ij}]_{i,j=1,\dots,K}.
\end{aligned}$$

With only Assumptions 3.1 and 3.2, it is difficult to estimate these structural parameters. In this section, estimators of the structural parameters above are derived for a specific spatial context.

Consider observed loss data X_{iu} for $i = 1, \dots, K$ and $u = 1, \dots, n$, where i denotes the ith region, and u denotes the uth year. Following Schnapp et al. (2000), the conditional means, $\boldsymbol{\mu}(\boldsymbol{\Theta}) = (\mu(\Theta_1), \dots, \mu(\Theta_K))'$, are assumed to be realizations of a spatial process and so are the corresponding observations $\boldsymbol{X}_{(u)}|\boldsymbol{\Theta} = (X_{1u}, \dots, X_{Ku})'|\boldsymbol{\Theta}$. To be precise, we make the following assumption, in addition to Assumptions 3.1 and 3.2.

Assumption 3.4. Let $d(i, j)$ denote the distance between regions i and j. Then, assume

$$\begin{aligned}
\mathrm{Cov}[\mathrm{E}[X_{iu}|\Theta_i], \mathrm{E}[X_{jv}|\Theta_j]] = \mathrm{Cov}[\mu(\Theta_i), \mu(\Theta_j)] &= a_{ij} = f(d(i, j)), \\
\mathrm{E}[\mathrm{Cov}[X_{iu}, X_{ju}|\Theta_i, \Theta_j]] = \mathrm{E}[\sigma^2(\Theta_i, \Theta_j)] &= v_{ij} = g(d(i, j)),
\end{aligned}$$

for $i, j = 1, \ldots, K$ and $u, v = 1, \ldots, n+1$.

This means that the spatial process represented by $\mu(\Theta_i) - \mu_i = \mathrm{E}[X_{iu}|\Theta_i] - \mu_i$ is second-order stationary with isotropic spatial covariance function (or covariogram) $f(\cdot)$. For $u = 1, \ldots, n$, this is also true for the conditional spatial process of $X_{iu} - \mu(\Theta_i)|\Theta_i$, but with isotropic conditional covariance $\sigma^2(\cdot, \cdot)$ that has expected value $g(\cdot)$. It should be noted that with Assumption 3.4,

$$\begin{aligned} f(0) &= a_{ii} = \mathrm{V}[\mathrm{E}[X_{iu}|\Theta_i]] = \mathrm{V}[\mu(\Theta_i)], \\ g(0) &= v_{ii} = \mathrm{E}[\mathrm{V}[X_{iu}|\Theta_i]] = \mathrm{E}[\sigma^2(\Theta_i)], \end{aligned}$$

for $i = 1, \ldots, K$ and $u = 1, \ldots, n$.

Moreover, it is convenient to further make the following assumption. Define a set of disjoint, exhaustive intervals $\mathcal{D} = \{d_0 = [0], d_1, \ldots, d_S\}$ on the set of non-negative real numbers such that each interval contains at least two unique pairs of regions with distances between regions that fall into the interval. The value of the spatial covariance functions $f(\cdot)$ and $g(\cdot)$ is then assumed to be constant over each interval of distances in $\mathcal{D}$ (i.e., $f(\cdot)$ and $g(\cdot)$ are step functions). In effect, this assumption forces the spatial correlation of the risk parameter and the expected spatial correlation of conditional losses to be equal for regions that are within d_s distance of each other. Note that the assumption of a piecewise constant spatial correlation function is only made to allow for exact unbiasedness in estimators; this assumption is not vital because without it the estimators proposed in this section are still approximately unbiased.

In the following theorems, relevant unbiased estimators that can be used to estimate structural parameters $\boldsymbol{\mu}$, $\boldsymbol{A}$, and $\boldsymbol{V}$ are presented. For $\boldsymbol{\mu}$, we have the following estimator:

Theorem 3.4. Under Assumptions 3.1 and 3.2, the estimator

$$\widehat{\mu}_i = \frac{1}{n} \sum_{u=1}^{n} X_{iu} \tag{3.5}$$

is unbiased for μ_i for $i = 1, \ldots, n$.

Proof. The estimator $\hat{\mu}_i$ is unbiased for μ_i because

$$\begin{aligned}
\mathrm{E}[\widehat{\mu}_i] &= \mathrm{E}\left[\frac{1}{n}\sum_{u=1}^{n} X_{iu}\right] \\
&= \frac{1}{n}\sum_{u=1}^{n} \mathrm{E}[X_{iu}] \\
&= \frac{1}{n}\sum_{u=1}^{n} \mu_i \\
&= \mu_i.
\end{aligned}$$

□

Of course, the estimator $\widehat{\mu} = \frac{1}{K}\sum_{i=1}^{K}\widehat{\mu}_i$ can be used when $\mu_i = \mu_j$ for all $i, j = 1, \ldots, K$.

For $\boldsymbol{V}$, we have the following estimator.

Theorem 3.5. Under Assumptions 3.1, 3.2, and 3.4, if it is further assumed that the spatial covariance function $g(\cdot)$ is equal to the constants $g(d_0), \ldots, g(d_S)$ over the intervals $d_0, \ldots, d_S$ respectively, the estimator

$$\widehat{g}(d_s) = \frac{1}{|D_s|(n-1)} \sum_{(i,j)\in D_s} \sum_{u=1}^{n} (X_{iu} - \overline{X}_i)(X_{ju} - \overline{X}_j) \tag{3.6}$$

is unbiased for $g(d_s)$ for $s = 0, \ldots, S$, where $D_s = \{(i,j) : d(i,j) \in d_s\}$ and $|D_s|$ is the number of pairs of regions with distances that are within the interval d_s.

Proof. Since

$$\begin{aligned}
\mathrm{Cov}[X_{iu}, X_{jv}] &= \mathrm{Cov}[\mathrm{E}[X_{iu}|\boldsymbol{\Theta}], \mathrm{E}[X_{jv}|\boldsymbol{\Theta}]] + \mathrm{E}[\mathrm{Cov}[X_{iu}, X_{jv}|\boldsymbol{\Theta}]] \\
&= \mathrm{Cov}[\mu(\Theta_i), \mu(\Theta_j)] + \delta_{uv}\,\mathrm{E}[\sigma^2(\Theta_i, \Theta_j)] \\
&= a_{ij} + \delta_{uv} v_{ij},
\end{aligned}$$

we have

$$\begin{aligned}
&\mathrm{E}\left[\sum_{u=1}^{n}(X_{iu} - \overline{X}_i)(X_{ju} - \overline{X}_j)\right] \\
&\quad= \mathrm{E}\Bigg[\sum_{u=1}^{n}(X_{iu} - \mu_i)(X_{ju} - \mu_j) + n(\overline{X}_i - \mu_i)(\overline{X}_j - \mu_j) \\
&\qquad -(\overline{X}_j - \mu_j)\sum_{u=1}^{n}(X_{iu} - \mu_i) - (\overline{X}_i - \mu_i)\sum_{v=1}^{n}(X_{jv} - \mu_j)\Bigg]
\end{aligned}$$

$$\begin{aligned}
&= \mathrm{E}\left[\sum_{u=1}^{n}(X_{iu}-\mu_i)(X_{ju}-\mu_j) - \frac{1}{n}\sum_{u=1}^{n}\sum_{v=1}^{n}(X_{iu}-\mu_i)(X_{jv}-\mu_j)\right] \\
&= \sum_{u=1}^{n}\mathrm{Cov}[X_{iu},X_{ju}] - \frac{1}{n}\sum_{u=1}^{n}\sum_{v=1}^{n}\mathrm{Cov}[X_{iu},X_{jv}] \\
&= \sum_{u=1}^{n}(a_{ij}+v_{ij}) - \frac{1}{n}\sum_{u=1}^{n}\sum_{v=1}^{n}(a_{ij}+\delta_{uv}v_{ij}) \\
&= (n-1)v_{ij}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{E}[\widehat{g}(d_s)] &= \mathrm{E}\left[\frac{1}{|D_s|(n-1)}\sum_{(i,j)\in D_s}\sum_{u=1}^{n}(X_{iu}-\overline{X}_i)(X_{ju}-\overline{X}_j)\right] \\
&= \frac{1}{|D_s|(n-1)}\sum_{(i,j)\in D_s}\mathrm{E}\left[\sum_{u=1}^{n}(X_{iu}-\overline{X}_i)(X_{ju}-\overline{X}_j)\right] \\
&= \frac{1}{|D_s|(n-1)}\sum_{(i,j)\in D_s}(n-1)v_{ij} \\
&= \frac{1}{|D_s|(n-1)}\sum_{(i,j)\in D_s}(n-1)g(d_s) \\
&= g(d_s).
\end{aligned}$$

□

Estimating the structural parameter $\boldsymbol{A}$ is more complicated. The following theorem allows one to construct an unbiased estimator of $\boldsymbol{A}$ assuming $\mu_i = \mu$ for all $i = 1, \ldots, K$.

Theorem 3.6. Under the assumptions of Theorem 3.5, if it is further assumed that the spatial covariance function $f(\cdot)$ is equal to the constants $f(d_0), \ldots, f(d_S)$ over the intervals $d_0, \ldots, d_S$ respectively and that $\mu_i = \mu$ for all $i = 1, \ldots, K$, then

$$\widehat{\gamma}_f(d_s) = \frac{1}{|D_s|\left(1-\frac{1}{n}\right)}\left(\sum_{(i,j)\in D_s}\frac{1}{2}(\overline{X}_i-\overline{X}_j)^2 - \frac{1}{n^2}\sum_{(i,j)\in D_s}\sum_{u=1}^{n}\frac{1}{2}(X_{iu}-X_{ju})^2\right)$$

is unbiased for the semivariogram of the spatial process of $\mu(\Theta_i)$, $\gamma_f(d_s) = f(d_0) - f(d_s),\ \forall (i,j) \in D_s$.

Proof. First,

$$\mathrm{E}\left[\sum_{(i,j)\in D_s}\frac{1}{2}(\overline{X}_i-\overline{X}_j)^2\right]$$

$$
\begin{aligned}
&= \mathrm{E}\left[\sum_{(i,j)\in D_s} \frac{1}{2}\left((\overline{X}_i-\mu_i)^2+(\overline{X}_j-\mu_j)^2-2(\overline{X}_i-\mu_i)(\overline{X}_j-\mu_j)\right)\right] \\
&\quad + \sum_{(i,j)\in D_s} \frac{1}{2}(\mu_i-\mu_j)\,\mathrm{E}\left[\overline{X}_i-\overline{X}_j\right] + \sum_{(i,j)\in D_s} \frac{1}{2}(\mu_i-\mu_j)^2 \\
&= \frac{1}{2}\sum_{(i,j)\in D_s}\left(\mathrm{V}[\overline{X}_i]+\mathrm{V}[\overline{X}_j]-2\,\mathrm{Cov}[\overline{X}_i,\overline{X}_j]\right) + \sum_{(i,j)\in D_s} \frac{1}{2}(\mu_i-\mu_j)^2 \\
&= \frac{1}{2}\sum_{(i,j)\in D_s}\sum_{u=1}^{n}\sum_{v=1}^{n}\frac{1}{n^2}\left(\mathrm{Cov}[X_{iu},X_{iv}]+\mathrm{Cov}[X_{ju},X_{jv}]-2\,\mathrm{Cov}[X_{iu},X_{jv}]\right) \\
&\quad + \sum_{(i,j)\in D_s} \frac{1}{2}(\mu_i-\mu_j)^2 \\
&= \frac{1}{2}\sum_{(i,j)\in D_s}\sum_{u=1}^{n}\sum_{v=1}^{n}\frac{1}{n^2}\left[a_{ii}+\delta_{uv}v_{ii}+a_{jj}+\delta_{uv}v_{jj}-2(a_{ij}+\delta_{uv}v_{ij})\right] \\
&\quad + \sum_{(i,j)\in D_s} \frac{1}{2}(\mu_i-\mu_j)^2 \\
&= \frac{1}{2}\sum_{(i,j)\in D_s}\sum_{u=1}^{n}\sum_{v=1}^{n}\frac{1}{n^2}\left[f(0)+f(0)-2f(d_s)+\delta_{uv}(g(0)+g(0)-2g(d_s))\right] \\
&\quad + \sum_{(i,j)\in D_s} \frac{1}{2}(\mu_i-\mu_j)^2 \\
&= \sum_{(i,j)\in D_s}\left(f(0)-f(d_s)+\frac{1}{n}(g(0)-g(d_s))\right) + \sum_{(i,j)\in D_s} \frac{1}{2}(\mu_i-\mu_j)^2 \\
&= |D_s|\left(\gamma_f(d_s)+\frac{1}{n}\gamma_g(d_s)\right) + \sum_{(i,j)\in D_s} \frac{1}{2}(\mu_i-\mu_j)^2,
\end{aligned}
$$

where $\gamma_g(d_s) = g(d_0) - g(d_s)$ is the expectation of the conditional semivariogram of the spatial process represented by $\boldsymbol{X}_{(u)}|\boldsymbol{\Theta}$. Also,

$$
\begin{aligned}
&\mathrm{E}\left[\sum_{(i,j)\in D_s}\sum_{u=1}^{n}\frac{1}{2}(X_{iu}-X_{ju})^2\right] \\
&= \mathrm{E}\left[\sum_{(i,j)\in D_s}\sum_{u=1}^{n}\frac{1}{2}\left((X_{iu}-\mu_i)-(X_{ju}-\mu_j)+(\mu_i-\mu_j)\right)^2\right] \\
&= \mathrm{E}\left[\sum_{(i,j)\in D_s}\sum_{u=1}^{n}\frac{1}{2}\left((X_{iu}-\mu_i)^2+(X_{ju}-\mu_j)^2-(X_{iu}-\mu_i)(X_{ju}-\mu_j)\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
&+ \sum_{(i,j)\in D_s} \frac{1}{2}(\mu_i - \mu_j)\,\mathrm{E}\left[\sum_{u=1}^{n}(X_{iu}-\mu_i)-(X_{ju}-\mu_j)\right] + \sum_{(i,j)\in D_s} \frac{n}{2}(\mu_i-\mu_j)^2 \\
= \;& \frac{1}{2}\sum_{(i,j)\in D_s}\sum_{u=1}^{n}\left(\mathrm{V}[X_{iu}]+\mathrm{V}[X_{ju}]-\mathrm{Cov}[X_{iu},X_{ju}]\right) + \sum_{(i,j)\in D_s} \frac{n}{2}(\mu_i-\mu_j)^2 \\
= \;& \frac{1}{2}\sum_{(i,j)\in D_s}\sum_{u=1}^{n}\left(a_{ii}+v_{ii}+a_{jj}+v_{jj}-2(a_{ij}+v_{ij})\right) + \sum_{(i,j)\in D_s} \frac{n}{2}(\mu_i-\mu_j)^2 \\
= \;& \frac{1}{2}\sum_{(i,j)\in D_s}\sum_{u=1}^{n}\left(f(0)+f(0)-2f(d_s)+g(0)+g(0)-2g(d_s)\right) + \sum_{(i,j)\in D_s} \frac{n}{2}(\mu_i-\mu_j)^2 \\
= \;& n\left(\sum_{(i,j)\in D_s}\left(f(0)-f(d_s)+g(0)-g(d_s)\right) + \sum_{(i,j)\in D_s}\frac{1}{2}(\mu_i-\mu_j)^2\right) \\
= \;& n\left(|D_s|\left(\gamma_f(d_s)+\gamma_g(d_s)\right) + \sum_{(i,j)\in D_s}\frac{1}{2}(\mu_i-\mu_j)^2\right).
\end{aligned}
$$

Then, Theorem 3.6 follows when $\mu_i = \mu$ for $i = 1, \cdots, K$. □

To construct an unbiased estimator of $f(\cdot)$ using Theorem 3.6, it is necessary to have an estimate of the variance $f(0)$. To do so, specify that d_{s^*} is the interval (d', ∞) for some $d' \in \mathbb{R}^+$ and that $f(d) = 0$ for $d \in d_{s^*}$. This means that d' is a pre-specified maximum distance of correlation such that regions with distances above d' are not correlated with regards to the spatial process of the conditional means $\boldsymbol{\mu}(\boldsymbol{\Theta})$. This effectively allows the estimator in Theorem 3.6 to be also an unbiased estimator of the variance $f(0)$. Then, it is straightforward to construct the following unbiased estimator of $f(\cdot)$ for other distances:

$$\widehat{f}(d_s) = \widehat{\gamma}_f(d_{s^*}) - \widehat{\gamma}_f(d_s),$$

for $s = 1, \ldots, S$.

It should be emphasized that Theorem 3.6 only yields unbiased estimates when $\mu_i = \mu_j$ for all $i, j = 1, \ldots, K$. If this is not the case, the estimator $\widehat{\gamma}_f(\cdot)$ has a non-negative bias:

$$\mathrm{E}\left[\widehat{\gamma}_f(d_s)\right] = \gamma_f(d_s) + \frac{1}{|D_s|}\sum_{(i,j)\in D_s}\frac{1}{2}(\mu_i-\mu_j)^2.$$

It is then necessary to gauge whether the average squared differences between unconditional means are sufficiently small so that bias is minimal.

As an additional note, in the (biased) estimation of the covariogram $\widehat{f}(d_s) = \widehat{\gamma}_f(d_{s^*}) - \widehat{\gamma}_f(d_s)$, the bias terms of each of the estimators $\widehat{\gamma}_f(d_{s^*})$ and $\widehat{\gamma}_f(d_s)$ may cancel to produce an overall small bias for $\widehat{f}(d_s)$, except in the case $d_s = 0$. This, of course, requires the bias terms to be approximately equal for different distance intervals. However, approximate equality in the bias terms may not be true when a spatial trend is present. Under a spatial trend, when two regions are farther apart, the difference between the means of the two regions may enlarge. Since the bias of the estimator depends only on the difference between means, there would be a positive relationship between distance and bias. Therefore, the bias of $\widehat{\gamma}_f(d_s)$ may not always cancel as desired in the estimation of covariances.

Remark 3.5. It is not necessary to assume covariances v_{ij} to be defined by the spatial covariance function $g(\cdot)$ for the estimator for v_{ij} in Theorem 3.5. As seen in its proof, the estimator

$$\widehat{v}_{ij} = \frac{\sum_{u=1}^{n} (X_{iu} - \overline{X}_i)(X_{ju} - \overline{X}_j)}{n-1}$$

is unbiased for v_{ij}.

Remark 3.6. Instead of using Theorem 3.5, it is possible to estimate $\boldsymbol{V}$ in a way similar to the estimation of $\boldsymbol{A}$ with Theorem 3.6 for the case of equal means $\mu_i = \mu$ for $i = 1, \ldots, K$. The method would involve first estimating the expectation of the semivariogram of the spatial process of $\boldsymbol{X}_{(u)}|\boldsymbol{\Theta}$ and then subtracting to obtain estimates of the expectation of conditional covariances, $\boldsymbol{V}$. However, this will require making the additional assumption that $g(d) = 0$ for $d \in d_{s^*}$.

Chapter 4

Bühlmann-Straub Credibility Model with General Dependence Structure and Conditional Cross-Sectional Dependence

In this section, the Bühlmann's credibility model with general dependence structure and conditional cross-sectional dependence in Chapter 3 is generalized to allow for different weights, which is one of the generalizations that Bühlmann and Straub (1970) made to the Bühlmann's credibility model. Next, non-parametric estimation methods in the spatial context are proposed to estimate the structural parameters of the generalized model.

4.1 The Credibility Estimator

In many applications, it is more appropriate that the covariance of losses $\boldsymbol{X}_{(u)}|\boldsymbol{\Theta}$ conditional on the risk parameters $\boldsymbol{\Theta}$ in the credibility model studied in Chapter 3 depends on weights. In the following assumption, this is made precise.

Assumption 4.1. Conditional on the vector of risk parameters $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_K)'$, the cross-sectional vectors of claims in a single period, $\boldsymbol{X}_{(u)} = (X_{1u}, \ldots, X_{Ku})'$, are independent and identically distributed for $u = 1, \ldots, n+1$ with conditional moments $\mathrm{E}[\boldsymbol{X}_{(u)}|\boldsymbol{\Theta}] = \mu(\boldsymbol{\Theta}) = (\mu(\Theta_1), \ldots, \mu(\Theta_K))'$ and $\mathrm{Cov}[\boldsymbol{X}_{(u)}|\boldsymbol{\Theta}] = \boldsymbol{W}_{(u)}^{-\frac{1}{2}}\boldsymbol{\Sigma}(\boldsymbol{\Theta})\boldsymbol{W}_{(u)}^{-\frac{1}{2}} = \left[\frac{\sigma^2(\Theta_i,\Theta_j)}{\sqrt{w_{iu}w_{ju}}}\right]_{i,j=1,\ldots,K}$, where for $u = 1, \dots, n$,

$$\boldsymbol{W}_{(u)} = \begin{bmatrix} w_{1u} & & 0 \\ & \ddots & \\ 0 & & w_{Ku} \end{bmatrix},$$

is a matrix with positive exposures or weights for all K entities in the uth period on its diagonal.

The assumptions used in this chapter are Assumptions 4.1 and 3.2, where the former specifies the conditional moments and the latter specifies the dependence structure of the risk parameters $\boldsymbol{\Theta}$.

For the results below, it is necessary to introduce additional notation. For $i = 1, \dots, K$, let $\boldsymbol{W}_i$ be a matrix with exposures in all n periods for the ith period on its diagonal:

$$\boldsymbol{W}_i = \begin{bmatrix} w_{i1} & & 0 \\ & \ddots & \\ 0 & & w_{in} \end{bmatrix},$$

and $\boldsymbol{W}_X$ be the matrix with all exposures for all risks and periods:

$$\boldsymbol{W}_X = \begin{bmatrix} \boldsymbol{W}_1 & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{W}_K \end{bmatrix}.$$

Then, we have the following lemma.

Lemma 4.1. Under Assumptions 4.1 and 3.2 and the notations in this section,

1. The expectations of $\boldsymbol{X}$ and $\boldsymbol{X}_{(n+1)}$ are

$$\boldsymbol{\mu}_X = E[\boldsymbol{X}] = \boldsymbol{\mu} \otimes \boldsymbol{1}_n \quad \text{and} \quad \boldsymbol{\mu}_{X_{(n+1)}} = E[\boldsymbol{X}_{(n+1)}] = \boldsymbol{\mu}.$$

2. The covariance matrix of $\boldsymbol{X}$ is

$$\boldsymbol{\Sigma}_{XX} = \mathrm{Cov}[\boldsymbol{X}] = \boldsymbol{U}\boldsymbol{A}\boldsymbol{U}' + \boldsymbol{W}_X^{-\frac{1}{2}}(\boldsymbol{V} \otimes \boldsymbol{I}_n)\boldsymbol{W}_X^{-\frac{1}{2}}.$$

3. The covariance matrix of $\boldsymbol{X}_{(n+1)}$ and $\boldsymbol{X}$ is

$$\boldsymbol{\Sigma}_{X_{(n+1)}X} = \mathrm{Cov}[\boldsymbol{X}_{(n+1)}, \boldsymbol{X}] = \boldsymbol{A}\boldsymbol{U}'.$$

4. The inverse of the covariance matrix $\boldsymbol{\Sigma}_{XX}$ is

$$\begin{aligned}\boldsymbol{\Sigma}_{XX}^{-1} &= \boldsymbol{W}_X^{\frac{1}{2}}(\boldsymbol{V}^{-1}\otimes\boldsymbol{I}_n)\boldsymbol{W}_X^{\frac{1}{2}} \\ &\quad -\boldsymbol{W}_X^{\frac{1}{2}}(\boldsymbol{V}^{-1}\otimes\boldsymbol{1}_n)\boldsymbol{W}_X^{\frac{1}{2}}(\boldsymbol{A}^{-1}+\boldsymbol{S})^{-1}\boldsymbol{W}_X^{\frac{1}{2}}(\boldsymbol{V}^{-1}\otimes\boldsymbol{1}_n')\boldsymbol{W}_X^{\frac{1}{2}},\end{aligned}$$

where

$$\boldsymbol{S}=\begin{bmatrix} v_{11}^{(-1)}\sum_u w_{1u} & v_{12}^{(-1)}\sum_u\sqrt{w_{1u}w_{2u}} & \dots & v_{1K}^{(-1)}\sum_u\sqrt{w_{1u}w_{Ku}} \\ v_{21}^{(-1)}\sum_u\sqrt{w_{2u}w_{1u}} & v_{22}^{(-1)}\sum_u w_{2u} & \dots & v_{2K}^{(-1)}\sum_u\sqrt{w_{2u}w_{Ku}} \\ \vdots & \vdots & \vdots & \vdots \\ v_{K1}^{(-1)}\sum_u\sqrt{w_{Ku}w_{1u}} & v_{K2}^{(-1)}\sum_u\sqrt{w_{Ku}w_{2u}} & \dots & v_{KK}^{(-1)}\sum_u w_{Ku}\end{bmatrix},$$

and $v_{iu}^{(-1)}$ is the (i,u)th element of the matrix $\boldsymbol{V}^{-1}$. The summation over u is through $u=1,\dots,n$.

Proof. 1. Same as in Lemma 3.1.

2. Since $\text{Cov}[\text{E}[\boldsymbol{X}_i|\boldsymbol{\Theta}],\text{E}[\boldsymbol{X}_i|\boldsymbol{\Theta}]]=a_{ij}\boldsymbol{1}_n\boldsymbol{1}_n'$ and $\text{E}[\text{Cov}[\boldsymbol{X}_i,\boldsymbol{X}_j|\boldsymbol{\Theta}]]=W_i^{-\frac{1}{2}}v_{ij}\boldsymbol{I}_nW_j^{-\frac{1}{2}}$ for $i,j=1,\dots,K$, we have

$$\begin{aligned}\boldsymbol{\Sigma}_{XX} &= \text{Cov}[\boldsymbol{X}] \\ &= \text{Cov}[\text{E}[\boldsymbol{X}|\boldsymbol{\Theta}]]+\text{E}[\text{Cov}[\boldsymbol{X}|\boldsymbol{\Theta}]] \\ &= \left(\boldsymbol{A}\otimes\boldsymbol{1}_n\boldsymbol{1}_n'\right)+\begin{bmatrix}\boldsymbol{W}_1^{-\frac{1}{2}}v_{11}\boldsymbol{W}_1^{-\frac{1}{2}} & \dots & \boldsymbol{W}_1^{-\frac{1}{2}}v_{1K}\boldsymbol{W}_K^{-\frac{1}{2}} \\ \vdots & \vdots & \vdots \\ \boldsymbol{W}_1^{-\frac{1}{2}}v_{1K}\boldsymbol{W}_K^{-\frac{1}{2}} & \dots & \boldsymbol{W}_K^{-\frac{1}{2}}v_{KK}\boldsymbol{W}_K^{-\frac{1}{2}}\end{bmatrix} \\ &= \boldsymbol{U}\boldsymbol{A}\boldsymbol{U}'+\boldsymbol{W}_X^{-\frac{1}{2}}(\boldsymbol{V}\otimes\boldsymbol{I}_n)\boldsymbol{W}_X^{-\frac{1}{2}}.\end{aligned}$$

3. Same as in Lemma 3.1.

4. By setting $\boldsymbol{E}=\boldsymbol{W}_X^{-\frac{1}{2}}(\boldsymbol{V}\otimes\boldsymbol{I}_n)\boldsymbol{W}_X^{-\frac{1}{2}},\boldsymbol{F}=\boldsymbol{U},\boldsymbol{G}=\boldsymbol{A}$, and $\boldsymbol{H}=\boldsymbol{U}'$ in the relation

$$(\boldsymbol{E}+\boldsymbol{F}\boldsymbol{G}\boldsymbol{H})^{-1}=\boldsymbol{E}^{-1}-\boldsymbol{E}^{-1}\boldsymbol{F}(\boldsymbol{G}^{-1}+\boldsymbol{H}\boldsymbol{E}^{-1}\boldsymbol{F})^{-1}\boldsymbol{H}\boldsymbol{E}^{-1},$$

$$\begin{aligned}
\boldsymbol{\Sigma}_{XX}^{-1} &= \left(\boldsymbol{U}\boldsymbol{A}\boldsymbol{U}'+\boldsymbol{W}_X^{-\frac{1}{2}}(\boldsymbol{V}\otimes\boldsymbol{I}_n)\boldsymbol{W}_X^{-\frac{1}{2}}\right)^{-1}\\
&= \boldsymbol{W}_X^{\frac{1}{2}}(\boldsymbol{V}^{-1}\otimes\boldsymbol{I}_n)\boldsymbol{W}_X^{\frac{1}{2}}\\
&\quad -\boldsymbol{W}_X^{\frac{1}{2}}(\boldsymbol{V}^{-1}\otimes\boldsymbol{I}_n)\boldsymbol{W}_X^{\frac{1}{2}}\boldsymbol{U}\left(\boldsymbol{A}^{-1}+\boldsymbol{U}'\left(\boldsymbol{W}_X^{-\frac{1}{2}}(\boldsymbol{V}\otimes\boldsymbol{I}_n)\boldsymbol{W}_X^{-\frac{1}{2}}\right)^{-1}\boldsymbol{U}\right)^{-1}\\
&\quad \times\boldsymbol{U}'\boldsymbol{W}_X^{\frac{1}{2}}(\boldsymbol{V}^{-1}\otimes\boldsymbol{I}_n')\boldsymbol{W}_X^{\frac{1}{2}}\\
&= \boldsymbol{W}_X^{\frac{1}{2}}(\boldsymbol{V}^{-1}\otimes\boldsymbol{I}_n)\boldsymbol{W}_X^{\frac{1}{2}}\\
&\quad -\boldsymbol{W}_X^{\frac{1}{2}}(\boldsymbol{V}^{-1}\otimes\boldsymbol{1}_n)\boldsymbol{W}_X^{\frac{1}{2}}(\boldsymbol{A}^{-1}+\boldsymbol{S})^{-1}\boldsymbol{W}_X^{\frac{1}{2}}(\boldsymbol{V}^{-1}\otimes\boldsymbol{1}_n')\boldsymbol{W}_X^{\frac{1}{2}},
\end{aligned}$$

since

$$\begin{aligned}
&\boldsymbol{U}'\left(\boldsymbol{W}_X^{-\frac{1}{2}}(\boldsymbol{V}\otimes\boldsymbol{I}_n)\boldsymbol{W}_X^{-\frac{1}{2}}\right)^{-1}\boldsymbol{U}\\
&= \boldsymbol{U}'\boldsymbol{W}_X^{\frac{1}{2}}(\boldsymbol{V}^{-1}\otimes\boldsymbol{I}_n)\boldsymbol{W}_X^{\frac{1}{2}}\boldsymbol{U}\\
&= \begin{bmatrix}
\boldsymbol{1}_n'\boldsymbol{W}_1v_{11}^{(-1)}\boldsymbol{I}_n\boldsymbol{W}_1\boldsymbol{1}_n & \boldsymbol{1}_n'\boldsymbol{W}_1v_{12}^{(-1)}\boldsymbol{I}_n\boldsymbol{W}_2\boldsymbol{1}_n & \ldots & \boldsymbol{1}_n'\boldsymbol{W}_1v_{1K}^{(-1)}\boldsymbol{I}_n\boldsymbol{W}_K\boldsymbol{1}_n\\
\boldsymbol{1}_n'\boldsymbol{W}_2v_{21}^{(-1)}\boldsymbol{I}_n\boldsymbol{W}_1\boldsymbol{1}_n & \boldsymbol{1}_n'\boldsymbol{W}_2v_{22}^{(-1)}\boldsymbol{I}_n\boldsymbol{W}_2\boldsymbol{1}_n & \ldots & \boldsymbol{1}_n'\boldsymbol{W}_2v_{2K}^{(-1)}\boldsymbol{I}_n\boldsymbol{W}_K\boldsymbol{1}_n\\
\vdots & \vdots & \vdots & \vdots\\
\boldsymbol{1}_n'\boldsymbol{W}_Kv_{K1}^{(-1)}\boldsymbol{I}_n\boldsymbol{W}_1\boldsymbol{1}_n & \boldsymbol{1}_n'\boldsymbol{W}_Kv_{K2}^{(-1)}\boldsymbol{I}_n\boldsymbol{W}_2\boldsymbol{1}_n & \ldots & \boldsymbol{1}_n'\boldsymbol{W}_Kv_{KK}^{(-1)}\boldsymbol{I}_n\boldsymbol{W}_K\boldsymbol{1}_n
\end{bmatrix}\\
&= \boldsymbol{S}
\end{aligned}$$

□

Using Lemma 4.1, the following theorem can be obtained.

Theorem 4.2. Under Assumptions 4.1 and 3.2 and notations in this section, the inhomogeneous credibility premium of $\boldsymbol{X}_{(n+1)}$, denoted by $\widehat{\boldsymbol{X}}_{(n+1)}^{(bs)}$, that is obtained by minimizing (2.1) in the class of inhomogeneous linear functions $L(\boldsymbol{X},1)$ is given by

$$\widehat{\boldsymbol{X}}_{(n+1)}^{(bs)}=\sum_{i=1}^{K}\left[\boldsymbol{Z}_i\overline{\boldsymbol{X}}_i^{(bs)}+\frac{1}{K}\boldsymbol{I}_K\boldsymbol{\mu}-\boldsymbol{Z}_i\mu_i\boldsymbol{1}_K\right], \tag{4.1}$$

where

$$\begin{aligned}
\boldsymbol{Z}_i &= \left(\boldsymbol{A}^{-1}+\boldsymbol{S}\right)^{-1}\boldsymbol{S}_i\\
\boldsymbol{S}_i &= \begin{bmatrix} v_{1i}^{(-1)}\sum_u\sqrt{w_{1u}w_{iu}} & & 0\\ & \ddots & \\ 0 & & v_{Ki}^{(-1)}\sum_u\sqrt{w_{Ku}w_{iu}}\end{bmatrix}\\
\overline{\boldsymbol{X}}_i^{(bs)} &= \begin{bmatrix} \frac{\sum_u\sqrt{w_{1u}w_{iu}}X_{iu}}{\sum_u\sqrt{w_{1u}w_{iu}}}\\ \vdots \\ \frac{\sum_u\sqrt{w_{Ku}w_{iu}}X_{iu}}{\sum_u\sqrt{w_{Ku}w_{iu}}}\end{bmatrix}.
\end{aligned}$$

The associated mean square prediction error matrix is

$$\begin{aligned}
&\mathrm{E}\left[\left(\boldsymbol{X}_{(n+1)}-\widehat{\boldsymbol{X}}_{(n+1)}^{(bs)}\right)\left(\boldsymbol{X}_{(n+1)}-\widehat{\boldsymbol{X}}_{(n+1)}^{(bs)}\right)'\right]\\
&= \left(\boldsymbol{A}+\boldsymbol{W}_{(n+1)}^{-\frac{1}{2}}\boldsymbol{V}\boldsymbol{W}_{(n+1)}^{-\frac{1}{2}}\right)+\left(\sum_{i=1}^{K}\sum_{j=1}^{K}\boldsymbol{Z}_i\left(a_{ij}\boldsymbol{1}_K\boldsymbol{1}_K'+v_{ij}\boldsymbol{W}_{ij}^{*}\right)\boldsymbol{Z}_j'\right)\\
&\quad -\left(\sum_{j=1}^{K}\boldsymbol{A}_{(j)}\boldsymbol{1}_K\boldsymbol{1}_K'\boldsymbol{Z}_j'\right)-\left(\sum_{j=1}^{K}\boldsymbol{Z}_j\boldsymbol{1}_K\boldsymbol{1}_K'\boldsymbol{A}_{(j)}'\right),
\end{aligned}\tag{4.2}$$

where

$$\boldsymbol{A}_{(j)}=\begin{bmatrix} a_{1j} & & 0\\ & \ddots & \\ 0 & & a_{Kj}\end{bmatrix} \quad\text{and}\quad \boldsymbol{W}_{ij}^{*}=\left[\frac{\sum_u\sqrt{w_{xu}w_{yv}}}{\left(\sum_u\sqrt{w_{xu}w_{iu}}\right)\left(\sum_v\sqrt{w_{yv}w_{jv}}\right)}\right]_{x,y=1,\ldots,K}.$$

Proof. By Lemma 4.1,

$$\begin{aligned}
&\boldsymbol{\Sigma}_{X_{(n+1)}X}\boldsymbol{\Sigma}_{XX}^{-1}(\boldsymbol{X}-\boldsymbol{\mu}_X)\\
&= \boldsymbol{A}\boldsymbol{U}'\Big[\boldsymbol{W}_X^{\frac{1}{2}}(\boldsymbol{V}^{-1}\otimes\boldsymbol{I}_n)\boldsymbol{W}_X^{\frac{1}{2}}\\
&\quad -\boldsymbol{W}_X^{\frac{1}{2}}(\boldsymbol{V}^{-1}\otimes\boldsymbol{1}_n)\boldsymbol{W}_X^{\frac{1}{2}}(\boldsymbol{A}^{-1}+\boldsymbol{S})^{-1}\boldsymbol{W}_X^{\frac{1}{2}}(\boldsymbol{V}^{-1}\otimes\boldsymbol{1}_n')\boldsymbol{W}_X^{\frac{1}{2}}\Big](\boldsymbol{X}-\boldsymbol{\mu}_X)\\
&= \left[\boldsymbol{A}-\boldsymbol{U}'\boldsymbol{W}_X^{\frac{1}{2}}(\boldsymbol{V}^{-1}\otimes\boldsymbol{1}_n)\boldsymbol{W}_X^{\frac{1}{2}}(\boldsymbol{A}^{-1}+\boldsymbol{S})^{-1}\right]\boldsymbol{W}_X^{\frac{1}{2}}(\boldsymbol{V}^{-1}\otimes\boldsymbol{1}_n')\boldsymbol{W}_X^{\frac{1}{2}}(\boldsymbol{X}-\boldsymbol{\mu}_X)\\
&= \left[\boldsymbol{A}-\boldsymbol{S}(\boldsymbol{A}^{-1}+\boldsymbol{S})^{-1}\right]\boldsymbol{W}_X^{\frac{1}{2}}(\boldsymbol{V}^{-1}\otimes\boldsymbol{1}_n')\boldsymbol{W}_X^{\frac{1}{2}}(\boldsymbol{X}-\boldsymbol{\mu}_X).
\end{aligned}$$

Using the relation $(\boldsymbol{E}+\boldsymbol{F})^{-1} = \boldsymbol{E}^{-1} - \boldsymbol{E}^{-1}\boldsymbol{F}(\boldsymbol{E}+\boldsymbol{F})^{-1}$,

$$
\begin{aligned}
&\boldsymbol{\Sigma}_{X_{(n+1)}X}\boldsymbol{\Sigma}_{XX}^{-1}(\boldsymbol{X}-\boldsymbol{\mu}_X) \\
&= \left[\boldsymbol{A} - \boldsymbol{A}S(\boldsymbol{A}^{-1}+\boldsymbol{S})^{-1}\right]\boldsymbol{W}_X^{\frac{1}{2}}(\boldsymbol{V}^{-1}\otimes \boldsymbol{1}_n')\boldsymbol{W}_X^{\frac{1}{2}}(\boldsymbol{X}-\boldsymbol{\mu}_X) \\
&= (\boldsymbol{A}^{-1}+\boldsymbol{S})^{-1}\begin{bmatrix} v_{11}^{(-1)}\sum_u \sqrt{w_{1u}w_{1u}}(X_{1u}-\mu_1)+\cdots+v_{1K}^{(-1)}\sum_u\sqrt{w_{1u}w_{Ku}}(X_{Ku}-\mu_K) \\ \vdots \\ v_{K1}^{(-1)}\sum_u \sqrt{w_{Ku}w_{1u}}(X_{1u}-\mu_1)+\cdots+v_{KK}^{(-1)}\sum_u\sqrt{w_{Ku}w_{Ku}}(X_{Ku}-\mu_K)\end{bmatrix} \\
&= (\boldsymbol{A}^{-1}+\boldsymbol{S})^{-1}\left\{\sum_{i=1}^{K}\boldsymbol{S}_i\left(\overline{\boldsymbol{X}}_i^{(bs)}-\mu_i\boldsymbol{1}_K\right)\right\} \\
&= \sum_{i=1}^{K}\boldsymbol{Z}_i\left(\overline{\boldsymbol{X}}_i^{(bs)}-\mu_i\boldsymbol{1}_K\right).
\end{aligned}
$$

Then, using Lemma 2.1, we have

$$
\begin{aligned}
\widehat{\boldsymbol{X}}_{(n+1)}^{(bs)} &= \boldsymbol{\mu}_Y + \boldsymbol{\Sigma}_{X_{(n+1)}X}\boldsymbol{\Sigma}_{XX}(\boldsymbol{X}-\boldsymbol{\mu}_X) \\
&= \boldsymbol{\mu} + \sum_{i=1}^{K}\boldsymbol{Z}_i\left(\overline{\boldsymbol{X}}_i^{(bs)}-\mu_i\boldsymbol{1}_K\right) \\
&= \sum_{i=1}^{K}\frac{1}{K}\boldsymbol{I}_K\boldsymbol{\mu} + \sum_{i=1}^{K}\boldsymbol{Z}_i\left(\overline{\boldsymbol{X}}_i^{(bs)}-\mu_i\boldsymbol{1}_K\right) \\
&= \sum_{i=1}^{K}\left[\boldsymbol{Z}_i\overline{\boldsymbol{X}}_i^{(bs)} + \frac{1}{K}\boldsymbol{I}_K\boldsymbol{\mu} - \boldsymbol{Z}_i\mu_i\boldsymbol{1}_K\right].
\end{aligned}
$$

For the mean square prediction error, since

$$
\begin{aligned}
\mathrm{Cov}[X_{i,n+1},\overline{\boldsymbol{X}}_j^{(bs)}] &= \begin{bmatrix} \sum_{v=1}^{n}\sqrt{w_{1v}w_{jv}}\frac{\mathrm{Cov}[X_{i,n+1},X_{jv}]}{\sum_v\sqrt{w_{1v}w_{jv}}} \\ \vdots \\ \sum_{v=1}^{n}\frac{\sqrt{w_{Kv}w_{jv}}\,\mathrm{Cov}[X_{i,n+1},X_{jv}]}{\sum_v\sqrt{w_{Kv}w_{jv}}}\end{bmatrix}' \\
&= a_{ij}\boldsymbol{1}_K'
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{Cov}[\overline{\boldsymbol{X}}_i^{bs}, \overline{\boldsymbol{X}}_j^{bs}] &= \left[\sum_{u=1}^{n}\sum_{v=1}^{n} \frac{\sqrt{w_{xu}w_{iu}w_{yv}w_{jv}}\,\mathrm{Cov}[X_{iu}, X_{jv}]}{\sum_u \sum_v \sqrt{w_{xu}w_{iu}w_{yv}w_{jv}}}\right]_{x,y=1,\ldots,K} \\
&= \left[\sum_{u=1}^{n}\sum_{v=1}^{n} \frac{\sqrt{w_{xu}w_{iu}w_{yv}w_{jv}}\left(a_{ij} + \delta uv \frac{v_{ij}}{\sqrt{w_{iu}w_{jv}}}\right)}{\sum_u \sum_v \sqrt{w_{xu}w_{iu}w_{yv}w_{jv}}}\right]_{x,y=1,\ldots,K} \\
&= \left[a_{ij} + v_{ij}\frac{\sum_u \sqrt{w_{xu}w_{yv}}}{(\sum_u \sqrt{w_{xu}w_{iu}})(\sum_v \sqrt{w_{yv}w_{jv}})}\right]_{x,y=1,\ldots,K} \\
&= a_{ij}\mathbf{1}_K\mathbf{1}_K' + v_{ij}\boldsymbol{W}_{ij}^*,
\end{aligned}
$$

for $i, j = 1, \ldots, K$, we have

$$
\begin{aligned}
&\mathrm{E}\left[\left(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}_{(n+1)}^{(bs)}\right)\left(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}_{(n+1)}^{(bs)}\right)'\right] \\
&= \mathrm{Cov}\left[\boldsymbol{X}_{(n+1)}\right] + \mathrm{Cov}\left[\widehat{\boldsymbol{X}}_{(n+1)}^{(bs)}\right] - \mathrm{Cov}\left[\boldsymbol{X}_{(n+1)}, \widehat{\boldsymbol{X}}_{(n+1)}^{(bs)}\right] - \mathrm{Cov}\left[\widehat{\boldsymbol{X}}_{(n+1)}^{(bs)}, \boldsymbol{X}_{(n+1)}\right] \\
&\quad + \mathrm{E}\left[\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}_{(n+1)}^{(bs)}\right]\mathrm{E}\left[\left(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}_{(n+1)}^{(bs)}\right)'\right] \\
&= \mathrm{Cov}\left[\boldsymbol{X}_{(n+1)}\right] + \mathrm{Cov}\left[\sum_{i=1}^{K}\boldsymbol{Z}_i\overline{\boldsymbol{X}}_i^{(bs)}\right] \\
&\quad - \mathrm{Cov}\left[\boldsymbol{X}_{(n+1)}, \sum_{i=1}^{K}\boldsymbol{Z}_i\overline{\boldsymbol{X}}_i^{(bs)}\right] - \mathrm{Cov}\left[\sum_{i=1}^{K}\boldsymbol{Z}_i\overline{\boldsymbol{X}}_i^{(bs)}, \boldsymbol{X}_{(n+1)}\right] + 0 \\
&= \left(\boldsymbol{A} + \boldsymbol{W}_{(n+1)}^{-\frac{1}{2}}\boldsymbol{V}\boldsymbol{W}_{(n+1)}^{-\frac{1}{2}}\right) + \left(\sum_{i=1}^{K}\sum_{j=1}^{K}\boldsymbol{Z}_i\left(a_{ij}\mathbf{1}_K\mathbf{1}_K' + v_{ij}\boldsymbol{W}_{ij}^*\right)\boldsymbol{Z}_j'\right) \\
&\quad - \sum_{j=1}^{K}\boldsymbol{A}_{(j)}\mathbf{1}_K\mathbf{1}_K'\boldsymbol{Z}_j' - \sum_{j=1}^{K}\boldsymbol{Z}_j\mathbf{1}_K\mathbf{1}_K'\boldsymbol{A}_{(j)}'.
\end{aligned}
$$

□

Remark 4.1. When $\mu_i = \mu$ for all $i = 1, \ldots, K$, the credibility estimator $\widehat{\boldsymbol{X}}_{(n+1)}^{(bs)}$ can be written as

$$
\widehat{\boldsymbol{X}}_{(n+1)}^{(bs)} = \sum_{i=1}^{K}\left[\boldsymbol{Z}_i\overline{\boldsymbol{X}}_i^{(bs)} + \left(\frac{1}{K}\boldsymbol{I}_K - \boldsymbol{Z}_i\right)\boldsymbol{\mu}\right]. \tag{4.3}
$$

This way, the term $\boldsymbol{Z}_i\overline{\boldsymbol{X}}_i^{(bs)}$ and the term $(\frac{1}{K}\boldsymbol{I}_K - \boldsymbol{Z}_i)\boldsymbol{\mu}$ can be more easily interpreted as the contribution of the past losses of entity i and the expected loss of entity i, respectively, to the credibility estimator $\widehat{\boldsymbol{X}}_{(n+1)}^{(bs)}$.

In the following theorem, the homogeneous credibility premium is given.

Theorem 4.3. Under Assumptions 4.1, 3.2, and 3.3 and the notations in this section, the homogeneous credibility premium of $\boldsymbol{X}_{(n+1)}$, denoted by $\widehat{\boldsymbol{X}}_{(n+1)}^{(bs)hom}$, that is obtained by minimizing (2.1) in the class of homogeneous linear functions $L_e(\boldsymbol{X})$ is given by

$$\widehat{\boldsymbol{X}}_{(n+1)}^{(bs)hom} = \sum_{i=1}^{K} \left[\boldsymbol{Z}_i \overline{\boldsymbol{X}}_i^{(bs)} + \left(\frac{1}{K} \boldsymbol{I}_K - \boldsymbol{Z}_i \right) \widehat{\mu}^{(bs)hom} \boldsymbol{1}_K \right], \tag{4.4}$$

where $\boldsymbol{Z}_i = \left(\boldsymbol{A}^{-1} + \boldsymbol{S} \right)^{-1} \boldsymbol{S}_i$ and

$$\widehat{\mu}^{(bs)hom} = \frac{\boldsymbol{1}_K' \boldsymbol{A}^{-1} \sum_i \boldsymbol{Z}_i \overline{\boldsymbol{X}}_i^{(bs)}}{\boldsymbol{1}_K' \boldsymbol{A}^{-1} \sum_i \boldsymbol{Z}_i \boldsymbol{1}_K}.$$

The associated mean square prediction error matrix is

$$\begin{aligned}
&\mathrm{E}\left[\left(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}_{(n+1)}^{(bs)hom} \right) \left(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}_{(n+1)}^{(bs)hom} \right)' \right] \\
&= \left(\boldsymbol{A} + \boldsymbol{W}_{(n+1)}^{-\frac{1}{2}} \boldsymbol{V} \boldsymbol{W}_{(n+1)}^{-\frac{1}{2}} \right) + \left(\sum_{i=1}^{K} \sum_{j=1}^{K} \boldsymbol{Z}_i^{hom} \left(a_{ij} \boldsymbol{1}_K \boldsymbol{1}_K' + v_{ij} \boldsymbol{W}_{ij}^{*} \right) \left(\boldsymbol{Z}_j^{hom} \right)' \right) \\
&\quad - \left(\sum_{j=1}^{K} \boldsymbol{A}_{(j)} \boldsymbol{1}_K \boldsymbol{1}_K' \left(\boldsymbol{Z}_j^{hom} \right)' \right) - \left(\sum_{j=1}^{K} \boldsymbol{Z}_j^{hom} \boldsymbol{1}_K \boldsymbol{1}_K' \boldsymbol{A}_{(j)}' \right),
\end{aligned} \tag{4.5}$$

where

$$\boldsymbol{Z}_i^{hom} = \left[\boldsymbol{I}_K + \left(\sum_{j=1}^{K} \left(\frac{1}{K} \boldsymbol{I}_K - \boldsymbol{Z}_j \right) \boldsymbol{1}_K \right) \left(\frac{\boldsymbol{1}_K' \boldsymbol{A}^{-1}}{\boldsymbol{1}_K' \boldsymbol{A}^{-1} \sum_j \boldsymbol{Z}_j \boldsymbol{1}_K} \right) \right] \boldsymbol{Z}_i.$$

Proof. Noting that

$$\begin{aligned}
\boldsymbol{\Sigma}_{X_{(n+1)}X} \boldsymbol{\Sigma} \boldsymbol{X} &= \sum_i \boldsymbol{Z}_i \overline{\boldsymbol{X}}_i^{(bs)} \\
\boldsymbol{\Sigma}_{X_{(n+1)}X} \boldsymbol{\Sigma} \mu_X &= \sum_i \boldsymbol{Z}_i \mu_i \boldsymbol{1}_K \\
\boldsymbol{U}' \boldsymbol{\Sigma}_{XX}^{-1} \boldsymbol{X} &= \boldsymbol{A}^{-1} \sum_i \boldsymbol{Z}_i \overline{\boldsymbol{X}}_i^{(bs)} \\
\boldsymbol{U}' \boldsymbol{\Sigma}_{XX}^{-1} \boldsymbol{\mu}_X &= \boldsymbol{A}^{-1} \sum_i \boldsymbol{Z}_i \mu_i \boldsymbol{1}_K
\end{aligned}$$

from the proof of Theorem 4.2 and that $\mu_i = \mu$ for all $i = 1, \ldots, K$ from Assumption 3.3,

$$\widehat{\boldsymbol{X}}_{(n+1)}^{(bs)hom} = \left(\boldsymbol{\Sigma}_{X_{(n+1)}X} + \frac{\left(\boldsymbol{\mu}_{X_{(n+1)}} - \boldsymbol{\Sigma}_{X_{(n+1)}X} \boldsymbol{\Sigma}_{XX}^{-1} \boldsymbol{\mu}_X \right) \boldsymbol{\mu}_X'}{\boldsymbol{\mu}_X' \boldsymbol{\Sigma}_{XX}^{-1} \boldsymbol{\mu}_X} \right) \boldsymbol{\Sigma}_{XX}^{-1} \boldsymbol{X}$$

$$
\begin{aligned}
&= \sum_{i=1}^{K} \boldsymbol{Z}_i \overline{\boldsymbol{X}}_i^{(bs)} + \frac{(\boldsymbol{\mu} - \sum_i \boldsymbol{Z}_i \mu_i \mathbf{1}_K)\boldsymbol{\mu}'\boldsymbol{U}'\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{X}}{\boldsymbol{\mu}'\boldsymbol{U}'\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\mu}_X} \\
&= \sum_{i=1}^{K} \boldsymbol{Z}_i \overline{\boldsymbol{X}}_i^{(bs)} + \frac{\mu^2(\boldsymbol{I}_K - \sum_i \boldsymbol{Z}_i)\mathbf{1}_K \mathbf{1}_K' \boldsymbol{A}^{-1} \sum_i \boldsymbol{Z}_i \overline{\boldsymbol{X}}_i^{(bs)}}{\mu^2 \mathbf{1}_K' \boldsymbol{A}^{-1} \sum_i \boldsymbol{Z}_i \mathbf{1}_K} \\
&= \sum_{i=1}^{K} \left[\boldsymbol{Z}_i \overline{\boldsymbol{X}}_i^{(bs)} + \left(\frac{1}{K}\boldsymbol{I}_K - \boldsymbol{Z}_i \right) \mathbf{1}_K \frac{\mathbf{1}_K' \boldsymbol{A}^{-1} \sum_j \boldsymbol{Z}_j \overline{\boldsymbol{X}}_j^{(bs)}}{\mathbf{1}_K' \boldsymbol{A}^{-1} \sum_j \boldsymbol{Z}_j \mathbf{1}_K} \right],
\end{aligned}
$$

which leads to the homogeneous credibility premium in (4.4). For the mean square prediction error, note that

$$
\begin{aligned}
\widehat{\boldsymbol{X}}_{(n+1)}^{(bs)hom} &= \sum_{i=1}^{K} \boldsymbol{Z}_i \overline{\boldsymbol{X}}_i^{(bs)} + \sum_{i=1}^{K} \left(\sum_{j=1}^{K} \left(\frac{1}{K}\boldsymbol{I}_K - \boldsymbol{Z}_j \right) \mathbf{1}_K \right) \left(\frac{\mathbf{1}_K' \boldsymbol{A}^{-1}}{\mathbf{1}_K' \boldsymbol{A}^{-1} \sum_j \boldsymbol{Z}_j \mathbf{1}_K} \right) \boldsymbol{Z}_i \overline{\boldsymbol{X}}_i^{(bs)} \\
&= \sum_{i=1}^{K} \left[\boldsymbol{I}_K + \left(\sum_{j=1}^{K} \left(\frac{1}{K}\boldsymbol{I}_K - \boldsymbol{Z}_j \right) \mathbf{1}_K \right) \left(\frac{\mathbf{1}_K' \boldsymbol{A}^{-1}}{\mathbf{1}_K' \boldsymbol{A}^{-1} \sum_j \boldsymbol{Z}_j \mathbf{1}_K} \right) \right] \boldsymbol{Z}_i \overline{\boldsymbol{X}}_i^{(bs)} \\
&= \sum_{i=1}^{K} \boldsymbol{Z}_i^{hom} \overline{\boldsymbol{X}}_i^{(bs)}.
\end{aligned}
$$

Hence, the proof of Theorem 4.2 applies, but with $\boldsymbol{Z}_i$ replaced by the matrix $\boldsymbol{Z}_i^{hom}$. □

Remark 4.2. If we set $\boldsymbol{V} = diag(v_{11}, \ldots, v_{KK})$, then we have

$$
\begin{aligned}
\boldsymbol{S} &= \begin{bmatrix} v_{11}^{-1} \sum_u w_{1u} & & 0 & \\ & & \ddots & \\ & 0 & \ldots & v_{KK}^{-1} \sum_u w_{Ku} \end{bmatrix} \\
\boldsymbol{S}_i &= \begin{bmatrix} \ddots & & 0 \\ & v_{ii}^{-1} \sum_u w_{iu} & \\ 0 & & \ddots \end{bmatrix}.
\end{aligned}
$$

Performing some additional matrix manipulation, we obtain the inhomogeneous Bühlmann-Straub estimator without conditional cross-sectional dependence in Wen and Wu (2011) as a special case of the inhomogeneous credibility estimator in Theorem 4.2. This is also true for the corresponding homogeneous credibility estimator in Theorem 4.3.

4.2 Estimation of Structural Parameters in a Spatial Context

In this section, results analogous to Section 3.2 are presented for the Bühlmann-Straub model with spatial dependence structure among risks and spatial conditional dependence. To reiterate, the structural parameters that are estimated are $\boldsymbol{\mu}$, $\boldsymbol{V}$, and $\boldsymbol{A}$. Assumptions 4.1, 3.2, and 3.4 are again used. Therefore, the structural parameters $\boldsymbol{V}$ and $\boldsymbol{A}$ are fully specified by the covariograms $g(\cdot)$ and $f(\cdot)$, respectively.

Additional notation is needed:

$$w_{i\cdot} = \sum_{u=1}^{n} w_{iu} \qquad \text{and} \qquad w_{\cdot\cdot} = \sum_{i=1}^{K} w_{i\cdot}.$$

An unbiased estimator of the unconditional mean $\boldsymbol{\mu}$ is given below.

Theorem 4.4. Under Assumptions 4.1 and 3.2, the estimator

$$\widehat{\mu}_i^{(bs)} = \left(\frac{1}{w_{i\cdot}}\right) \sum_{u=1}^{n} w_{iu} X_{iu} \tag{4.6}$$

is unbiased for μ_i for $i = 1, \dots, n$.

Proof.

$$\begin{aligned} \mathrm{E}\left[\widehat{\mu}_i^{(bs)}\right] &= \mathrm{E}\left[\frac{1}{w_{i\cdot}} \sum_{u=1}^{n} w_{iu} X_{iu}\right] \\ &= \frac{1}{w_{i\cdot}} \sum_{u=1}^{n} w_{iu} \, \mathrm{E}[X_{iu}] \\ &= \frac{1}{w_{i\cdot}} \sum_{u=1}^{n} w_{iu} \mu_i \\ &= \mu_i . \end{aligned}$$

□

Under the assumption of equal means $\mu_i = \mu$ for all $i = 1, \dots, K$, the unbiased estimator $\widehat{\mu}^{(bs)} = \frac{1}{w_{\cdot\cdot}} \sum_{i=1}^{K} w_{i\cdot} \widehat{\mu}_i^{(bs)}$ should be used to estimate μ.

For the expected process covariance matrix $\boldsymbol{V}$, we have the following estimator.

Theorem 4.5. Under Assumptions 4.1, 3.2, and 3.4, if it is further assumed that the spatial covariance function $g(\cdot)$ is equal to the constants $g(d_0), \dots, g(d_S)$ over the intervals

$d_0, \ldots, d_S$ respectively, the estimator

$$\widehat{g}^{(bs)}(d_s) = \frac{1}{|D_s|} \sum_{(i,j)\in D_s} \frac{1}{n_{ij}^*} \sum_{u=1}^{n} \sqrt{w_{iu}w_{ju}}(X_{iu} - \overline{X}_i^*)(X_{ju} - \overline{X}_j^*)$$

is unbiased for $g(d_s)$ for $s = 0, \ldots, S$, where

$$\begin{aligned} n_{ij}^* &= n - 2 + \frac{\left(\sum_u \sqrt{w_{iu}w_{ju}}\right)\left(\sum_v p_{iv}p_{jv}/\sqrt{w_{iv}w_{jv}}\right)}{\sum_u p_{iu} \sum_v p_{jv}} \\ \overline{X}_i^* &= \frac{\sum_u p_{iu}X_{iu}}{\sum_u p_{iu}}, \end{aligned}$$

for some set of positive constants $p_{iu}, i = 1, \ldots, K, u = 1, \ldots, n$.

Proof. Since

$$\begin{aligned} \mathrm{Cov}[X_{iu}, X_{jv}] &= \mathrm{Cov}[\mathrm{E}[X_{iu}|\boldsymbol{\Theta}], \mathrm{E}[X_{jv}|\boldsymbol{\Theta}]] + \mathrm{E}[\mathrm{Cov}[X_{iu}, X_{jv}|\boldsymbol{\Theta}]] \\ &= \mathrm{Cov}[\mu(\Theta_i), \mu(\Theta_j)] + \delta_{uv}\, \mathrm{E}\left[\frac{\sigma^2(\Theta_i, \Theta_j)}{\sqrt{w_{iu}w_{jv}}}\right] \\ &= a_{ij} + \delta_{uv}\frac{v_{ij}}{\sqrt{w_{iu}w_{jv}}}, \end{aligned}$$

we have

$$\begin{aligned} &\mathrm{E}\left[\sum_{u=1}^{n} \sqrt{w_{iu}w_{ju}}(X_{iu} - \overline{X}_i^*)(X_{ju} - \overline{X}_j^*)\right] \\ =&\,\mathrm{E}\left[\sum_{u=1}^{n} \sqrt{w_{iu}w_{ju}}(X_{iu} - \mu_i)(X_{ju} - \mu_j) + \sum_{u=1}^{n} \sqrt{w_{iu}w_{ju}}(\overline{X}_i^* - \mu_i)(\overline{X}_j^* - \mu_j)\right. \\ &\left. -(\overline{X}_j^* - \mu_j)\sum_{u=1}^{n} \sqrt{w_{iu}w_{ju}}(X_{iu} - \mu_i) - (\overline{X}_i^* - \mu_i)\sum_{u=1}^{n} \sqrt{w_{iu}w_{ju}}(X_{ju} - \mu_j)\right] \\ =&\,\mathrm{E}\left[\sum_{u=1}^{n} \sqrt{w_{iu}w_{ju}}(X_{iu} - \mu_i)(X_{ju} - \mu_j)\right. \\ &+ \frac{\sum_u \sqrt{w_{iu}w_{ju}}}{\sum_u p_{iu}\sum_v p_{jv}} \sum_{u=1}^{n}\sum_{v=1}^{n} p_{iu}p_{ju}(X_{iu} - \mu_i)(X_{jv} - \mu_j) \\ &- \frac{1}{\sum_u p_{ju}} \sum_{u=1}^{n}\sum_{v=1}^{n} \sqrt{w_{iu}w_{ju}}p_{jv}(X_{iu} - \mu_i)(X_{jv} - \mu_j) \\ &\left. - \frac{1}{\sum_u p_{iu}} \sum_{u=1}^{n}\sum_{v=1}^{n} \sqrt{w_{iu}w_{ju}}p_{iv}(X_{ju} - \mu_i)(X_{iv} - \mu_i)\right] \end{aligned}$$

$$
\begin{aligned}
=&\sum_{u=1}^{n}\sqrt{w_{iu}w_{ju}}\left(a_{ij}+\frac{v_{ij}}{\sqrt{w_{iu}w_{ju}}}\right)\\
&+\frac{\sum_u\sqrt{w_{iu}w_{ju}}}{\sum_u p_{iu}\sum_v p_{jv}}\sum_{u=1}^{n}\sum_{v=1}^{n}p_{iu}p_{jv}\left(a_{ij}+\delta_{uv}\frac{v_{ij}}{\sqrt{w_{iu}w_{jv}}}\right)\\
&-\frac{1}{\sum_u p_{ju}}\sum_{u=1}^{n}\sum_{v=1}^{n}\sqrt{w_{iu}w_{ju}}p_{jv}\left(a_{ij}+\delta_{uv}\frac{v_{ij}}{\sqrt{w_{iu}w_{jv}}}\right)\\
&-\frac{1}{\sum_u p_{iu}}\sum_{u=1}^{n}\sum_{v=1}^{n}\sqrt{w_{iu}w_{ju}}p_{iv}\left(a_{ij}+\delta_{uv}\frac{v_{ij}}{\sqrt{w_{iv}w_{ju}}}\right)\\
=&\sum_{u=1}^{n}\sqrt{w_{iu}w_{ju}}\frac{v_{ij}}{\sqrt{w_{iu}w_{ju}}}\\
&+\frac{\sum_u\sqrt{w_{iu}w_{ju}}}{\sum_u p_{iu}\sum_v p_{jv}}\sum_{u=1}^{n}p_{iu}p_{ju}\frac{v_{ij}}{\sqrt{w_{iu}w_{ju}}}\\
&-\frac{1}{\sum_u p_{ju}}\sum_{u=1}^{n}\sqrt{w_{iu}w_{ju}}p_{ju}\frac{v_{ij}}{\sqrt{w_{iu}w_{ju}}}\\
&-\frac{1}{\sum_u p_{iu}}\sum_{u=1}^{n}\sqrt{w_{iu}w_{ju}}p_{iu}\frac{v_{ij}}{\sqrt{w_{iu}w_{ju}}}\\
=&\left(n-2+\frac{\sum_u\sqrt{w_{iu}w_{ju}}\sum_v\frac{p_{iv}p_{jv}}{\sqrt{w_{iv}w_{jv}}}}{\sum_u p_{iu}\sum_v p_{jv}}\right)v_{ij}\\
=&n_{ij}^{*}v_{ij}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathrm{E}\left[\widehat{g}^{(bs)}(d_s)\right] &= \mathrm{E}\left[\frac{1}{|D_s|}\sum_{(i,j)\in D_s}\frac{1}{n_{ij}^{*}}\sum_{u=1}^{n}\sqrt{w_{iu}w_{ju}}(X_{iu}-\overline{X}_i^{*})(X_{ju}-\overline{X}_j^{*})\right]\\
&= \frac{1}{|D_s|}\sum_{(i,j)\in D_s}\frac{1}{n_{ij}^{*}}\mathrm{E}\left[\sum_{u=1}^{n}\sqrt{w_{iu}w_{ju}}(X_{iu}-\overline{X}_i^{*})(X_{ju}-\overline{X}_j^{*})\right]\\
&= \frac{1}{|D_s|}\sum_{(i,j)\in D_s}\frac{1}{n_{ij}^{*}}n_{ij}^{*}v_{ij}\\
&= \frac{1}{|D_s|}\sum_{(i,j)\in D_s}g(d_s)\\
&= g(d_s).
\end{aligned}
$$

□

To use Theorem 4.5, it is necessary to specify the constant p_{iu} for $i = 1, \ldots, K$ and $u = 1, \ldots, n$. The choice of p_{iu}'s may affect the efficiency of the estimator. The optimal selection of these weights is left as an open problem. A potential choice of p_{iu} is $p_{iu} = w_{iu}$ to be consistent with the weights used in $\widehat{\mu}_i$.

The following theorem allows one to construct unbiased estimator of the covariance matrix of hypothetical means $\boldsymbol{A}$ assuming $\mu_i = \mu$ for all $i = 1, \ldots, K$.

Theorem 4.6. Under the assumptions in Theorem 4.5, if it is further assumed that the spatial covariance function $f(\cdot)$ is equal to the constants $f(d_0), \ldots, f(d_S)$ over the intervals $d_0, \ldots, d_S$ respectively and that $\mu_i = \mu$ for all $i = 1, \ldots, K$, then

$$\widehat{\gamma}_f(d_s) = \left(\sum_{(i,j)\in D_s} \sqrt{w_{i\cdot} w_{j\cdot}} \right)^{-1} \left(\sum_{(i,j)\in D_s} \sqrt{w_{i\cdot} w_{j\cdot}} \frac{1}{2} (\overline{X}_i^* - \overline{X}_j^*)^2 - \widehat{\gamma}_g^*(d_s) \right)$$

is unbiased for the semivariogram of the spatial process of $\mu(\Theta_i)$, $\gamma_f(d_s) = f(d_0) - f(d_s)$, where

$$\begin{aligned} \widehat{\gamma}_g^*(d_s) &= \sum_{(i,j)\in D_s} \sqrt{w_{i\cdot} w_{j\cdot}} \frac{1}{2} \left(\frac{\sum_u p_{iu}^2 / w_{iu}}{(\sum_u p_{iu})^2} \widehat{v}_{ii}^{(bs)} + \frac{\sum_u p_{ju}^2 / w_{ju}}{(\sum_u p_{ju})^2} \widehat{v}_{jj}^{(bs)} \right. \\ &\qquad \left. -2 \frac{\sum_u p_{iu} p_{ju} / \sqrt{w_{iu} w_{ju}}}{\sum_u p_{iu} \sum_v p_{jv}} \widehat{v}_{ij}^{(bs)} \right) \end{aligned}$$

and

$$\widehat{v}_{ij}^{(bs)} = \frac{1}{n_{ij}^*} \sum_{u=1}^{n} \sqrt{w_{iu} w_{ju}} (X_{iu} - \overline{X}_i^*)(X_{ju} - \overline{X}_j^*).$$

Proof. First,

$$\begin{aligned} & \mathrm{E}\left[\sum_{(i,j)\in D_s} \frac{1}{2} \sqrt{w_{i\cdot} w_{j\cdot}} (\overline{X}_i^* - \overline{X}_j^*)^2 \right] \\ &= \mathrm{E}\left[\sum_{(i,j)\in D_s} \frac{1}{2} \sqrt{w_{i\cdot} w_{j\cdot}} \left((\overline{X}_i^* - \mu_i)^2 + (\overline{X}_j^* - \mu_j)^2 - 2(\overline{X}_i^* - \mu_i)(\overline{X}_j^* - \mu_j) \right) \right] \\ &\quad + \mathrm{E}\left[\sum_{(i,j)\in D_s} \frac{1}{2} \sqrt{w_{i\cdot} w_{j\cdot}} \cdot 2 \left(\overline{X}_i^* - \mu_i - (\overline{X}_j^* - \mu_j) \right) (\mu_i - \mu_j) \right] \\ &\quad + \sum_{(i,j)\in D_s} \frac{1}{2} \sqrt{w_{i\cdot} w_{j\cdot}} (\mu_i - \mu_j)^2 \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2}\sum_{(i,j)\in D_s}\sqrt{w_{i\cdot}w_{j\cdot}}\,\mathrm{E}\left[\frac{\sum_u\sum_v p_{iu}p_{iv}(X_{iu}-\mu_i)(X_{iv}-\mu_i)}{(\sum_u p_{iu})^2}\right.\\
&\quad\left.+\frac{\sum_u\sum_v p_{ju}p_{jv}(X_{ju}-\mu_j)(X_{jv}-\mu_j)}{(\sum_u p_{ju})^2}-2\frac{\sum_u\sum_v p_{iu}p_{jv}(X_{iu}-\mu_i)(X_{jv}-\mu_j)}{\sum_u p_{iu}\sum_v p_{jv}}\right]\\
&\quad+\frac{1}{2}\sum_{(i,j)\in D_s}\sqrt{w_{i\cdot}w_{j\cdot}}(\mu_i-\mu_j)^2\\
&= \frac{1}{2}\sum_{(i,j)\in D_s}\sqrt{w_{i\cdot}w_{j\cdot}}\left\{\frac{\sum_u\sum_v p_{iu}p_{iv}(a_{ii}+\delta_{uv}v_{ii}/\sqrt{w_{iu}w_{iv}})}{(\sum_u p_{iu})^2}\right.\\
&\quad\left.+\frac{\sum_u\sum_v p_{ju}p_{jv}(a_{jj}+\delta_{uv}v_{jj}/\sqrt{w_{ju}w_{jv}})}{(\sum_u p_{ju})^2}-2\frac{\sum_u\sum_v p_{iu}p_{jv}(a_{ij}+\delta_{uv}v_{ij}/\sqrt{w_{iu}w_{jv}})}{\sum_u p_{iu}\sum_v p_{jv}}\right\}\\
&\quad+\frac{1}{2}\sum_{(i,j)\in D_s}\sqrt{w_{i\cdot}w_{j\cdot}}(\mu_i-\mu_j)^2\\
&= \frac{1}{2}\sum_{(i,j)\in D_s}\sqrt{w_{i\cdot}w_{j\cdot}}\,\{a_{ii}+a_{jj}-2a_{ij}\}\\
&\quad+\frac{1}{2}\sum_{(i,j)\in D_s}\left\{v_{ii}\frac{\sum_u p_{iu}^2/w_{iu}}{(\sum_u p_{iu})^2}+v_{jj}\frac{\sum_u p_{ju}^2/w_{ju}}{(\sum_u p_{ju})^2}-2v_{ij}\frac{\sum_u p_{iu}p_{ju}/\sqrt{w_{iu}w_{ju}}}{\sum_u p_{iu}\sum_v p_{jv}}\right\}\\
&\quad+\frac{1}{2}\sum_{(i,j)\in D_s}\sqrt{w_{i\cdot}w_{j\cdot}}(\mu_i-\mu_j)^2\\
&= \sum_{(i,j)\in D_s}\sqrt{w_{i\cdot}w_{j\cdot}}\,\{f(0)-f(d_s)\}\\
&\quad+\frac{1}{2}\sum_{(i,j)\in D_s}\sqrt{w_{i\cdot}w_{j\cdot}}\left\{v_{ii}\frac{\sum_u p_{iu}^2/w_{iu}}{(\sum_u p_{iu})^2}+v_{jj}\frac{\sum_u p_{ju}^2/w_{ju}}{(\sum_u p_{ju})^2}-2v_{ij}\frac{\sum_u p_{iu}p_{ju}/\sqrt{w_{iu}w_{ju}}}{\sum_u p_{iu}\sum_v p_{jv}}\right\}\\
&\quad+\frac{1}{2}\sum_{(i,j)\in D_s}\sqrt{w_{i\cdot}w_{j\cdot}}(\mu_i-\mu_j)^2\\
&= \left(\sum_{(i,j)\in D_s}\sqrt{w_{i\cdot}w_{j\cdot}}\right)(f(0)-f(d_s))+\gamma_g^*(d_s)+\frac{1}{2}\sum_{(i,j)\in D_s}\sqrt{w_{i\cdot}w_{j\cdot}}(\mu_i-\mu_j)^2,
\end{aligned}
$$

where

$$
\gamma_g^*(d_s)=\frac{1}{2}\sum_{(i,j)\in D_s}\sqrt{w_{i\cdot}w_{j\cdot}}\left\{v_{ii}\frac{\sum_u p_{iu}^2/w_{iu}}{(\sum_u p_{iu})^2}+v_{jj}\frac{\sum_u p_{ju}^2/w_{ju}}{(\sum_u p_{ju})^2}-2v_{ij}\frac{\sum_u p_{iu}p_{ju}/\sqrt{w_{iu}w_{ju}}}{\sum_u p_{iu}\sum_v p_{jv}}\right\}.
$$

Note that $\gamma_g^*(d_s)$ is not a semivariogram. Now, since $\widehat{v}_{ij}^{(bs)}$ is unbiased for v_{ij} as shown in

Theorem 4.5, it is easy to see that $\widehat{\gamma}^*_g(d_s)$ is also unbiased for $\gamma^*_g(d_s)$. Therefore, we have

$$\begin{aligned}
\mathrm{E}[\widehat{\gamma}_f(d_s)] &= \left(\sum_{(i,j)\in D_s}\sqrt{w_{i\cdot}w_{j\cdot}}\right)^{-1}\mathrm{E}\left[\sum_{(i,j)\in D_s}\sqrt{w_{i\cdot}w_{j\cdot}}\frac{1}{2}(\overline{X}^*_i-\overline{X}^*_j)^2-\widehat{\gamma}^*_g(d_s)\right] \\
&= \left(\sum_{(i,j)\in D_s}\sqrt{w_{i\cdot}w_{j\cdot}}\right)^{-1}\left(\sum_{(i,j)\in D_s}\sqrt{w_{i\cdot}w_{j\cdot}}\right)(f(0)-f(d_s)) \\
&\quad+\left(\sum_{(i,j)\in D_s}\sqrt{w_{i\cdot}w_{j\cdot}}\right)^{-1}\frac{1}{2}\sum_{(i,j)\in D_s}\sqrt{w_{i\cdot}w_{j\cdot}}(\mu_i-\mu_j)^2 \\
&= f(0)-f(d_s)+\left(\sum_{(i,j)\in D_s}\sqrt{w_{i\cdot}w_{j\cdot}}\right)^{-1}\frac{1}{2}\sum_{(i,j)\in D_s}\sqrt{w_{i\cdot}w_{j\cdot}}(\mu_i-\mu_j)^2 \\
&= \gamma_f(d_s),
\end{aligned}$$

when equal means are assumed. □

Following the method discussed in Section 3.2, the following estimator can be used to estimate $f(\cdot)$:

$$\widehat{f}(d_s)=\widehat{\gamma}_f(d_{s^*})-\widehat{\gamma}_f(d_s),$$

for $s=0,\ldots,S$.

Again, the estimator $\widehat{\gamma}_f(d_s)$ (and therefore $\widehat{f}(\cdot)$) is only unbiased when the unconditional means are equal (i.e. $\mu_i=\mu$ for $i=1,\ldots,K$). When unconditional means are unequal, the expectation of $\widehat{\gamma}_f(d_s)$ is

$$\mathrm{E}[\widehat{\gamma}_f(d_s)]=\gamma_f(d_s)+\left(\sum_{(i,j)\in D_s}\sqrt{w_{i\cdot}w_{j\cdot}}\right)^{-1}\sum_{(i,j)\in D_s}\frac{1}{2}\sqrt{w_{i\cdot}w_{j\cdot}}(\mu_i-\mu_j)^2,$$

as seen in the proof of Theorem 4.6.

Chapter 5

Regression Credibility Model with General Dependence Structure and Spatio-Temporal Dependence

In this chapter, a regression credibility model with general dependence among risk parameters and spatio-temporal dependence is considered. In the following section, the assumptions are stated and the credibility estimator is derived.

5.1 The Credibility Estimator

Compared to the previous models with fixed means, the mean of an observation X_{iu} is now assumed to be a linear combinations of predictors. Also, conditional temporal dependence is considered in addition to the conditional spatial dependence already assumed in the previous models. Mathematically, we have the following assumption.

Assumption 5.1. Assume $X_{iu} = \boldsymbol{Y}_{iu}\boldsymbol{\beta}_i(\Theta_i) + \epsilon_{iu}, i = 1, \ldots, K, u = 1, \ldots, n+1$, where $\boldsymbol{Y}_{iu}$ is a $1 \times p$ design matrix with $p < n$, $\boldsymbol{\beta}_i(\Theta_i) = (\beta_{i1}(\Theta_i), \ldots, \beta_{ip}(\Theta_i))'$ is a column vector of p linear coefficients that are dependent on the risk parameter Θ_i, and ϵ_{iu} is a random fluctuation. Denote $\boldsymbol{\epsilon}_{(u)} = (\epsilon_{1u}, \ldots, \epsilon_{Ku})'$. Assume for $i = 1, \ldots, K, u, v = 1, \ldots, n+1$ that ϵ_{iu} has conditional expectation $\mathrm{E}[\epsilon_{iu}|\boldsymbol{\Theta}] = \mathrm{E}[\epsilon_{iu}|\Theta_i] = 0$ and conditional covariance matrix $\mathrm{Cov}[\boldsymbol{\epsilon}_{(u)}, \boldsymbol{\epsilon}_{(v)}|\boldsymbol{\Theta}] = \boldsymbol{W}_{(u)}^{-\frac{1}{2}} v_{(uv)} \boldsymbol{V}_S \boldsymbol{W}_{(v)}^{-\frac{1}{2}} = \left[\frac{v_{(uv)}v_{ij}}{\sqrt{w_{iu}w_{jv}}}\right]_{i,j=1,\ldots,K}$, where $\boldsymbol{V}_T = \left[v_{(uv)}\right]_{u,v=1,\ldots,n}$

is a temporal covariance matrix and $\boldsymbol{V}_S = \left[v_{ij}\right]_{i,j=1,\ldots,K}$ is a cross-sectional covariance matrix. Finally, assume that $\mathrm{E}[\boldsymbol{\beta}_i(\Theta_i)] = \boldsymbol{b}_i$, $\mathrm{Cov}[\boldsymbol{\beta}_i(\Theta_i), \boldsymbol{\beta}_j(\Theta_j)] = \boldsymbol{M}_{ij}$.

In Assumption 5.1, the error terms ϵ_{iu} no longer have covariances that are conditional on the risk parameters $\boldsymbol{\Theta}$. Also, it can be seen that the spatio-temporal covariance of the errors ϵ_{iu} is separable in the sense that

$$w_{iu} w_{jv} \, \mathrm{Cov}[\epsilon_{iu}, \epsilon_{jv}] = v_{ij} \cdot v_{(uv)}.$$

This assumption is imposed to facilitate the inversion of the spatio-temporal matrix of all the error terms by the use of separable spatio-temporal covariance matrix when the dimensions of matrix may be too large to invert numerically. However, it should be noted that although separable spatio-temporal covariance functions remain in common use, they are criticized to be often unrealistic in applications due to their restrictive nature in representing different forms of spatio-temporal dependence. See Cressie and Wilkie (2011) for more information on separable spatio-temporal covariance functions and other methods to facilitate inversion of large dimension spatio-temporal covariance matrices.

The following notations are used in this section. For $i = 1, \ldots, K$

$$\boldsymbol{Y}_i = \begin{bmatrix} \boldsymbol{Y}_{i1} \\ \vdots \\ \boldsymbol{Y}_{in} \end{bmatrix}, \qquad \boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_1 & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{Y}_K \end{bmatrix}, \qquad \boldsymbol{Y}_{(n+1)} = \begin{bmatrix} \boldsymbol{Y}_{1,n+1} & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{Y}_{K,n+1} \end{bmatrix},$$

$$\boldsymbol{\beta}(\boldsymbol{\Theta}) = \begin{bmatrix} \boldsymbol{\beta}_1(\Theta_1) \\ \vdots \\ \boldsymbol{\beta}_K(\Theta_K) \end{bmatrix}, \quad \boldsymbol{b} = \mathrm{E}[\boldsymbol{\beta}(\boldsymbol{\Theta})] = \begin{bmatrix} \boldsymbol{b}_1 \\ \vdots \\ \boldsymbol{b}_K \end{bmatrix}, \quad \boldsymbol{M} = \mathrm{Cov}[\boldsymbol{\beta}(\boldsymbol{\Theta})] = \begin{bmatrix} \boldsymbol{M}_{11} & \ldots & \boldsymbol{M}_{1K} \\ \vdots & \ddots & \vdots \\ \boldsymbol{M}_{K1} & \ldots & \boldsymbol{M}_{KK} \end{bmatrix},$$

and

$$\boldsymbol{\epsilon}_i = \begin{bmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{in} \end{bmatrix}, \qquad \boldsymbol{\epsilon} = \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \vdots \\ \boldsymbol{\epsilon}_K \end{bmatrix}, \qquad \boldsymbol{\epsilon}_{(n+1)} = \begin{bmatrix} \epsilon_{1,n+1} \\ \vdots \\ \epsilon_{K,n+1} \end{bmatrix}.$$

Further denote the temporal covariance vector by $\boldsymbol{v}_{(n+1)} = \big(v_{(1,n+1)}, \ldots, v_{(n,n+1)}\big)'$.

Theorem 5.1. Under Assumptions 5.1 and the notations introduced previously, the inhomogeneous credibility premium of $\boldsymbol{X}_{(n+1)}$, denoted by $\widehat{\boldsymbol{X}}^{(reg)}_{(n+1)}$, that is obtained by minimizing (2.1) in the class of inhomogeneous linear functions $L(X,1)$ is given by

$$\widehat{\boldsymbol{X}}^{(reg)}_{(n+1)} = \boldsymbol{Z}\boldsymbol{X} + (\boldsymbol{\pi} - \boldsymbol{Z})\widehat{\boldsymbol{Y}\boldsymbol{\beta}}, \tag{5.1}$$

and the inhomogenous credibility estimator of the conditional mean $\boldsymbol{Y}\boldsymbol{\beta}(\boldsymbol{\Theta})$ is

$$\widehat{\boldsymbol{Y}\boldsymbol{\beta}} = \boldsymbol{Y}\left\{\boldsymbol{Q}\widehat{\boldsymbol{\beta}}^{(GLS)}_{\Theta} + (\boldsymbol{I}_{Kp} - \boldsymbol{Q})\boldsymbol{b}\right\},$$

where

$$\widehat{\boldsymbol{\beta}}^{(GLS)}_{\Theta} = (\boldsymbol{Y}'\boldsymbol{V}^{-1}_{XX}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{V}^{-1}_{XX}\boldsymbol{X}$$

is the generalized least square estimator of $\boldsymbol{\beta}(\boldsymbol{\Theta})$ given knowledge of $\boldsymbol{\Theta}$,

$$\begin{aligned}
\boldsymbol{Z} &= \boldsymbol{V}_{X_{(n+1)}X}\boldsymbol{V}^{-1}_{XX} && (5.2)\\
\boldsymbol{Q} &= \boldsymbol{M}\left(\boldsymbol{M} + \left(\boldsymbol{Y}'\boldsymbol{V}^{-1}_{XX}\boldsymbol{Y}\right)^{-1}\right)^{-1} && (5.3)\\
\boldsymbol{V}^{-1}_{XX} &= \boldsymbol{W}^{\frac{1}{2}}_{X}(\boldsymbol{V}^{-1}_{S} \otimes \boldsymbol{V}^{-1}_{T})\boldsymbol{W}^{\frac{1}{2}}_{X}\\
\boldsymbol{V}_{X_{(n+1)}X} &= \boldsymbol{W}^{-\frac{1}{2}}_{(n+1)}(\boldsymbol{V}_{S} \otimes \boldsymbol{v}'_{(n+1)})\boldsymbol{W}^{-\frac{1}{2}}_{X}\\
\boldsymbol{\pi} &= \begin{bmatrix} \boldsymbol{\pi}_1 & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{\pi}_K \end{bmatrix}, && (5.4)
\end{aligned}$$

and the $1 \times n$ matrix $\boldsymbol{\pi}_i$ is the solution of the equation $\boldsymbol{\pi}_i\boldsymbol{Y}_i = \boldsymbol{Y}_{i,n+1}$, $i = 1, \ldots, K$. The associated mean square prediction error matrix is

$$\begin{aligned}
&\mathrm{E}\left[\left(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}^{(reg)}_{(n+1)}\right)\left(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}^{(reg)}_{(n+1)}\right)'\right]\\
&= \boldsymbol{Y}_{(n+1)}\left(\boldsymbol{I}_{Kp} - \boldsymbol{Q}\right)\boldsymbol{M}\boldsymbol{Y}'_{(n+1)} + \boldsymbol{W}^{-\frac{1}{2}}_{(n+1)}v_{(n+1,n+1)}\boldsymbol{V}_S\boldsymbol{W}^{-\frac{1}{2}}_{(n+1)}\\
&\quad -\boldsymbol{Z}\boldsymbol{Y}\left(\boldsymbol{I}_{Kp} - \boldsymbol{Q}\right)\boldsymbol{M}\boldsymbol{Y}'\boldsymbol{Z}' - \boldsymbol{Z}\boldsymbol{V}'_{X_{(n+1)}X}\\
&\quad -\boldsymbol{V}_{X_{(n+1)}X}\left((\boldsymbol{\pi} - \boldsymbol{Z})\boldsymbol{Y}\boldsymbol{Q}(\boldsymbol{Y}'\boldsymbol{V}^{-1}_{XX}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{V}^{-1}_{XX}\right)'\\
&\quad -\left((\boldsymbol{\pi} - \boldsymbol{Z})\boldsymbol{Y}\boldsymbol{Q}(\boldsymbol{Y}'\boldsymbol{V}^{-1}_{XX}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{V}^{-1}_{XX}\right)\boldsymbol{V}'_{X_{(n+1)}X}. && (5.5)
\end{aligned}$$

Proof. First, suppose

$$\mathrm{Proj}(\boldsymbol{X}_{(n+1)}|L(\boldsymbol{X},1)) = \boldsymbol{c}_0 + \sum_{i=1}^{K}\boldsymbol{C}_i\boldsymbol{X}_i,$$

where $\boldsymbol{c}_0$ is a vector of length K and $\boldsymbol{C}_i$ is a $K \times K$ matrix for $i = 1, \ldots, K$. Denote the projection of $\boldsymbol{X}_{(n+1)}$ conditional on $\boldsymbol{\Theta}$ by

$$\widehat{\boldsymbol{X}}_{(n+1)|\boldsymbol{\Theta}} = \text{Proj}(\boldsymbol{X}_{(n+1)}|L(\boldsymbol{X}, 1), \boldsymbol{\Theta}).$$

By the definition of orthogonal projection onto $L(1, \boldsymbol{X})|\boldsymbol{\Theta}$, for $i = 1, \ldots, K$

$$\mathrm{E}\left[\left(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}_{(n+1)|\boldsymbol{\Theta}}\right) \boldsymbol{X}_i'|\boldsymbol{\Theta}\right] = \boldsymbol{0},$$

we have

$$\mathrm{E}\left[\left(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}_{(n+1)|\boldsymbol{\Theta}}\right)\left(\boldsymbol{c}_0 + \sum_{i=1}^{K} \boldsymbol{C}_i \boldsymbol{X}_i - \widehat{\boldsymbol{X}}_{(n+1)|\boldsymbol{\Theta}})\right)'\right] = \boldsymbol{0}.$$

Therefore, we have

$$\begin{aligned}
&\mathrm{E}[(\boldsymbol{X}_{(n+1)} - \text{Proj}(\boldsymbol{X}_{(n+1)}|L(\boldsymbol{X}, 1)))(\boldsymbol{X}_{(n+1)} - \text{Proj}(\boldsymbol{X}_{(n+1)}|L(\boldsymbol{X}, 1)))']\\
&= \mathrm{E}\left[\left(\boldsymbol{X}_{(n+1)} - \left(\boldsymbol{c}_0 + \sum_{i=1}^{K} \boldsymbol{C}_i \boldsymbol{X}_i\right)\right)\left(\boldsymbol{X}_{(n+1)} - \left(\boldsymbol{c}_0 + \sum_{i=1}^{K} \boldsymbol{C}_i \boldsymbol{X}_i\right)\right)'\right]\\
&= \mathrm{E}\left[\left(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}_{(n+1)|\boldsymbol{\Theta}} - \left(\boldsymbol{c}_0 + \sum_{i=1}^{K} \boldsymbol{C}_i \boldsymbol{X}_i - \widehat{\boldsymbol{X}}_{(n+1)|\boldsymbol{\Theta}}\right)\right)\right.\\
&\qquad \left.\left(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}_{(n+1)|\boldsymbol{\Theta}} - \left(\boldsymbol{c}_0 + \sum_{i=1}^{K} \boldsymbol{C}_i \boldsymbol{X}_i - \widehat{\boldsymbol{X}}_{(n+1)|\boldsymbol{\Theta}}\right)\right)'\right]\\
&= \mathrm{E}[(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}_{(n+1)|\boldsymbol{\Theta}})(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}_{(n+1)|\boldsymbol{\Theta}})'] - \boldsymbol{0} - \boldsymbol{0}\\
&\quad + \mathrm{E}\left[\left(\boldsymbol{c}_0 + \sum_{i=1}^{K} \boldsymbol{C}_i \boldsymbol{X}_i - \widehat{\boldsymbol{X}}_{(n+1)|\boldsymbol{\Theta}})\right)\left(\boldsymbol{c}_0 + \sum_{i=1}^{K} \boldsymbol{C}_i \boldsymbol{X}_i - \widehat{\boldsymbol{X}}_{(n+1)|\boldsymbol{\Theta}})\right)'\right].
\end{aligned}$$

It can be seen that the first term is independent of $\boldsymbol{c}_0$ and $\boldsymbol{C}_i$ for $i = 1, \ldots, K$. That is, in the projection of $\boldsymbol{X}_{(n+1)}$ onto $L(1, \boldsymbol{X})$, the minimization of the expected quadratic loss from $\boldsymbol{X}_{(n+1)}$ by finding optimal values of $\boldsymbol{c}_0$ and $\boldsymbol{C}_i$ also minimizes the expected quadratic loss from $\widehat{\boldsymbol{X}}_{(n+1)|\boldsymbol{\Theta}}$. Therefore, $\text{Proj}(\boldsymbol{X}_{(n+1)}|L(\boldsymbol{X}, 1)) = \text{Proj}(\text{Proj}(\boldsymbol{X}_{(n+1)}|L(\boldsymbol{X}, 1), \boldsymbol{\Theta})|L(\boldsymbol{X}, 1))$. Note that this proof of iterativity of orthogonal projection (see Lemma 2.3) comes from Wen and Wu (2011).

Now, solving for $\mathrm{Proj}(\boldsymbol{X}_{(n+1)}|L(\boldsymbol{X},1),\boldsymbol{\Theta})$ gives

$$\begin{aligned}
\boldsymbol{\Sigma}_{X_{(n+1)}X|\Theta} &= \mathrm{Cov}[\boldsymbol{X}_{(n+1)},\boldsymbol{X}|\boldsymbol{\Theta}] \\
&= \mathrm{Cov}[\boldsymbol{Y}_{(n+1)}\boldsymbol{\beta}(\boldsymbol{\Theta})+\boldsymbol{\epsilon}_{(n+1)},\boldsymbol{Y}\boldsymbol{\beta}(\boldsymbol{\Theta})+\boldsymbol{\epsilon}|\boldsymbol{\Theta}] \\
&= \mathrm{Cov}[\boldsymbol{\epsilon}_{(n+1)},\boldsymbol{\epsilon}|\boldsymbol{\Theta}] \\
&= \boldsymbol{W}_{(n+1)}^{-\frac{1}{2}}(\boldsymbol{V}_S\otimes\boldsymbol{v}'_{(n+1)})\boldsymbol{W}_X^{-\frac{1}{2}} \\
&= \boldsymbol{V}_{X_{(n+1)}X}, \\
\boldsymbol{\Sigma}_{XX|\Theta} &= \mathrm{Cov}[\boldsymbol{X}|\boldsymbol{\Theta}] \\
&= \mathrm{Cov}[\boldsymbol{Y}\boldsymbol{\beta}(\boldsymbol{\Theta})+\boldsymbol{\epsilon}|\boldsymbol{\Theta}] \\
&= \mathrm{Cov}[\boldsymbol{\epsilon}|\boldsymbol{\Theta}] \\
&= \boldsymbol{W}_X^{-\frac{1}{2}}(\boldsymbol{V}_S\otimes\boldsymbol{V}'_T)\boldsymbol{W}_X^{-\frac{1}{2}} \\
&= \boldsymbol{V}_{XX}.
\end{aligned}$$

Hence, by Lemma 2.1,

$$\begin{aligned}
\mathrm{Proj}(\boldsymbol{X}_{(n+1)}|L(\boldsymbol{\Theta},1),\boldsymbol{\Theta}) &= \mathrm{E}[\boldsymbol{X}_{(n+1)}|\boldsymbol{\Theta}] \\
&\quad + \mathrm{Cov}[\boldsymbol{X}_{(n+1)},\boldsymbol{X}|\boldsymbol{\Theta}]\left(\mathrm{Cov}[\boldsymbol{X}|\boldsymbol{\Theta}]\right)^{-1}(\boldsymbol{X}-\mathrm{E}[\boldsymbol{X}|\boldsymbol{\Theta}]) \\
&= \boldsymbol{Y}_{(n+1)}\boldsymbol{\beta}(\boldsymbol{\Theta})+\boldsymbol{V}_{X_{(n+1)}X}\boldsymbol{V}_{XX}^{-1}(\boldsymbol{X}-\boldsymbol{Y}\boldsymbol{\beta}(\boldsymbol{\Theta})) \\
&= \boldsymbol{Z}\boldsymbol{X}+(\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{Y}\boldsymbol{\beta}(\boldsymbol{\Theta}),
\end{aligned}$$

where $\boldsymbol{\pi}$, given by (5.4), satisfies $\boldsymbol{\pi}\boldsymbol{Y}=\boldsymbol{Y}_{(n+1)}$. Then, we have

$$\begin{aligned}
\mathrm{Proj}(\boldsymbol{X}_{(n+1)}|L(\boldsymbol{X},1)) &= \mathrm{Proj}(\mathrm{Proj}(\boldsymbol{X}_{(n+1)}|L(\boldsymbol{X},1),\boldsymbol{\Theta})|L(\boldsymbol{X},1)) \\
&= \mathrm{Proj}(\boldsymbol{Z}\boldsymbol{X}+(\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{Y}\boldsymbol{\beta}(\boldsymbol{\Theta})|L(\boldsymbol{X},1)) \\
&= \boldsymbol{Z}\boldsymbol{X}+(\boldsymbol{\pi}-\boldsymbol{Z})\,\mathrm{Proj}(\boldsymbol{Y}\boldsymbol{\beta}(\boldsymbol{\Theta})|L(\boldsymbol{X},1)).
\end{aligned}$$

Now, to obtain $\mathrm{Proj}(\boldsymbol{Y}\boldsymbol{\beta}(\boldsymbol{\Theta})|L(\boldsymbol{X},1))=\widehat{\boldsymbol{Y}\boldsymbol{\beta}}$, we have

$$\begin{aligned}
\boldsymbol{\Sigma}_{\boldsymbol{Y}\boldsymbol{\beta}(\boldsymbol{\Theta}),X} &= \mathrm{Cov}[\boldsymbol{Y}\boldsymbol{\beta}(\boldsymbol{\Theta}),\boldsymbol{X}] \\
&= \mathrm{Cov}[\mathrm{E}[\boldsymbol{Y}\boldsymbol{\beta}(\boldsymbol{\Theta})|\boldsymbol{\Theta}],\mathrm{E}[\boldsymbol{X}|\boldsymbol{\Theta}]]+\mathrm{E}[\mathrm{Cov}[\boldsymbol{Y}\boldsymbol{\beta}(\boldsymbol{\Theta}),\boldsymbol{X}|\boldsymbol{\Theta}]] \\
&= \mathrm{Cov}[\boldsymbol{Y}\boldsymbol{\beta}(\boldsymbol{\Theta})] \\
&= \boldsymbol{Y}\boldsymbol{M}\boldsymbol{Y}',
\end{aligned}$$

and

$$\begin{aligned}\boldsymbol{\Sigma}_{XX} &= \mathrm{Cov}[\boldsymbol{X}] \\ &= \mathrm{Cov}[\mathrm{E}[\boldsymbol{X}|\boldsymbol{\Theta}]] + \mathrm{E}[\mathrm{Cov}[\boldsymbol{X}|\boldsymbol{\Theta}]] \\ &= \mathrm{Cov}[\boldsymbol{Y}\boldsymbol{\beta}(\boldsymbol{\Theta})] + \mathrm{E}[\mathrm{Cov}[\boldsymbol{\epsilon}|\boldsymbol{\Theta}]] \\ &= \boldsymbol{Y}\boldsymbol{M}\boldsymbol{Y}' + \boldsymbol{V}_{XX}.\end{aligned}$$

Using $(\boldsymbol{E}+\boldsymbol{F}\boldsymbol{G}\boldsymbol{H})^{-1} = \boldsymbol{E}^{-1} - \boldsymbol{E}^{-1}\boldsymbol{F}(\boldsymbol{G}^{-1} + \boldsymbol{H}\boldsymbol{E}^{-1}\boldsymbol{F})^{-1}\boldsymbol{H}\boldsymbol{E}^{-1}$,

$$\begin{aligned}\boldsymbol{\Sigma}_{XX}^{-1} &= \left(\boldsymbol{Y}\boldsymbol{M}\boldsymbol{Y}' + \boldsymbol{V}_{XX}\right)^{-1} \\ &= \boldsymbol{V}_{XX}^{-1} - \boldsymbol{V}_{XX}^{-1}\boldsymbol{Y}(\boldsymbol{M}^{-1} + \boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1},\end{aligned}$$

and using $(\boldsymbol{E}+\boldsymbol{F})^{-1} = \boldsymbol{E}^{-1} - \boldsymbol{E}^{-1}\boldsymbol{F}(\boldsymbol{E}+\boldsymbol{F})^{-1}$, we further obtain

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{Y}\boldsymbol{\beta}(\boldsymbol{\Theta}),X}\boldsymbol{\Sigma}_{XX}^{-1} &= \boldsymbol{Y}(\boldsymbol{M} - \boldsymbol{M}\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\boldsymbol{Y}(\boldsymbol{M}^{-1} + \boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\boldsymbol{Y})^{-1})\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1} \\ &= \boldsymbol{Y}(\boldsymbol{M}^{-1} + \boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}.\end{aligned}$$

It follows that

$$\begin{aligned}&\boldsymbol{\Sigma}_{\boldsymbol{Y}\boldsymbol{\beta}(\boldsymbol{\Theta}),X}\boldsymbol{\Sigma}_{XX}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}_X) \\ &= \boldsymbol{Y}(\boldsymbol{M}^{-1} + \boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}(\boldsymbol{X} - \boldsymbol{Y}\boldsymbol{b}) \\ &= \boldsymbol{Y}(\boldsymbol{M}^{-1} + \boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\boldsymbol{Y})^{-1}((\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\boldsymbol{Y})(\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\boldsymbol{X} - \boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\boldsymbol{Y}\boldsymbol{b}) \\ &= \boldsymbol{Y}(\boldsymbol{M}^{-1} + \boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\boldsymbol{Y}\left(\widehat{\boldsymbol{\beta}}_{\Theta}^{(GLS)} - \boldsymbol{b}\right) \\ &= \boldsymbol{Y}\boldsymbol{M}\left(\boldsymbol{M} + \left(\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\boldsymbol{Y}\right)^{-1}\right)^{-1}\left(\widehat{\boldsymbol{\beta}}_{\Theta}^{(GLS)} - \boldsymbol{b}\right) \\ &= \boldsymbol{Y}\boldsymbol{Q}\left(\widehat{\boldsymbol{\beta}}_{\Theta}^{(GLS)} - \boldsymbol{b}\right),\end{aligned}$$

where the relation $(\boldsymbol{E}+\boldsymbol{F})^{-1}\boldsymbol{F} = \boldsymbol{E}^{-1}(\boldsymbol{F}^{-1} + \boldsymbol{E}^{-1})^{-1}$ is used to obtain the second last equality.

Finally, by Lemma 2.1, we have

$$\begin{aligned}\widehat{\boldsymbol{Y}\boldsymbol{\beta}} = \mathrm{Proj}(\boldsymbol{Y}\boldsymbol{\beta}(\boldsymbol{\Theta})|L(\boldsymbol{X},1)) &= \boldsymbol{\mu}_{\boldsymbol{Y}\boldsymbol{\beta}(\boldsymbol{\Theta})} + \boldsymbol{\Sigma}_{\boldsymbol{Y}\boldsymbol{\beta}(\boldsymbol{\Theta}),X}\boldsymbol{\Sigma}_{XX}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}_X) \\ &= \boldsymbol{Y}\boldsymbol{b} + \boldsymbol{Y}\boldsymbol{Q}\left(\widehat{\boldsymbol{\beta}}_{\Theta}^{(GLS)} - \boldsymbol{b}\right) \\ &= \boldsymbol{Y}\left\{\boldsymbol{Q}\widehat{\boldsymbol{\beta}}_{\Theta}^{(GLS)} + (\boldsymbol{I}_{Kp} - \boldsymbol{Q})\boldsymbol{b}\right\},\end{aligned}$$

from which the result desired, $\boldsymbol{X}_{(n+1)}^{(reg)}$, is obtained.

For the mean square prediction error, note that

$$\begin{aligned}
\widehat{\boldsymbol{X}}^{(reg)}_{(n+1)} &= \boldsymbol{Z}\boldsymbol{X} + (\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{Y}\left\{\boldsymbol{Q}(\boldsymbol{Y}'\boldsymbol{V}^{-1}_{XX}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{V}^{-1}_{XX}\boldsymbol{X} + (\boldsymbol{I}_{Kp}-\boldsymbol{Q})\boldsymbol{b}\right\} \\
&= \left(\boldsymbol{Z} + (\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{Y}\boldsymbol{Q}(\boldsymbol{Y}'\boldsymbol{V}^{-1}_{XX}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{V}^{-1}_{XX}\right)\boldsymbol{X} \\
&\quad + (\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{Y}(\boldsymbol{I}_{Kp}-\boldsymbol{Q})\boldsymbol{b}.
\end{aligned}$$

Hence,

$$\begin{aligned}
&\mathrm{E}\left[\left(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}^{(reg)}_{(n+1)}\right)\left(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}^{(reg)}_{(n+1)}\right)'\right] \\
&= \mathrm{Cov}\left[\boldsymbol{X}_{(n+1)}\right] + \mathrm{Cov}\left[\widehat{\boldsymbol{X}}^{(reg)}_{(n+1)}\right] \\
&\quad - \mathrm{Cov}\left[\boldsymbol{X}_{(n+1)}, \widehat{\boldsymbol{X}}^{(reg)}_{(n+1)}\right] - \mathrm{Cov}\left[\widehat{\boldsymbol{X}}^{(reg)}_{(n+1)}, \boldsymbol{X}_{(n+1)}\right] \\
&\quad + \mathrm{E}\left[\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}^{(reg)}_{(n+1)}\right]\mathrm{E}\left[\left(\boldsymbol{X}_{(n+1)} - \widehat{\boldsymbol{X}}^{(reg)}_{(n+1)}\right)'\right] \\
&= \boldsymbol{Y}_{(n+1)}\boldsymbol{M}\boldsymbol{Y}'_{(n+1)} + \boldsymbol{W}^{-\frac{1}{2}}_{(n+1)}v_{(n+1,n+1)}\boldsymbol{V}_S\boldsymbol{W}^{-\frac{1}{2}}_{(n+1)} \\
&\quad + \left(\boldsymbol{Z} + (\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{Y}\boldsymbol{Q}(\boldsymbol{Y}'\boldsymbol{V}^{-1}_{XX}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{V}^{-1}_{XX}\right)\left(\boldsymbol{Y}\boldsymbol{M}\boldsymbol{Y}' + \boldsymbol{V}_{XX}\right) \\
&\quad \times \left(\boldsymbol{Z} + (\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{Y}\boldsymbol{Q}(\boldsymbol{Y}'\boldsymbol{V}^{-1}_{XX}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{V}^{-1}_{XX}\right)' \\
&\quad - \left(\boldsymbol{Y}_{(n+1)}\boldsymbol{M}\boldsymbol{Y}' + \boldsymbol{V}_{X_{(n+1)}X}\right)\left(\boldsymbol{Z} + (\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{Y}\boldsymbol{Q}(\boldsymbol{Y}'\boldsymbol{V}^{-1}_{XX}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{V}^{-1}_{XX}\right)' \\
&\quad - \left(\boldsymbol{Z} + (\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{Y}\boldsymbol{Q}(\boldsymbol{Y}'\boldsymbol{V}^{-1}_{XX}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{V}^{-1}_{XX}\right)\left(\boldsymbol{Y}_{(n+1)}\boldsymbol{M}\boldsymbol{Y}' + \boldsymbol{V}_{X_{(n+1)}X}\right)' \\
&= \boldsymbol{Y}_{(n+1)}\boldsymbol{M}\boldsymbol{Y}'_{(n+1)} + \boldsymbol{W}^{-\frac{1}{2}}_{(n+1)}v_{(n+1,n+1)}\boldsymbol{V}_S\boldsymbol{W}^{-\frac{1}{2}}_{(n+1)} \\
&\quad + \boldsymbol{Z}\boldsymbol{Y}\boldsymbol{M}\boldsymbol{Y}'\boldsymbol{Z}' + (\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{Y}\boldsymbol{Q}\boldsymbol{M}\boldsymbol{Q}'\boldsymbol{Y}'(\boldsymbol{\pi}-\boldsymbol{Z})' \\
&\quad + \boldsymbol{Z}\boldsymbol{Y}\boldsymbol{M}\boldsymbol{Q}'\boldsymbol{Y}'(\boldsymbol{\pi}-\boldsymbol{Z})' + (\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{Y}\boldsymbol{Q}\boldsymbol{M}\boldsymbol{Y}'\boldsymbol{Z}' \\
&\quad + \boldsymbol{Z}\boldsymbol{V}_{XX}\boldsymbol{Z}' + (\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{Y}\boldsymbol{Q}(\boldsymbol{Y}'\boldsymbol{V}^{-1}_{XX}\boldsymbol{Y})^{-1}\boldsymbol{Q}'\boldsymbol{Y}'(\boldsymbol{\pi}-\boldsymbol{Z})' \\
&\quad + \boldsymbol{Z}\boldsymbol{Y}(\boldsymbol{Y}'\boldsymbol{V}^{-1}_{XX}\boldsymbol{Y})^{-1}\boldsymbol{Q}'\boldsymbol{Y}'(\boldsymbol{\pi}-\boldsymbol{Z})' + (\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{Y}\boldsymbol{Q}(\boldsymbol{Y}'\boldsymbol{V}^{-1}_{XX}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{Z}' \\
&\quad - \boldsymbol{Y}_{(n+1)}\boldsymbol{M}\boldsymbol{Y}'\boldsymbol{Z}' - \boldsymbol{Y}_{(n+1)}\boldsymbol{M}\boldsymbol{Q}'\boldsymbol{Y}'(\boldsymbol{\pi}-\boldsymbol{Z})' \\
&\quad - \boldsymbol{V}_{X_{(n+1)}X}\left(\boldsymbol{Z} + (\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{Y}\boldsymbol{Q}(\boldsymbol{Y}'\boldsymbol{V}^{-1}_{XX}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{V}^{-1}_{XX}\right)' \\
&\quad - \boldsymbol{Z}\boldsymbol{Y}\boldsymbol{M}\boldsymbol{Y}'_{(n+1)} - (\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{Y}\boldsymbol{Q}\boldsymbol{M}\boldsymbol{Y}'_{(n+1)} \\
&\quad - \left(\boldsymbol{Z} + (\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{Y}\boldsymbol{Q}(\boldsymbol{Y}'\boldsymbol{V}^{-1}_{XX}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{V}^{-1}_{XX}\right)\boldsymbol{V}'_{X_{(n+1)}X} \\
&= \boldsymbol{Y}_{(n+1)}\boldsymbol{M}\boldsymbol{Y}'_{(n+1)} + \boldsymbol{W}^{-\frac{1}{2}}_{(n+1)}v_{(n+1,n+1)}\boldsymbol{V}_S\boldsymbol{W}^{-\frac{1}{2}}_{(n+1)} \\
&\quad + \boldsymbol{Z}\left(\boldsymbol{Y}\boldsymbol{M}\boldsymbol{Y}' + \boldsymbol{V}_{XX}\right)\boldsymbol{Z}' + (\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{Y}\boldsymbol{M}\boldsymbol{Q}'\boldsymbol{Y}'(\boldsymbol{\pi}-\boldsymbol{Z})'
\end{aligned}$$

$$
\begin{aligned}
& +\boldsymbol{ZYMY}'(\boldsymbol{\pi}-\boldsymbol{Z})' + (\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{YMY}'\boldsymbol{Z}' \\
& -\boldsymbol{\pi YMY}'\boldsymbol{Z}' - \boldsymbol{\pi YMQ}'\boldsymbol{Y}'(\boldsymbol{\pi}-\boldsymbol{Z})' \\
& -\boldsymbol{V}_{X_{(n+1)}X}\left(\boldsymbol{Z} + (\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{YQ}(\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\right)' \\
& -\boldsymbol{ZYMY}'\boldsymbol{\pi}' - (\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{YQMY}'\boldsymbol{\pi}' \\
& -\left(\boldsymbol{Z} + (\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{YQ}(\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\right)\boldsymbol{V}'_{X_{(n+1)}X} \\
= \; & \boldsymbol{Y}_{(n+1)}\boldsymbol{MY}'_{(n+1)} + \boldsymbol{W}_{(n+1)}^{-\frac{1}{2}} v_{(n+1,n+1)}\boldsymbol{V}_S\boldsymbol{W}_{(n+1)}^{-\frac{1}{2}} \\
& +\boldsymbol{ZV}_{XX}\boldsymbol{Z}' - \boldsymbol{ZYMQ}'\boldsymbol{Y}'(\boldsymbol{\pi}-\boldsymbol{Z})' \\
& -\boldsymbol{ZYMY}'\boldsymbol{Z}' - (\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{YQMY}'\boldsymbol{\pi}' \\
& -\boldsymbol{V}_{X_{(n+1)}X}\left(\boldsymbol{Z} + (\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{YQ}(\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\right)' \\
& -\left(\boldsymbol{Z} + (\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{YQ}(\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\right)\boldsymbol{V}'_{X_{(n+1)}X} \\
= \; & \boldsymbol{Y}_{(n+1)}\boldsymbol{MY}'_{(n+1)} + \boldsymbol{W}_{(n+1)}^{-\frac{1}{2}} v_{(n+1,n+1)}\boldsymbol{V}_S\boldsymbol{W}_{(n+1)}^{-\frac{1}{2}} \\
& -\boldsymbol{ZYMQ}'\boldsymbol{Y}'(\boldsymbol{\pi}-\boldsymbol{Z})' - \boldsymbol{ZYMY}'\boldsymbol{Z}' - (\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{YQMY}'\boldsymbol{\pi}' \\
& -\boldsymbol{V}_{X_{(n+1)}X}\left((\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{YQ}(\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\right)' \\
& -\left((\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{YQ}(\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\right)\boldsymbol{V}'_{X_{(n+1)}X} - \boldsymbol{ZV}'_{X_{(n+1)}X} \\
= \; & \boldsymbol{Y}_{(n+1)}\boldsymbol{MY}'_{(n+1)} + \boldsymbol{W}_{(n+1)}^{-\frac{1}{2}} v_{(n+1,n+1)}\boldsymbol{V}_S\boldsymbol{W}_{(n+1)}^{-\frac{1}{2}} \\
& +\boldsymbol{ZYMQ}'\boldsymbol{Y}'\boldsymbol{Z}' - \boldsymbol{ZYMY}'\boldsymbol{Z}' - \boldsymbol{\pi YQMY}'\boldsymbol{\pi}' \\
& -\boldsymbol{V}_{X_{(n+1)}X}\left((\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{YQ}(\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\right)' \\
& -\left((\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{YQ}(\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\right)\boldsymbol{V}'_{X_{(n+1)}X} - \boldsymbol{ZV}'_{X_{(n+1)}X} \\
= \; & \boldsymbol{Y}_{(n+1)}\boldsymbol{MY}'_{(n+1)} + \boldsymbol{W}_{(n+1)}^{-\frac{1}{2}} v_{(n+1,n+1)}\boldsymbol{V}_S\boldsymbol{W}_{(n+1)}^{-\frac{1}{2}} \\
& +\boldsymbol{ZYQMY}'\boldsymbol{Z}' - \boldsymbol{ZYMY}'\boldsymbol{Z}' - \boldsymbol{Y}_{(n+1)}\boldsymbol{QMY}'_{(n+1)} \\
& -\boldsymbol{V}_{X_{(n+1)}X}\left((\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{YQ}(\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\right)' \\
& -\left((\boldsymbol{\pi}-\boldsymbol{Z})\boldsymbol{YQ}(\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{V}_{XX}^{-1}\right)\boldsymbol{V}'_{X_{(n+1)}X} - \boldsymbol{ZV}'_{X_{(n+1)}X}.
\end{aligned}
$$

Note that in the process of deriving the mean square prediction error, the definition of $\boldsymbol{Q}$ given by (5.3) is used. □

Remark 5.1. To obtain the inhomogeneous credibility predictor without conditional temporal dependence, set $\boldsymbol{V}_T = \boldsymbol{I}_n$ and $\boldsymbol{Z} = \boldsymbol{0}$. Then, Theorem 5.1 can be used directly. Theorem 5.1 can also be applied for the case when the covariance matrix of the error terms is conditional on the risk parameters $\boldsymbol{\Theta}$ (mathematically, $\mathrm{Cov}[\boldsymbol{\epsilon}_{(u)}|\boldsymbol{\Theta}] = \boldsymbol{W}_{(u)}^{-\frac{1}{2}}\boldsymbol{\Sigma}_S(\boldsymbol{\Theta})\boldsymbol{W}_{(u)}^{-\frac{1}{2}}$

and $\mathrm{E}[\boldsymbol{\Sigma}_S(\boldsymbol{\Theta})] = \boldsymbol{V}_S$).

Remark 5.2. When $n_i = n$, $i = 1, \ldots, K$, the inhomogeneous credibility predictor for the regression credibility model with general dependence among risks and temporal dependence in Wen and Wu (2011) can be obtained by setting $\boldsymbol{V}_S = \boldsymbol{I}_K$ in Theorem 5.1. However, note that Theorem 5.1 relies on the restriction that the temporal covariance matrix $\boldsymbol{V}_T$ is the same for all K entities, which is necessary to obtain a separable spatio-temporal covariance matrix. This restriction is not imposed in Wen and Wu (2011).

Chapter 6

Application to Multiple Peril Crop Insurance Data

In this chapter, the Bühlmann-Straub credibility model with general dependence structure among risks and conditional cross-sectional dependence is applied to Multiple Peril Crop Insurance (MPCI) indemnity data as an illustration to the proposed models. The main quantity of concern is the loss rate (loss per exposure) for corn of each county in the state of Iowa. Structural parameter estimates are obtained from the loss rates and future loss rates are predicted. Estimation and prediction are compared to previous credibility models. For a brief introduction to MPCI, see Section 1.1.1. More information can also be found in Schnapp et al. (2000) or on the Risk Management Agency (RMA) website[1].

6.1 Description of Data and Exploratory Analysis

Indemnity amounts and exposure information for corn in Iowa from 2000 to 2011 are collected from the U.S. Department of Agriculture Risk Management Agency website. In total, there are $K = 100$ counties in Iowa as used by RMA. The K counties are numbered as $i = 1, \ldots, 100$ and the $n = 12$ years are numbered as $u = 1, \ldots, 12$. The loss rate for county i and year u, X_{iu}, is calculated as

$$X_{iu} = \frac{\text{indemnity amount of county } i \text{ in year } u}{\text{liability of county } i \text{ in year } u},$$

[1] `http://www.rma.usda.gov`

where liability is defined as the maximum dollar amount payable. The weight used, w_{iu}, is then defined as liability of county i in year u.

The longitude and latitude of the geographic center of each county is obtained from Census 2000 U.S. Gazetteer Files on the United States Census Bureau website[2]. The distance between each pair of counties is then calculated using the Vincenty inverse formula for ellipsoids approximation (Pineda-Krch, 2010). The minimum distance between counties is found to be 13km and the maximum is 505km .

A map of annual loss rates of each county averaged over all n years,

$$\overline{\boldsymbol{X}}^{(bs)} = \left(\overline{X}_1^{(bs)}, \ldots, \overline{X}_K^{(bs)}\right)', \text{where } \overline{X}_i^{(bs)} = \widehat{\mu}_i^{(bs)} = \frac{1}{w_{i\cdot}} \sum_{u=1}^{n} w_{iu} X_{iu},$$

can be found in Figure 6.1.

A histogram of annual county loss rates for all n years, $\boldsymbol{X}$, is shown in Figure 6.2. It can be seen clearly that the distribution of the loss rates is positively skewed and perhaps heavy-tailed. The overall weighted average of loss rates, $\overline{X}^{(bs)} = \widehat{\mu}^{(bs)} = \frac{1}{w_{\cdot\cdot}} \sum_i w_{i\cdot} \overline{X}_i^{(bs)}$, is 0.0291. The overall weighted standard deviation of loss rates is 0.0461.

The Moran's I test (Moran, 1950) is performed on the vector of county average loss rates, $\overline{\boldsymbol{X}}^{(bs)}$. It is also performed on the vector of loss rates $\boldsymbol{X}_{(u)} - \overline{\boldsymbol{X}}^{(bs)}$ across all counties in period u, for $u = 1, \ldots, n$. The former is designed as a rough test for spatial dependence of the hypothetical means, $\text{Cov}[\mu(\boldsymbol{\Theta})]$, and the latter is designed as a rough test for spatial dependence of the loss rates conditional on the risk parameters, $\text{Cov}[\boldsymbol{X}_{(u)}|\boldsymbol{\Theta}]$. It is found that the p-value is less than 0.001 for all cases, indicating rejection of the null hypothesis of no spatial clustering of the values.

As additional confirmation to the Moran's I test, the Mantel test (see, for example, Scheiner and Gurevitch (2001)) is performed on $\overline{\boldsymbol{X}}^{(bs)}$ and $\boldsymbol{X}_{(u)} - \overline{\boldsymbol{X}}^{(bs)}$ for $u = 1, \ldots, n$ on 9999 permutations. The tests for the vectors of loss rates on 10 of 12 years and the vector of county average loss rates returned p-values of less than 0.001, again indicating the rejection of the null hypothesis of no relationship between the geographic distance matrix and loss rate distance matrix.

[2] `http://www.census.gov/geo/www/gazetteer/places2k.html`

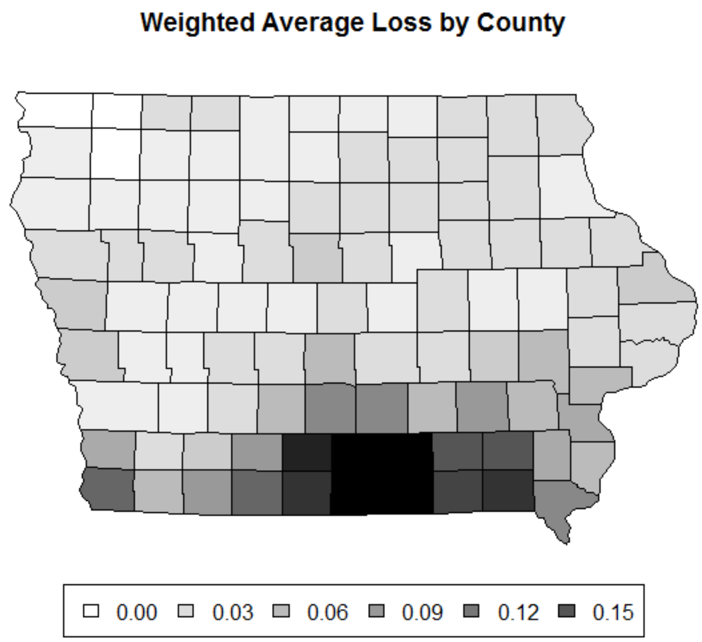

Figure 6.1: Map of weighted average annual loss rates, 2000 to 2011

6.2 Estimation of Structural Parameters

In this section and the next, three credibility models are compared:

1. The classical Bühlmann-Straub credibility model (Bühlmann and Straub, 1970),
2. The Bühlmann-Straub credibility model with general dependence structure among risks (Wen and Wu, 2011), and
3. The Bühlmann-Straub credibility model with general dependence structure among risks and conditional cross-sectional dependence.

These three models are called "classical", "reduced", and "full" for short. As a reminder, compared to the full model as presented in Chapter 4, the reduced model simply assumes that the process covariance matrix $\boldsymbol{\Sigma}(\Theta)$ (and therefore $\boldsymbol{V}$) is a diagonal matrix, and the classical model further assumes that the covariance matrix of the hypothetical mean $\boldsymbol{A}$ is a diagonal matrix.

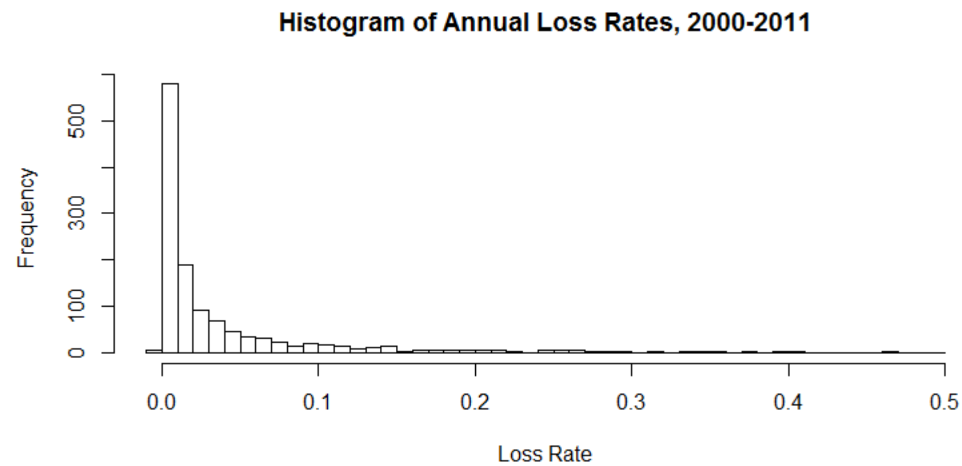

Figure 6.2: Histogram of annual loss rates, 2000 to 2011.

Assumption 3.4 that specifies the form of spatial dependence (isotropic second order stationarity) is also assumed as a requirement of the estimation procedure for the full model described in Section 4.2. The distance intervals used are

$$\mathcal{D} = \{0, (0, 37.5], (37.5, 42.5], (42.5, 47.5], \ldots, (472.5, 477.5], (477.5, 510.0]\}$$

for both $g(\cdot)$ and $f(\cdot)$. Note that with these distance intervals, each distance interval from 0 up to and including $(332.5, 337.5]$ has 30 or more pairs of counties, except for one distance interval which has 28 pairs. It is a rule of thumb to have more than 30 pairs per distance interval for estimation of spatial variograms and covariograms (Cressie, 1993). The average distance of pairs of counties is used as the midpoint of each distance interval.

Estimators of structural parameters in the classical model can be found in Bühlmann and Gisler (2005). The same estimators, $\widehat{\mu}^{(bs)}$, $\widehat{g}^{(bs)}(\cdot)$, and $\widehat{\gamma}_f^{(bs)}(\cdot)$, in Section 4.2 can be used for the reduced model, except with the following change in Theorem 4.6 that allows the estimator $\widehat{\gamma}_f^{(bs)}(\cdot)$ to be unbiased under the assumptions of the reduced model:

$$\begin{aligned} \widehat{\gamma}_g^*(d_s) &= \sum_{(i,j)\in D_s} \sqrt{w_{i\cdot}w_{j\cdot}}\frac{1}{2}\left(\frac{\sum_u p_{iu}^2/w_{iu}}{(\sum_u p_{iu})^2}\widehat{v}_{ii}^{(bs)} + \frac{\sum_u p_{ju}^2/w_{ju}}{(\sum_u p_{ju})^2}\widehat{v}_{jj}^{(bs)}\right. \\ &\qquad \left. -2\delta_{ij}\frac{\sum_u p_{iu}p_{ju}/\sqrt{w_{iu}w_{ju}}}{\sum_u p_{iu}\sum_v p_{jv}}\widehat{v}_{ij}^{(bs)}\right). \end{aligned} \tag{6.1}$$

Naturally, the modified estimator $\widehat{\gamma}_f^{(bs)}(\cdot)$ for the reduced model produces a smaller or equal result than the corresponding estimator for the full model if the expected process covariance estimators $\widehat{v}_{ij}^{(bs)}$ are non-negative. This can cause overestimation of $f(\cdot)$ when the modified estimator is used and the full model is the true model.

The estimators of the overall means are the same for all three models. As mentioned before, $\widehat{\mu}^{(bs)} = 0.0291$, representing a weighted average loss of 29 per every 1000 dollar insured.

The expected process covariogram estimator $\widehat{g}^{(bs)}(\cdot)$ gives estimates that are shown in Figure 6.3(a). These point estimates of the covariogram are nicely shaped indicating falling spatial correlation as distance increases. To obtain a smooth estimate of the covariogram and guarantee a positive-definite estimate of $\boldsymbol{V}$, least squares fitting of parametric covariograms is used as recommended by Cressie (1993) for its non-parametric approach that is consistent with credibility theory. However, generalized least squares is not used because of the lack of knowledge of the second moments of the covariogram estimator; to obtain second moments of the covariogram estimator, it will involve and require additional assumptions on the fourth moments of the loss rates. Therefore, a simple weighted least squares fit using the number of observations as weights is performed instead. Note that only point estimates up to a distance of 257.5 are used in the least squares fit, since it is recommended in spatial statistics to only use estimates for distances smaller than half of the maximum distance between pairs of regions due to variability issues with estimates for larger distances (Cressie, 1993). Note that when only half the distance intervals are used, the point estimates do not reach a covariance of zero yet. As a result, the exponential covariogram with a range of infinity (see the Appendix) is chosen.

The results of the fitted exponential covariogram are shown in Table 6.1. The nugget parameter a is fixed to zero because without fixing the nugget, its estimate is negative. The sill parameter s indicates that an estimate of the expected process variance is 77685. Finally, the effective range parameter r indicates that at around a distance of 600, the process correlation between counties drop to around 95%. This estimate of the effective range is a bit different from what the point estimates show. An estimate of 400 would be more suitable if looking at the whole graph of point estimates in Figure 6.3(a), but since points estimates beyond half of maximum distances should be ignored, in the next section, prediction is performed with the covariogram as fitted by least squares.

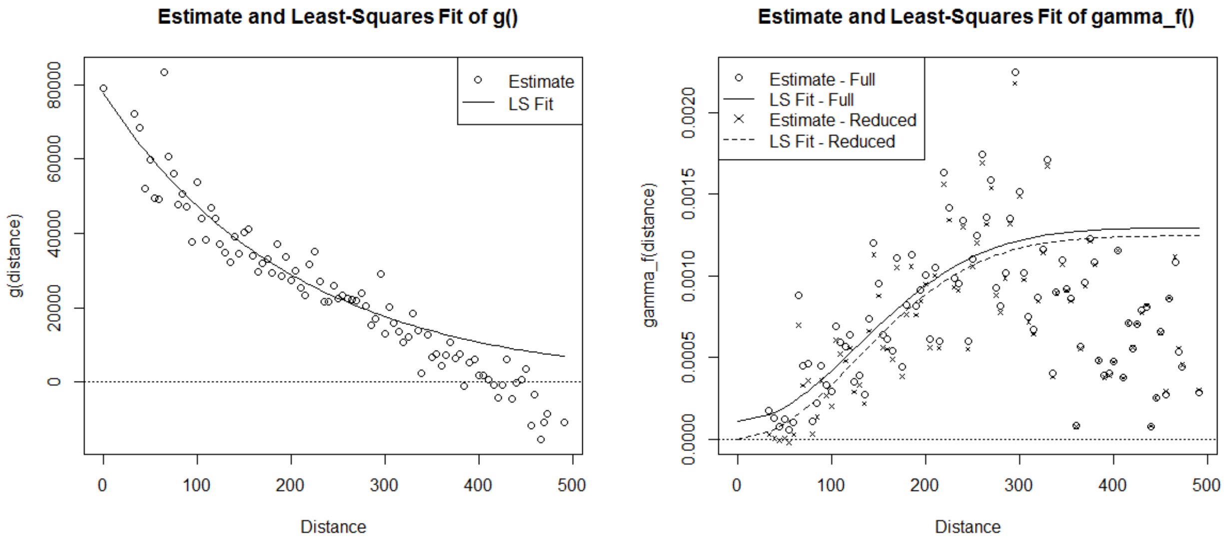

Figure 6.3: (a) Estimated expected process covariogram, $g(\cdot)$, and its exponential covariogram fit. (b) Estimated variogram of hypothetical means, $f(\cdot)$, fitted with the gaussian variogram with free nugget and the gaussian variogram with a fixed small positive nugget for the full model and reduced model, respectively.

The estimated expected process variance for both the reduced model and the classical model is 78976, which is the same as the point estimate of the expected process variance for the full model since the estimator used is the same.

Estimates of the variogram of hypothetical means computed with $\widehat{\gamma}_f(\cdot)$ are shown in Figure 6.3(b) for both full and reduced models. For most distance intervals, the estimates for the reduced model is smaller than the estimate for the full model. This is due to the modification made to $\widehat{\gamma}_f(\cdot)$ to restore unbiasedness under the reduced model, as mentioned earlier.

It can be seen in Figure 6.3(b) that point estimates of the variograms exhibit a quadratic pattern. Typical empirical variograms are thought to plateau at the sill beyond a certain distance when spatial correlation drops to zero. However, there is a dip in the point estimates beyond roughly two thirds of the maximum distance between counties. This can due to unreliable estimates at larger distances, edge effects, or an unexplained phenomenon. If the rule of thumb is followed and any point estimates for distances larger than half the maximum distance are ignored, it is not known whether the point estimates reach a sill. An

	Full	Full	Reduced
Fitted	$g(\cdot)$	$\gamma_f(\cdot)$	$\gamma_f(\cdot)$
Model	Exponential	Gaussian	Gaussian
s	77685	0.00130	0.00125
a	0^*	0.00011	$1 \times 10^{-8*}$
r	607.6	315.2	311.4

Table 6.1: Parameter estimates of least squares fits of the expected process covariogram, $g(\cdot)$, and the variogram of the hypothetic means, $\gamma_f(\cdot)$. The superscript " $*$ " indicates a parameter that is fixed and not estimated by least squares.

ever-increasing trend may be indicative of hidden spatial trend. Further analyses involving a wider geographical area and longer observation period may perhaps reveal the true cause for the shape of the variogram.

The parametric gaussian variogram (see the Appendix) is fitted to point estimates of both the full and the reduced model. The fitted variograms are shown in Figure 6.3(b) and the parameter estimates are shown in Table 6.1. The nugget parameter a for the reduced model is fixed due to negative fitted nuggets. Since the Gaussian variograms often create numerical errors with a zero nugget (Chilès and Delfiner, 1999), a small positive nugget is chosen instead.

The fitted parameters of the gaussian variograms are similar for the full and reduced models. Both suggest a sill s of roughly 0.00130 and an effective range of roughly 315. The nugget a for the full model, however, is roughly $\frac{1}{12}$ of the sill.

The estimated variance of the hypothetical mean for the classical model is 0.000421. This is $\frac{1}{3}$ the size of the corresponding estimate of roughly 0.00130 for the other two models. If the full model or the reduced model were true, it can be shown that the estimator of the variance of the hypothetical mean has a non-positive bias if the covariance between different counties is non-negative. This may be one cause of the discrepancy between the classical estimator and the proposed estimator. Another possible cause of the discrepancy is from the unapparent sill of the point estimates in the full and reduced model.

A comparison of the fitted covariogram of the hypothetical means and the fitted expected process covariogram is shown in Figure 6.4. Note that the actual expected spatial covariance of the conditional losses depends on the weights which changes through time and counties, so the expected process covariogram, averaged over time periods and counties, is shown. It can be seen that the spatial covariance of conditional losses is larger than the covariance

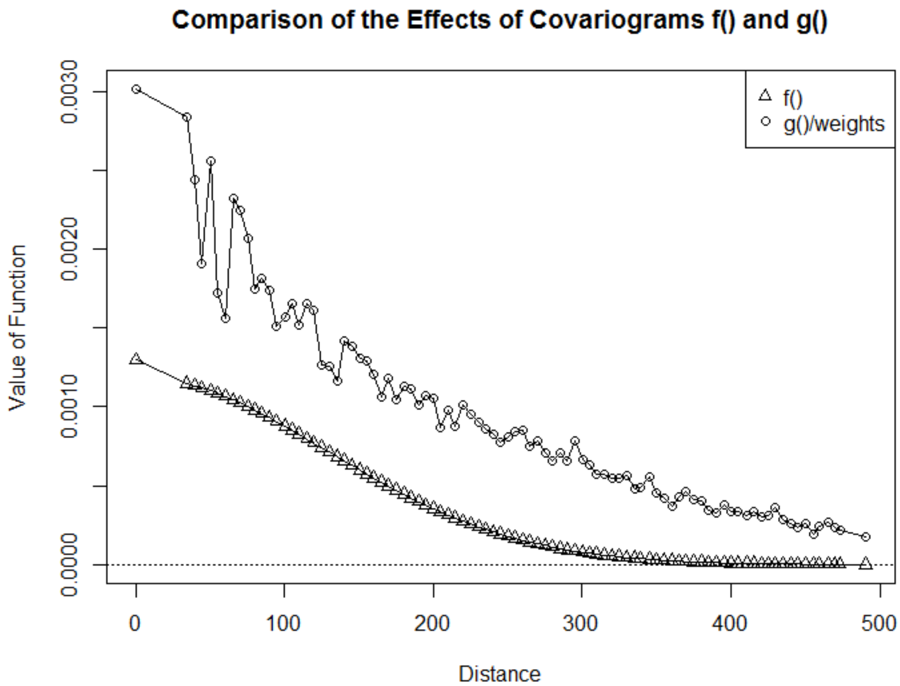

Figure 6.4: Comparison of the covariogram of the hypothetical mean, $f(\cdot)$ and expected process covariogram with weights applied, $\frac{g(\cdot)}{\sqrt{w_{iu}w_{jv}}}$.

of the hypothetical means indicating that more variation and covariation come from yearly fluctuations than the time-independent unknown hypothetical means.

6.3 Credibility Prediction

In this section, the estimates from the previous section is used to obtain predictions for loss rates in a future period, $\boldsymbol{X}_{(n+1)}$. Credibility predictors investigated are

1. Classical Model Inhomogeneous Credibility Premium (Bühlmann and Straub, 1970),
2. Reduced Model Inhomogeneous Credibility Premium (Wen and Wu, 2011),
3. Reduced Model Homogeneous Credibility Premium (Wen and Wu, 2011),
4. Full Model Inhomogeneous Credibility Premium, and

5. Full Model Homogeneous Credibility Premium.

The simple weighted average $\overline{X}_i^{(bs)}$, as a predictor, is used as a reference for comparison. Constant unconditional mean (Assumption 3.3) is assumed to facilitate equal comparisons of the five credibility predictors.

Histograms of the predicted values of the 100 counties are shown in Figure 6.5 for the five credibility predictors. The histogram of weighted averages of loss rates is also shown as a benchmark. It can be seen that the predictions of the full model and the classical model have more compact distributions around their center than the distribution of the weighted averages. The predictions of the reduced model has only a slightly more compact distribution than the averages.

The actual predicted values are shown in Figure 6.6. All five credibility predictors produce predicted values that are less extreme than the weighted average, as seen previously in the histograms. It can also be seen that the predicted values produced by the full and reduced models are much more spatially smoothed than the predicted values produced by classical model and the weighted average. For these two models with spatial dependence, nearby counties are predicted to have similar loss rates.

Comparing the full model to the reduced model, the former model smooths the extremes more evenly. There are three main differences between the two models.

- The estimates of $\boldsymbol{A}$ are smaller in the reduced model than in the full model.
- Point estimates are used directly for $\boldsymbol{V}$ in the reduced model and the fitted covariogram is used instead for $\boldsymbol{V}$ in the full model.
- The non-diagonal elements of $\boldsymbol{V}$ are zero in the reduced model and non-zero in the full model.

By performing each of these changes one by one in sequence, it is found that the smoothing of the extremes in the full model can be attributed mainly to the non-zero non-diagonal elements of $\boldsymbol{V}$.

Intuitively, the predictions should transition in sequence from weighted averages to classical model, reduced model, and finally full model or another similar sequence. However, it appears that the classical model is out of line; the classical model does not seem to transition smoothly to the reduced model and then to the full model. The estimates used for the process covariance and unconditional mean are similar among all three models. The

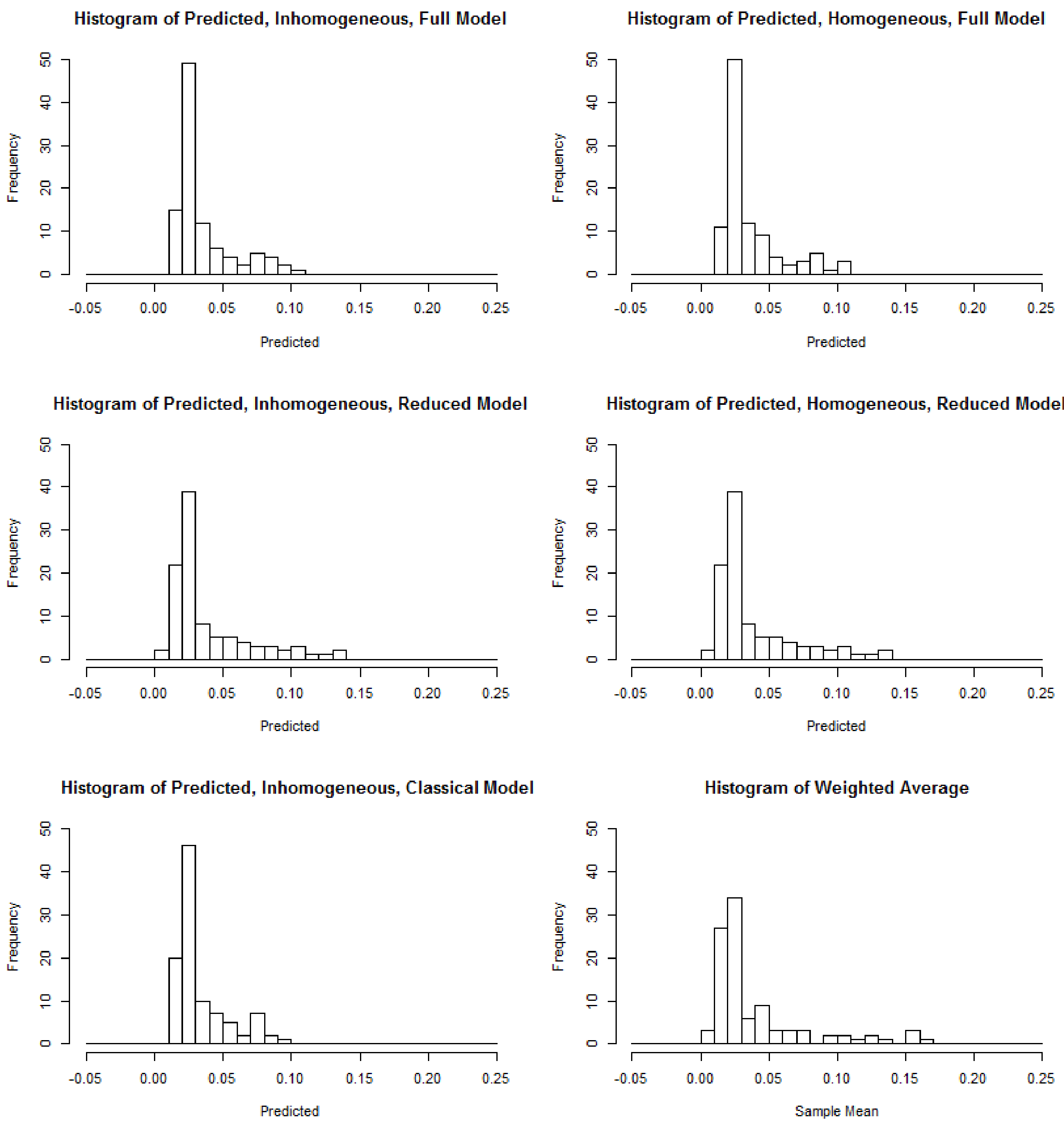

Figure 6.5: Histograms of predicted loss rates of various credibility predictors and simple weighted average.

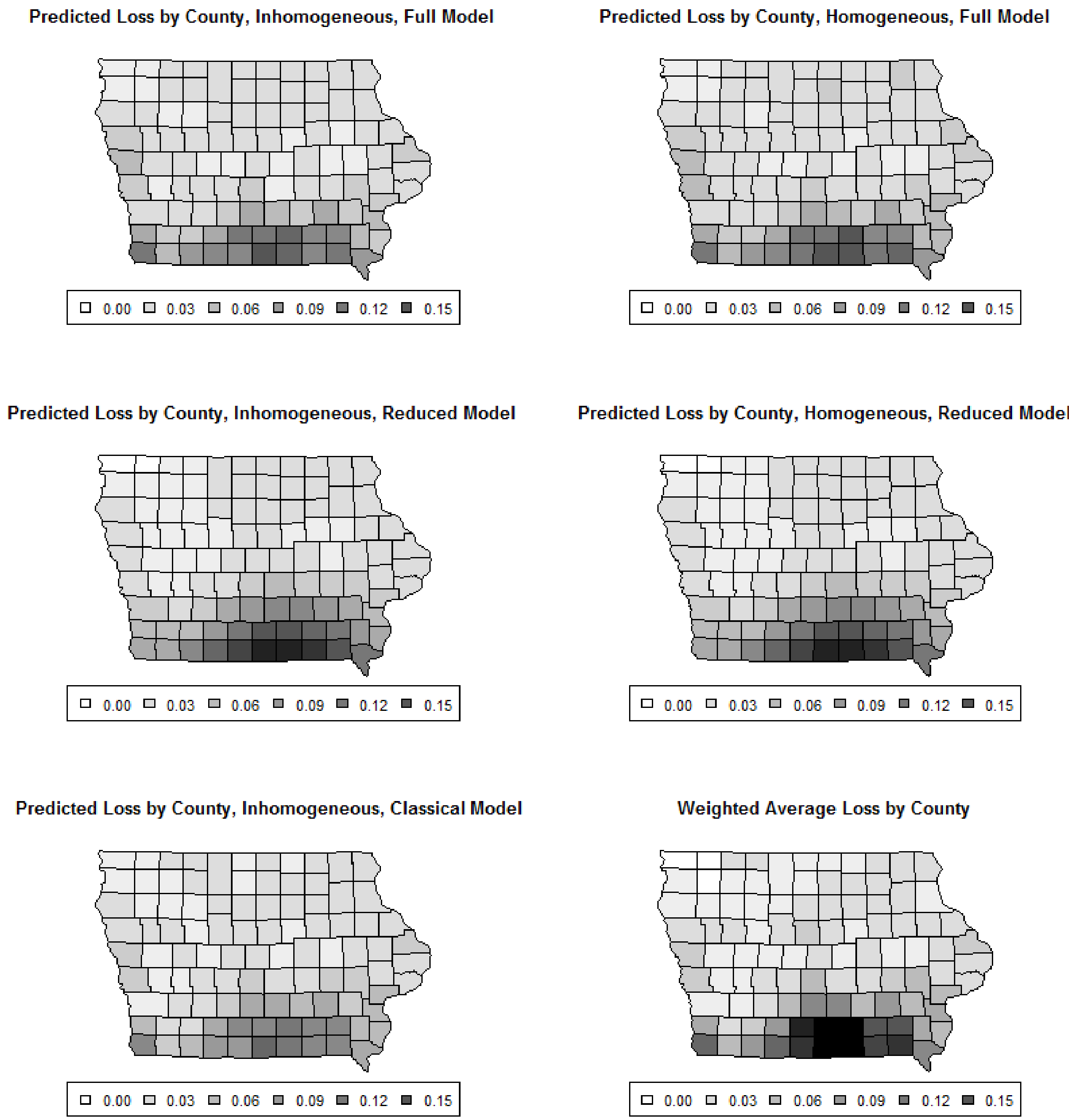

Figure 6.6: Predicted loss rates of various credibility predictors and simple weighted average.

only source of difference is the estimate of the variance of the hypothetical mean, and this is found to be the source of the discrepancy. As mentioned previously, the estimate of the variance of the hypothetical mean is much smaller than the corresponding estimates of the reduced and full model. This leads to less credibility assigned to the observed county averages and more credibility assigned to the overall mean, and as a result, predictions of the classical model are closer to the overall mean and hence out of line with the other models.

The inhomogeneous and homogeneous credibility estimators yield very similar predictions. Predictions for the full model increased by 2% to 12% when using homogeneous credibility premiums, while for the reduced model most changes are within -3% to 6%. Note that the difference between the inhomogeneous credibility premium and homogeneous credibility premium in this context is

$$\begin{aligned}
\widehat{\boldsymbol{X}}^{(bs)}_{(n+1)} - \widehat{\boldsymbol{X}}^{(bs)hom}_{(n+1)} &= \sum_{i=1}^{K} \left[\boldsymbol{Z}_i \overline{\boldsymbol{X}}_i^{(bs)} + \left(\frac{1}{K} \boldsymbol{I}_K - \boldsymbol{Z}_i \right) \mu \mathbf{1}_K \right] \\
&\quad - \sum_{i=1}^{K} \left[\boldsymbol{Z}_i \overline{\boldsymbol{X}}_i^{(bs)} + \left(\frac{1}{K} \boldsymbol{I}_K - \boldsymbol{Z}_i \right) \widehat{\mu}^{(bs)hom} \mathbf{1}_K \right] \\
&= \sum_{i=1}^{K} \left(\frac{1}{K} \boldsymbol{I}_K - \boldsymbol{Z}_i \right) \left(\mu - \widehat{\mu}^{(bs)hom} \right) \mathbf{1}_K .
\end{aligned}$$

The estimate of overall mean μ used in the inhomogeneous credibility estimator is 0.0291 for both the full and the reduced model, as mentioned before. On the other hand, the homogeneous credibility estimate of the overall mean μ is 0.0482 for the full model and 0.0515 for the reduced model. Hence, it can be seen that although there is over 50% change in the estimated mean used, the effects on the resulting predictions are small, suggesting large credibility assigned to the experience of each county rather than to the overall experience of all K counties.

Chapter 7

Simulation Study of Estimation and Prediction

In Chapter 6, an application of the models and estimators proposed in this project is reviewed. In this chapter, estimates from the application are used as a guide to create a simulation model. The simulation model is then used to compare the performance of estimators and predictors.

7.1 The Simulation Model

The model of loss rates used is

$$X_{iu} = \mu + \Theta_i + \epsilon_{iu}. \tag{7.1}$$

The vector of risk parameters $\boldsymbol{\Theta}$ is assumed to follow a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{A}$. The error vectors $\boldsymbol{\epsilon}_{(u)} = (\epsilon_{1u}, \ldots, \epsilon_{Ku})'$, $u = 1, \cdots, n+1$ are independent and each follows an identical multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{W}_{(u)}^{-\frac{1}{2}} \boldsymbol{V} \boldsymbol{W}_{(u)}^{-\frac{1}{2}}$. In addition, $\boldsymbol{\Theta}$ and each $\boldsymbol{\epsilon}_{(u)}$ are assumed to be independent. With these assumptions, it is easy to see that $\boldsymbol{\mu} = \mu \mathbf{1}_K$, $\mathrm{Cov}[\boldsymbol{\mu}(\boldsymbol{\Theta})] = \boldsymbol{A}$, and $\mathrm{E}[\mathrm{Cov}[\boldsymbol{X}_{(u)} | \boldsymbol{\Theta}]] = \mathrm{E}[\mathrm{Cov}[\boldsymbol{\epsilon}_{(u)}]] = \boldsymbol{W}_{(u)}^{-\frac{1}{2}} \boldsymbol{V} \boldsymbol{W}_{(u)}^{-\frac{1}{2}}$.

The numerical values of the parameters of the simulation model are specified to be as close to the estimates obtained in Chapter 6 as possible. This is because this simulation study is intended to be simple yet realistic. However, it should be cautioned that the MPCI loss rates in Chapter 6 are actually positively skewed which is a feature of the data that the multivariate normal distribution does not capture.

First, the number of counties K is set to 100 and the number of observation periods per sample is $n = 12$. The actual weight matrix of the MPCI loss rate data is used as the weight matrix in each simulation sample. The actual distance matrix for Iowa counties is also retained in the simulation. The overall mean μ is set to be 0.0291.

The covariogram fits for $f(\cdot)$ and $g(\cdot)$ are used as the true covariogram of the hypothetical mean and the expected process covariogram in the simulation model. However, the covariogram fits are not retained in their original form; the covariogram fits $f(\cdot)$ and $g(\cdot)$ are converted into step-functions in the simulation model. Precisely, the value of the true $f(\cdot)$ of the simulation model within the distance interval d_s is set constant to its point estimate $\widehat{f}(d_s)$ obtained in Chapter 6. This modification is made to remove any excess variability introduced due to non-constant covariogram values and to satisfy the conditions of Theorem 4.6 when examining the estimator $\widehat{\gamma}_f(\cdot)$. The same can be said for the conversion of the covariogram fit of $g(\cdot)$ into a step-function for use as the true expected process covariogram of the simulation model.

A total of 1000 samples of loss rates for the 100 counties and 12 periods are generated. The number of samples is relatively small and this is a limitation of this simulation study.

7.2 Performance of Estimators

For the estimator of the overall mean, $\widehat{\mu}$, it is found that it has a standard deviation of 0.022 and a high coefficient of variation at around 75% under the simulation model.

The standard deviation and the coefficient of variation of the estimator of the expected process covariogram, $\widehat{g}(\cdot)$, for different distance intervals are shown in Figure 7.1. It can be seen that beyond a distance of around 400, the standard deviation of the estimator rises. This is presumably due to the low number of pairs of regions that are in each distance interval beyond 400, since fewer pairs leads to higher variability. Hence, ignoring number of pairs as distance increases, it can be said that the standard deviation of the estimator stays constant with respect to distance. On the other hand, the coefficient of variation of the estimator of the expected process covariogram rises steadily as distance increases. In particular, the coefficient of variation hits 50% at about half the maximum distance. This perhaps is a confirmation of the rule of thumb that suggests using only covariogram estimates up to half the maximum distance between counties.

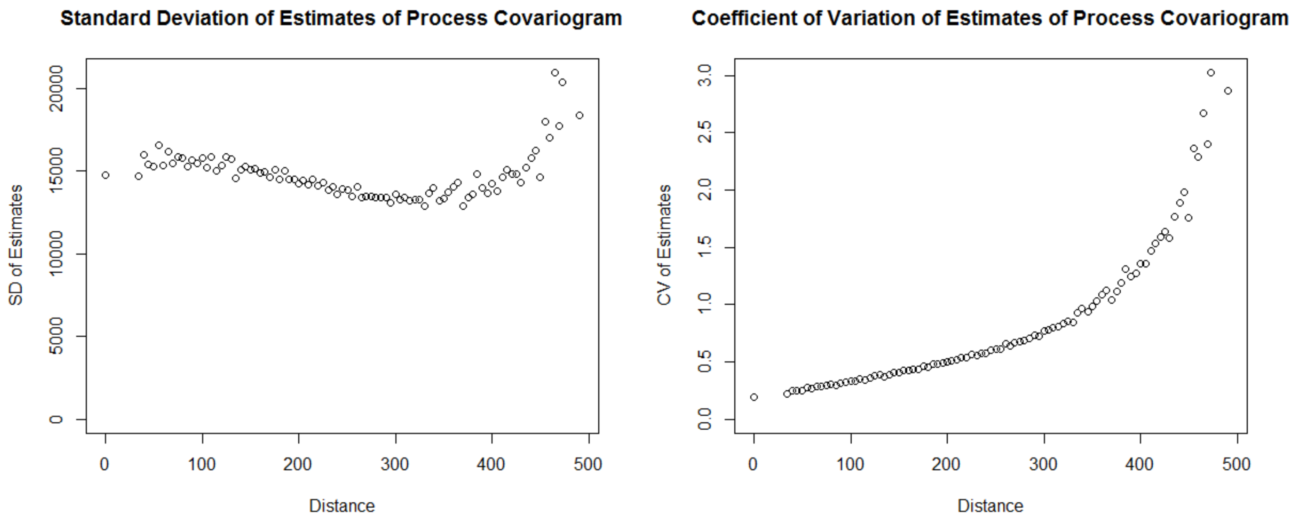

Figure 7.1: (a) The standard deviation of estimator of the expected process covariogram, $\widehat{g}(\cdot)$, at different distances. (b) The coefficient of variation of the estimator at different distances.

The standard deviation and the coefficient of variation of the estimator of the variogram of hypothetical means, $\widehat{\gamma}_f(\cdot)$, is shown in Figure 7.2. Unlike the estimator of the expected process covariogram, the standard deviation of the estimator of the variogram of hypothetical means rises linearly starting from 0. The coefficient of variation also rises steadily with increasing distance.

The estimator of the variogram of hypothetical means modified for the reduced model (6.1) has a negative bias in the simulation. This is expected, since under the assumptions of the full model, this modified estimator can be easily shown to be negatively biased when counties are positively correlated. The modified estimator would only be unbiased if the reduced model were the true model. The same can be said for the estimator of the variance of the hypothetical mean for the classical model.

One way to reduce the large standard deviations observed in this simulation study is to increase the number of periods. However, the gain in standard deviation of a larger number of periods may be limited for both estimators of the overall mean μ and variogram of the hypothetical mean $\gamma_f(\cdot)$. In the case of μ, having more periods only reveal the actual value of $\mathbf{\Theta}$ but the positive correlation among $\mathbf{\Theta}$ may make μ more difficult to estimate. In the case of $\gamma_f(\cdot)$, simply increasing the number of periods of data does not increase replication

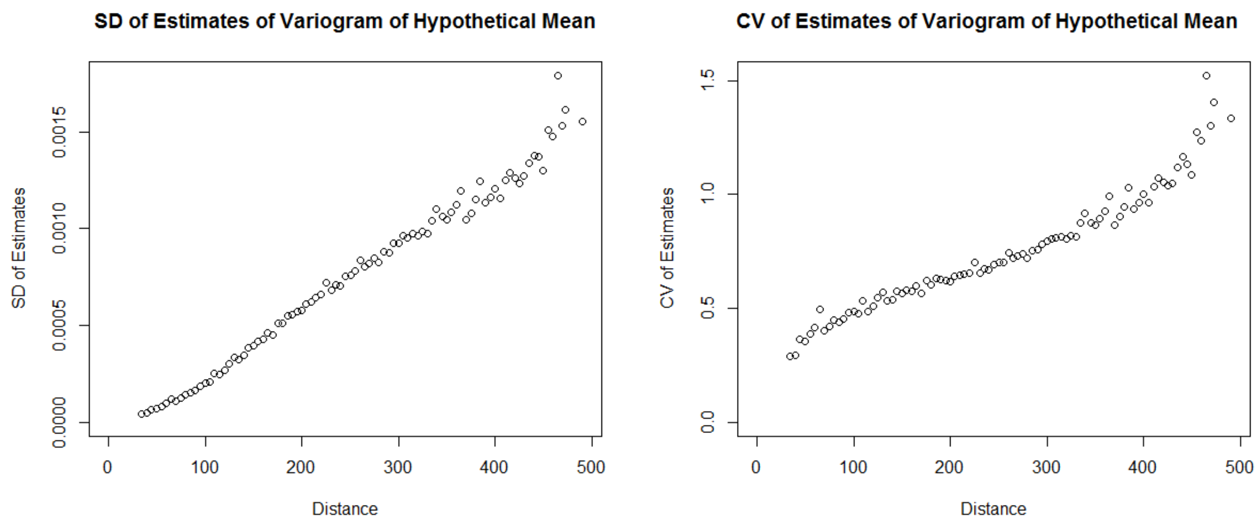

Figure 7.2: (a) The standard deviation of estimator of variogram of hypothetical means, $\widehat{\gamma}_f(\cdot)$, at different distances. (b) The coefficient of variation of the estimator at different distances.

for Θ. In an alternative simulation study with $n = 120$, it is indeed observed that increasing the number of periods observed does not lead to substantially smaller standard deviations with $\widehat{\mu}$ and $\widehat{\gamma}_f(\cdot)$. To improve the variability of these two estimators, it may be more helpful to have more counties by creating finer partitions or adding counties from nearby states.

7.3 Performance of Predictors

The same prediction procedure used in Chapter 6 is followed for all predictors. To obtain valid simulation results for predictions, it is necessary to check that the estimates of $\boldsymbol{A}$ and $\boldsymbol{V}$ in each simulated sample are positive-definite. Also, least-squares fitting of variograms and covariograms can generate negative or infinite parameter estimates. All samples for which any one of these problems occurs are discarded when evaluating predictions. The total number of samples after discarding those with issues is 942 instead of 1000.

In addition to testing the full estimation and prediction procedure which captures errors from both estimation and prediction, predictions made with the true values of the structural parameters μ, $\boldsymbol{A}$, and $\boldsymbol{V}$ are also tested to isolate the performance of predictors.

The main measure of performance used here is the mean square error (MSE) in the

Lag	End Point	Midpoint	K^*	F1	F2	F1*	F2*
1	50	38.3	6	0.000088	0.000086	0.000065	0.000068
2	100	79.6	15	0.000123	0.000121	0.000089	0.000091
3	150	127.3	27	0.000152	0.000150	0.000110	0.000113
4	200	172.4	30	0.000169	0.000166	0.000125	0.000128
5	300	227.5	22	0.000150	0.000147	0.000114	0.000118
Lag	R1	R2	R1*	R2*	C1	C1*	Wtd Avg
1	0.000190	0.000189	0.000118	0.000119	0.000107	0.000096	0.000104
2	0.000225	0.000225	0.000159	0.000159	0.000156	0.000142	0.000161
3	0.000250	0.000250	0.000186	0.000187	0.000187	0.000167	0.000205
4	0.000263	0.000262	0.000200	0.000201	0.000205	0.000177	0.000216
5	0.000231	0.000231	0.000161	0.000164	0.000170	0.000142	0.000162

Table 7.1: Simulated mean square error of various predictors. "F", "R", and "C" are short for full, reduced, and classical, respectively. The postfix "1" indicates a inhomogeneous predictor, and the postfix "2" indicates a homogeneous predictor. The superscript "*" indicates prediction that is made with true structural parameter values that is not estimated from the simulation sample.

estimation of the hypothetical mean. For a group of K^* counties, the simulated mean square error of the predictor $g(\boldsymbol{X})$ in the estimation of the hypothetical mean, averaged over the K^* counties, is calculated as

$$\mathrm{MSE}(g(\boldsymbol{X})) = \frac{1}{K^*}\sum_{i=1}^{K^*}\left\{\sum_{l=1}^{L}\left(\left(\mu+\Theta_i^{(l)}\right)-g_i^{(l)}(\boldsymbol{X})\right)^2\right\},$$

where $\mu+\Theta_i^{(l)}$ is the hypothetical mean and $g_i^{(l)}(\boldsymbol{X})$ is the value of the predictor of the lth simulated sample for the ith county. It is well-known that for a estimator linear in $\boldsymbol{X}$, minimizing the mean square prediction error (2.1) also minimizes the mean square error for the hypothetical means $\boldsymbol{\mu}(\boldsymbol{\Theta})$. In this simulation study, it is decided to evaluate the predictions based on mean square error instead of mean square prediction error because the former requires fewer simulated values and therefore less simulation error.

Table 7.1 shows the simulated mean square error for the 5 predictors considered. The 100 counties are grouped in to 5 different groups based on their distance from the center of Iowa. The five groups are $(0,50]$, $(50,100]$, $(100,150]$, $(150,200]$ and $(200,300]$. The mean square error is calculated separately for each group to check the presence of edge effects on prediction.

The maximum efficiency in prediction, in a sense, is achieved when true structural

Lag	F1	F2	R1	R2	C1
1	34%	27%	61%	59%	11%
2	38%	33%	41%	41%	10%
3	39%	33%	34%	34%	12%
4	36%	30%	31%	30%	16%
5	31%	25%	44%	41%	20%

Table 7.2: Percentage difference in simulated MSE when estimates of structural parameters are used over true values. The postfixes "1" and "2" indicate inhomogeneous and homogeneous predictors, respectively. A smaller percentage difference indicates a more favourable result.

parameter values are used. If structural parameters are unknown and hence have to be estimated, the efficiency of the predictor drops. Table 7.2 shows the percentage difference[1] in mean square error when structural parameters have to be estimated from the simulation sample. However, it should be noted that this interpretation of maximum efficiency is less applicable when the model is misspecified, as in the case of the predictions of the reduced model and the classical model. The true model in the simulation is the full model, but the underlying models are actually misspecified in the estimation and prediction of the results of the reduced model and classical model. In some cases, if estimation and prediction are robust, the estimated values and prediction can adapt to the true model and perform better than when the true parameters are actually used. Thus, the interpretation of maximum efficiency is not exact in cases of model misspecification.

In Table 7.2, it can be seen that credibility predictors of the reduced model have the highest percentage increases in mean square error when structural parameters are estimated, and predictors of the full model have the second highest increases. The inhomogeneous credibility predictor of the classical model has the least percentage increase in mean square error when structural parameters are estimated. Although the estimator of the variance of the hypothetical mean, a, is substantially unbiased, one possible reason for the smallest increase in mean square error of the classical model predictor may be the parsimonious number of estimates needed for this model. More estimated parameters can lead to higher variability. Another possible reason is that for each county, the predictor of the classical model only attempts to interpolate between the weighted average of that particular county

[1] Defined as $(\mathrm{MSE} - \mathrm{MSE}^*) \div \mathrm{MSE}^* \times 100\%$, where MSE denotes the simulated mean square error when estimated parameter values are used and MSE^* denotes the simulated mean square error when true parameter values are known.

Lag	F1	F2	F1*	F2*	R1	R2	R1*	R2*	C1	C1*
1	-16%	-17%	-37%	-35%	83%	82%	15%	64%	3%	-7%
2	-24%	-25%	-45%	-43%	39%	39%	-1%	-2%	-3%	-12%
3	-26%	-27%	-46%	-45%	22%	22%	-9%	-9%	-9%	-18%
4	-22%	-23%	-42%	-41%	21%	21%	-7%	-7%	-5%	-18%
5	-7%	-9%	-29%	-27%	43%	42%	-0%	1%	5%	-12%

Table 7.3: Percentage difference in simulated MSE when credibility predictors are used over the simple weighted average. The postfixes "1" and "2" indicate inhomogeneous and homogeneous predictors, respectively. The superscript "*" indicates use of true structural parameter values. A smaller percentage difference indicates a more favourable result.

and the overall mean, while the predictors of the full and reduced models actually use the weighted average of other counties in their interpolation, too. This way, estimation errors of the full and reduced model could be amplified. The overall simplicity of the classical model allows the mean square error to rise the least out of the three models when estimates are used.

The percentage difference[2] between each credibility predictor and the weighted average is shown in Table 7.3.

When true values of structural parameters are used, predictors of the full model show the largest decreases in mean square error when compared to the simple weighted average. This is expected since the model is specified correctly, and the predictors derived under the model have the smallest mean square error for any predictors of the same class. It is interesting to see that even when the true value of structural parameters are known, the predictors of the reduced model is only as good as the simple weighted average, while the predictor of the classical model is better than the simple weighted average. A possible reason for this is that when producing a prediction for a particular county, the predictors of the incorrectly specified reduced model may incorrectly use the experience of other counties, while the predictor of the classical model avoids this by using the experience of the particular county exclusively.

When structural parameters are unknown and have to be estimated, predictors of the full model once again show the largest decrease in mean square error. The predictor of the classical model is only as good as the sample weighted average in terms of mean square error, while predictors of the reduced model are much worse than the simple weighted average.

[2]Defined as $\left(\text{MSE}\Big(g(\boldsymbol{X})\Big) - \text{MSE}\left(\overline{\boldsymbol{X}}^{(bs)}\right)\right) \div \text{MSE}\left(\overline{\boldsymbol{X}}^{(bs)}\right) \times 100\%$.

This large increase in mean square error for predictors of the reduced model is found to be caused by the negative bias in the modified estimator of variogram of the hypothetical mean $\widehat{\gamma}_f(\cdot)$.

In Section 6.3, it is shown that the inhomogeneous and homogeneous credibility predictors produce very similar results with the MPCI data. In this simulation study which is specified very close to the MPCI data, it is once again confirmed that the inhomogeneous and homogeneous predictors perform very similarly, as seen in Table 7.1. The homogeneous predictors, which avoid estimation of the overall mean μ, attains only a decrease in mean square error of 2% at the best.

In a separate run, the same simulation study is performed with $n = 120$. It is found that the 10 times increase in periods observed decreases the mean square errors of the full, reduced, and classical predictors by 9, 2, and 10 times respectively. In comparison, the mean square error of the weighted averages decreases by exactly 10 times when the number of periods observed increases tenfold. The poor improvement in mean square error of the reduced model is again due to biased estimation of the variogram. Note that with $n = 120$, it is possible to observe the convergence of all predictors to the weighted averages as noted in Remark 3.1. In fact, the improvement in mean square error that is gained by using each of the credibility predictors in lieu of the weighted averages shrinks when the number of periods observed is increased ten times to $n = 120$.

As a final note, the actual mean square error consists of a variance-covariance term and a squared bias term:

$$\begin{aligned}
&\mathrm{E}\left[(\boldsymbol{\mu}(\boldsymbol{\Theta}) - g(\boldsymbol{X}))\,(\boldsymbol{\mu}(\boldsymbol{\Theta}) - g(\boldsymbol{X}))'\right] \\
&= \mathrm{Cov}[\boldsymbol{\mu}(\boldsymbol{\Theta}) - g(\boldsymbol{X})] + (\mathrm{E}[\boldsymbol{\mu}(\boldsymbol{\Theta})] - \mathrm{E}[g(\boldsymbol{X})])\,(\mathrm{E}[\boldsymbol{\mu}(\boldsymbol{\Theta})] - \mathrm{E}[g(\boldsymbol{X})])'\,.
\end{aligned}$$

The bias of each predictor is 0 theoretically, and this is demonstrated from results of the simulation study. Hence, the mean square errors studied here consist of the variance-covariance term only.

Chapter 8

Conclusion

In this project, credibility models with general dependence among risks and conditional dependence are studied. By extending the results of Wen and Wu (2011) and Schnapp et al. (2000), dependence among losses is introduced through dependence among risk parameters and cross-sectional conditional dependence. Inhomogeneous and homogeneous credibility predictors are derived using orthogonal projections in Hilbert space. Results obtained for the basic Bühlmann type model with dependence of both types are further generalized to Bühlmann-Straub and regression credibility model formulations. Also, to allow for practical use of these credibility predictors, non-parametric estimation methods for structural parameters in applications with spatial dependence are proposed.

The prediction and estimation methods proposed are tested with Multiple Peril Crop Insurance indemnities data. Various considerations for the use of the proposed estimation methods are described. Results are compared for the proposed model and two other special cases. The three models are found to yield different prediction results.

Finally, a small-scale simulation study using multivariate normal distributions is presented. In the first part of the study, estimators for structural parameters are compared. The estimators seem to have large coefficients of variation. Also, it is confirmed that the variability in estimates of variograms and covariograms may depend on distance. In the second part of the simulation study, mean square errors of credibility premiums of various Bühlmann-Straub-type models are compared. It is found that the correct specification of the underlying model is very important. If the true model is not specified correctly, credibility premiums can produce predictions that even yield higher mean square errors than the sample mean. Overall, under the circumstances of the simulation study, the proposed

Bühlmann-Straub model gives the best results.

This project can be extended in many ways. First, an estimation method for the structural parameters of the regression model with spatio-temporal dependence can be developed and tested. Second, more types of dependence through risk parameters and latent variables can be considered in addition to conditional spatio-temporal covariance. Third, further research can be done to relax the assumption of a separable spatio-temporal covariance matrix in the regression credibility model proposed. Lastly, other approaches to model spatial or spatio-temporal data such as Markov random fields can be explored to provide alternative methods for estimation.

Appendix.

Parametric Spatial Variogram Models

Here, definitions of parametric variograms and their interpretation are given. More information can be found in Cressie (1993) and Chilès and Delfiner (1999).

The exponential variogram model has the variogram

$$\gamma(h; a, s, r) = \begin{cases} 0, & h = 0 \\ a + (s - a)\left(1 - e^{-3h/r}\right), & h > 0 \end{cases}$$

and covariogram

$$C(h; a, s, r) = \begin{cases} s, & h = 0 \\ (s - a)e^{-3h/r}, & h > 0 \end{cases}.$$

The parameter a is called the nugget and the parameter s is called the sill. For a spatial process $Y(\cdot)$, sill is simply the variance of $Y(\cdot)$. The nugget is a discontinuity in the variogram and the corresponding covariogram at 0. This discontinuity can account for the variance of an additional white-noise process on top of a spatial process with a continuous variogram. It can also account for the spatial covariance that works on a smaller scale than the geographical scale considered and hence is unobserved. Finally, the parameter r is called the effective range. This is the distance at which the covariance function falls to roughly 5% of the maximum covariance $s - a$.

Note that the exponential variogram has the property that correlation never falls to zero as distance increases. Correlation only becomes arbitrarily closer to zero.

The gaussian variogram model has the variogram

$$\gamma(h;a,s,r) = \begin{cases} 0, & h=0 \\ a + (s-a)\left(1 - e^{-3h^2/r^2}\right), & h>0 \end{cases}$$

and covariogram

$$C(h;a,s,r) = \begin{cases} s, & h=0 \\ (s-a)e^{-3h^2/r^2}, & h>0 \end{cases}.$$

All parameters are interpreted the same way as the parameters in the exponential variogram model. Like the exponential variogram model, the correlation of the gaussian variogram model never falls to zero.

Bibliography

Banerjee, S., Carlin, B., and Gelfand, A. (2003). *Hierarchical modeling and analysis for spatial data*, volume 101. Chapman & Hall/CRC, Boca Raton, Florida.

Boskov, M. and Verrall, R. (1994). Premium rating by geographic area using spatial models. *Astin Bulletin*, 24(1):131–143.

Bühlmann, H. (1967). Experience rating and credibility. *Astin Bulletin*, 4:199–207.

Bühlmann, H. and Gisler, A. (2005). *A course in credibility theory and its applications*. Springer, Berlin.

Bühlmann, H. and Straub, E. (1970). Glaubwüdigkeit für schadensäze. *Bulletin of the Swiss Association of Actuaries*, 70(1):111–133.

Chilès, J.-P. and Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons, Inc., New York.

Cressie, N. (1993). *Statistics for spatial data*. John Wiley & Sons, Inc., New Jersey.

Cressie, N. and Wilkie, C. K. (2011). *Statistics for spatio-temporal data*. John Wiley & Sons, Inc., Hoboken, New Jersey.

Dannenburg, D. (1995). Crossed classification credibility models. In *Transactions of the 25th International Congress of Actuaries*, volume 4, pages 1–35.

Fahrmeir, L., Sagerer, F., and Sussmann, G. (2007). Geoadditive regression for analyzing small-scale geographical variability in car insurance. *Blätter der DGVFM*, 28(1):47–65.

Frees, E. and Wang, P. (2006). Copula credibility for aggregate loss models. *Insurance: Mathematics and Economics*, 38(2):360–373.

Frees, E. W., Young, V. R., and Luo, Y. (1999). A longitudinal data analysis interpretation of credibility models. *Insurance: Mathematics and Economics*, 24(3):229–247.

Goulet, V. (2001). A generalized crossed classification credibility model. *Insurance: Mathematics and Economics*, 28(2):205–216.

Gschlöbl, S. and Czado, C. (2007). Spatial modelling of claim frequency and claim size in non-life insurance. *Scandinavian Actuarial Journal*, 2007(3):202–225.

Hachemeister, C. A. (1975). Credibility for regression models with application to trend. In *Credibility, Theory and Application*, Proc. Berkeley Actuarial Research Conference on Credibility, pages 129–163, New York. Academic Press.

Jewell, W. S. (1975). The use of collateral data in credibility theory: a hierarchical model. *Giornale dell'Istituto Italiano degli Attuari*, 38:1–16.

Lo, C. H., Fung, W. K., and Zhu, Z. Y. (2006). Generalized estimating equations for variance and covariance parameters in regression credibility models. *Insurance: Mathematics and Economics*, 39(1):99–113.

Matheron, G. (1963). Principles of geostatistics. *Economic geology*, 58(8):1246–1266.

Moran, P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, pages 17–23.

Omre, H. (1987). Bayesian krigingmerging observations and qualified guesses in kriging. *Mathematical Geology*, 19(1):25–39.

Omre, H. and Halvorsen, K. (1989). The bayesian bridge between simple and universal kriging. *Mathematical Geology*, 21(7):767–786.

Pineda-Krch, M. (2010). Great-circle distance calculations in r. `http://www.r-bloggers.com/great-circle-distance-calculations-in-r/`.

Scheiner, S. and Gurevitch, J. (2001). *Design and analysis of ecological experiments*. Oxford University Press, New York.

Schnapp, F. F., Driscoll, J. L., Zacharias, T. P., and Josephson, G. R. (2000). Ratemaking considerations for multiple peril crop insurance. *Casualty Actuarial Society Forum*.

Taylor, G. (1989). Use of spline functions for premium rating by geographic area. *Astin Bulletin*, 19(1):91–122.

Taylor, G. (2001). Geographic premium rating by whittaker-spatial smoothing. *Astin Bulletin*, 31(1):147–160.

Wen, L. and Wu, X. (2011). The credibility estimator with general dependence structure over risks. *Communications in Statistics-Theory and Methods*, 40:1893–1910.

Wen, L., Wu, X., and Zhou, X. (2009). The credibility premiums for models with dependence induced by common effects. *Insurance: Mathematics and Economics*, 44:19–25.

Yeo, K. L. and Valdez, E. A. (2006). Claim dependence with common effects in credibility models. *Insurance: Mathematics and Economics*, 38(3):609–629.