Contribution to the weak convergence of empirical copula process : contribution to the stochastic claims reserving in general insurance

Przemyslaw Sloma

Thèse de Doctorat de l'Université Paris VI

Spécialité
Mathématiques Appliquées

Présentée par : Przemyslaw Sloma

Pour obtenir le titre de
Docteur en Sciences de l'Université Pierre et Marie Curie - Paris VI

Sujet de la thèse

Contribution to the weak convergence of empirical copula processes.
Contribution to the stochastic claims reserving in general insurance.

Soutenue le 30 septembre devant le jury composé de:

Directeur de thèse:
M. Paul DEHEUVELS *Professeur, Université Paris 6*

Rapporteurs:
Mme Vronique MAUME-DESCHAMPS *Professeur, Université Claude Bernard Lyon 1*
M. Arthur CHARPENTIER *Professeur, Université du Québec à Montréal*

Examinateurs:
M. Jean-David FERMANIAN *Professeur, Crest-Ensae*
M. Gérard BIAU *Professeur, Université Paris 6*
M. Michel BRONIATOWSKI *Professeur, Université Paris 6*

Remerciements

Je tiens à remercier en tout premier lieu Monsieur le Professeur Paul Deheuvels, mon directeur de thèse. Tout au long de ces années de notre collaboration, j'ai eu la chance de profiter de ses immenses connaissances mathématiques et statistiques, et de ses idées enrichissantes. Il a consacré de nombreuses heures à discuter avec moi et lire mes textes, jusqu'à corriger des fautes de grammaire. Je voudrais lui exprimer tous mes remerciements, pour sa disponibilité et pour son soutien.

Un remerciement spécial va à Johan Segers pour ces commentaires précieux sur mes travaux relatifs au processus empirique de copule ainsi que pour l'exemple de la *"very non-smooth copula"* (cf. Section 2.6.6).

Je remercie chaleureusement mes rapporteurs. Je leur suis très reconnaissant du temps qu'ils ont consacré à l'expertise de mes traveaux.

J'exprime également ma gratitude envers les membres du jury. Je leur adresse mes sincères remerciements pour l'intérêt qu'ils ont accordé à mes traveaux.

Enfin, je voudrais remercier ma famille: mon épouse *Lidia*, mon fils *Krystian*, *mes parents*, ma soeur *Dominika*, mon frère *Radek* et mon filleul *Seweryn*.

Résumé

Dans la première partie de la thèse, nous nous intéressons à la convergence faible du processus empirique pondéré des copules. Nous fournissons la condition suffisante pour que cette convergence ait lieu vers un processus Gaussien limite. Nos résultats sont obtenus dans un espace de Banach L^p $(1 \leq p < \infty)$. Nous donnons des applications statistiques de ces résultats aux tests d'adéquation (tests of goodness of fit) pour les copules. Une attention spéciale est portée aux tests basés sur des statistiques de type Cramér-von Mises.

Dans un second temps, nous étudions le problème de provisionnement stochastique pour une compagnie d'assurance non-vie. Les méthodes stochastiques sont utilisées afin d'évaluer la variabilité des réserves. Le point de départ pour cette thèse est une incohérence entre les méthodes utilisées en pratique et celles publiées dans la littérature. Pour remédier à cela, nous présentons un outil général de provisionnement stochastique à horizon ultime (Chapitre 3) et à un an (Chapitre 4), basé sur la méthode Chain Ladder.

Abstract

The aim of this thesis is twofold. First, we concentrate on the study of weak convergence of weighted empirical copula processes.

We provide sufficient conditions for this convergence to hold to a limiting Gaussian process. Our results are obtained in the framework of convergence in the Banach space L^p $(1 \leq p < \infty)$. Statistical applications to goodness of fit (GOF) tests for copulas are given to illustrate these results. We pay special attention to GOF tests based on Cramér-von Mises type statistics.

Second, we discuss the problem of stochastic claims reserving in general non-life insurance. Stochastic models are needed in order to assess the variability of the claims reserve. The starting point of this thesis is an observed inconsistency between the approaches used in practice and that suggested in the literature. To fill this gap, we present a general tool for measuring the uncertainty of reserves in the framework of ultimate (Chapter 3) and one-year time horizon (Chapter 4), based on the Chain-Ladder method.

Accompanying papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I Przemyslaw Sloma (2014):
Weak convergence of weighted empirical copula processes in L^p spaces.
Paper to be submitted.

II Przemyslaw Sloma (2013):
Generalized Framework of Mack Stochastic Chain Ladder Method.
Paper submitted to *Variance*.

III Przemyslaw Sloma (2014):
Generalized Framework for Measuring the Uncertainty of the Claims Development Result.
Paper to be submitted.

List of Abbreviations and Symbols

Paper I

$\dot{C}_\theta$ gradient of C_θ with respect to θ, page 62

$\ell^\infty([0,1]^d)$ the space of all uniformly bounded functions on $[0,1]^d$, page 55

$\mathbb{1}_A$ indicator function of event A, page 15

$\mathbb{B}$ C-Brownian bridge defined via (2.2.13), page 58

$\mathbb{B}^*$ centered Gaussian process defined by (2.2.14), page 58

$\mathbb{C}_n$ empirical copula process based on $\widetilde{C}_n$ and C, defined via (2.2.11) , page 57

$\overline{\mathbb{C}}_n$ empirical copula process based on $\overline{C}_n$ and C, defined via (2.4.1), page 61

$\overline{C}_n$ empirical copula function defined via (2.2.10), page 56

$\partial C_\theta^{(n)}/\partial u$ consistent estimator of $\partial C_\theta/\partial u$, page 176

$\partial C_\theta^{(n)}/\partial v$ consistent estimator of $\partial C_\theta/\partial v$, page 176

Π_i projection on coordinate i, page 68

$\rightsquigarrow$ weak convergence, see Definition 1.1.3 , page 18

$\stackrel{\mathcal{L}}{=}$ equality in distribution, page 55

Θ weak limit of Θ_n, defined in hypothesis H.1 , page 62

$\widetilde{C}_n$ empirical copula function defined via (2.2.9), page 56

$\xrightarrow{\mathbb{P}^*}$ (or $\xrightarrow{\mathbb{P}}$), convergence in outer probability, see Definition 1.1.3 , page 18

C^{Arch} Archimedean copula, page 20

C^{d-Arch} Multivariate Archimedean copula, page 85

C^{Ga} Gaussian copula, page 20

C^t Student copula, page 88

C_n empirical copula function defined via (2.2.8), page 55

C_j the j-th first-order partial derivative of C defined via (2.2.3), page 53

I identity function on $[0, 1]$, page 68

K^{-1} generalized inverse of K, page 14

$L^\infty([0, 1]^d)$ space of all essentially bounded functions, page 52

$L^p([0, 1]^d)$ space of all p^{th} power integrable functions, page 52

$w\mathbb{B}^{**}$ centered Gaussian process defined in Proposition 2.4.1, page 62

Paper II and Paper III

$\delta_{i,j}$ random variables defined by (3.3.2), page 100

$\gamma_{i,j}$ random variables defined by (3.3.4), page 101

$\sigma(C_{i,j})$ σ- field generated by $C_{i,j}$, page 115

$C_{i,j}$ cumulative payments for accident year i until development year j, page 99

$F_{i,k}$ individual development factors (or link ratios) defined via (3.3.1), page 100

MW Merz and Wüthrich, page 47

Contents

Chapter 1

General Introduction

The present thesis dissertation is divided in two main parts. The first part (**Paper I**), of theoretical character, concerns the study on the weak convergence of (weighted) empirical copula processes in L^p spaces. The second part (**Paper II and Paper III**), with more practical applications, is related to the problem of claims reserving in general non-life insurance. In **Paper II** we consider the topic of reserve risk under the current regulatory regime - **Solvency I**. In **Paper III** we study the same problem (reserve risk) in the framework of the future European insurance regulations - **Solvency II**. These two parts are treated separately. They are somehow linked in the sense that copula functions provide a useful tool to model multivariate reserve risk in non-life insurance. Presently, these joint aspect will not be considered. It turns out that we have considered here two separate problems, leaving for future research the goal of treating these questions in a unified framework.

1.1 General Introduction to Paper I

1.1.1 Notes on copula functions

Copula functions have recently become one of the most significant new tools to handle in a flexible way the dependence relations between random variables. They play an important role in the construction of multivariate distribution functions. As a consequence, copulas can be very useful for building the stochastic models having different properties that cannot be overlooked

in practice (e.g., heavy tails, asymmetries, etc.). We refer to Trivedi and Zimmer (2005) for general discussions of these questions.
Interest in copulas arises from several fields. First, econometricians often have more information about marginal distributions of related variables than on their joint distribution. The copula approach provides then a useful method for deriving joint distributions for these marginal distributions, especially when the variables are considered nonnormal. Second, in a bivariate context, copulas are instrumental to define nonparametric measures of dependence for pairs of random variables. Finally, copulas give useful extensions and generalizations of approaches for modeling joint distributions and dependence that have appeared in the literature.

The term copula was introduced by Sklar (1959). However, the idea of copula had previously appeared in a number of papers, most notably in Hoeffding (1940, 1941) who established best possible bounds for these functions and studied measures of dependence that are invariant under strictly increasing transformations. Copulas have proved themselves useful in a variety of modeling situations. Two of the most commonly used applications are briefly mentioned:

- Financial applications: copula functions are commonly used in financial risk assessment. They present an attractive alternative for pricing the assets in models parameterized by nonnormal marginals. Copulas are also used in: asset allocation, credit scoring, default risk modeling, derivative pricing, and risk management.

- Actuarial applications: actuaries have often used copulas, among others, in modeling dependent mortality, claims losses and pricing of (re)-insurance contracts.
  Concerning the first application, it is often of interest to examine the joint mortality pattern of groups of more than one single individual. This group could be, for example, a husband and wife, a family with children, or twins. In such cases, there is strong empirical evidence to support the dependence of mortality on pairs of individuals. For example, statistical analyses of mortality patterns of married couples are frequently made to test the "broken heart" syndrome. Intuitively, pairs of individuals exhibit dependence in mortality because they share common risk factors. These factors may be purely genetic, as in the case of twins, or environmental, as in the case of a married couple.

Concerning the second and third applications, it is common in actuarial practice to model the joint distribution of indemnity claims and an allocated loss adjustment expense (ALAE). ALAE are types of insurance company expenses that are specifically attributable to the settlement of individual claims such as lawyers fees and claims investigation expenses. One of the possible ways to describe the joint distribution of losses and expenses is to fit a bivariate copula function to the data. After having identified the joint distribution of vectors (loss, expense) by fitting the copula to the bivariate data, it is possible to analyze the distribution of arbitrary functions of the (loss, expense) vector. This leads us to applications of copulas in the pricing of the (re)insurance contracts. We refer to Frees and Valdez (1998) and Genest et al (1998) for more details about the actuarial applications of copula functions.

Moreover, the article of de Jong (2012) is one of the first papers dealing with the actuarial applications of copulas to a multivariate framework of claims reserving. The approach described in this work can be considered as a starting point for bridging the two separate parts of present thesis (Paper I on the one hand and Paper II together with Paper III on the other hand).

1.1.2 Copula functions

The problem of investigating stochastic dependence is then reduced to the problem of investigating bivariate distribution functions on the unit cube $[0,1]^d$ with uniform marginals, which leads, namely, to the copula distribution. For the sake of brevity we only state the results for $d = 2$.

Definition 1.1.1 (Copula) *A 2-dimensional copula* $C : [0,1]^2 \to [0,1]$ *is a bivariate cumulative distribution function on the unit cube with uniform marginals.*

The motivation of this definition is summarized in the following well-known theorem of Sklar which underlies most applications of copulas. It turns out that relaxing the assumption of continuity of the marginals results in the non-uniqueness of the associated copula.
For a distribution function K on the real line K^{-1} denotes the generalized

inverse of K and is formally defined by

$$K^{-1}(u) := \begin{cases} \inf\{\, x \in \mathbb{R} \,|\, K(x) \geq u \,\} & 0 < u \leq 1, \\ \sup\{\, x \in \mathbb{R} \,|\, K(x) = u \,\} & u = 0, \end{cases} \tag{1.1.1}$$

where $\inf \emptyset := \infty$ and $\sup \emptyset := -\infty$. The range of K is denoted by $ranK$. Let $\bar{\mathbb{R}}$ denotes the extended real line $[-\infty, \infty]$.

Theorem 1.1.1 (Sklar's Theorem) *Let F be a 2-dimensional distribution function with margins F_p for $p = 1, 2$. Then there exists a copula C such that for all $\mathbf{x} = (x_1, x_2) \in \bar{\mathbb{R}}^2$*

$$F(\mathbf{x}) = C(F_1(x_1), F_2(x_2)). \tag{1.1.2}$$

If all marginals are continuous then C is uniquely determined by

$$C(\mathbf{u}) = F(F_1^{-1}(u_1), F_2^{-1}(u_2)), \;\; \mathbf{u} = (u_1, u_2), \tag{1.1.3}$$

otherwise it is uniquely determined on the product of the ranges of the marginal distributions, namely $ranF_1 \times ranF_2$. Conversely, if C is a copula and F_1 and F_2 are distribution functions, then the function F defined by (1.1.2) is a joint distribution function with marginals F_1, F_2.

Proposition 1.1.1 *Let C be a 2-dimensional copula and suppose that $X = (X_1, X_2)^T \sim F = C(F_1, F_2)$ is a random vector with marginals F_1, F_2 and copula C. Then the following results hold:*

(i) ***Independence.*** *If F_1 and F_2 are continuous then X_1 and X_2 are independent if and only if $C(\mathbf{u}) = u_1 u_2 = \Pi(u)$. The latter distribution function is called the independence copula.*

(ii) ***Lipschitz-continuity.*** *C is Lipschitz-continuous with respect to the L^1-norm on $[0, 1]^2$ in the sense that*

$$|C(\mathbf{u}) - C(\mathbf{v})| \leq \|\mathbf{u} - \mathbf{v}\| = |u_1 - v_1| + |u_2 - v_2| \text{ for all } \mathbf{u}, \mathbf{v} \in [0, 1]^2.$$

(iii) ***Differentiability.*** *For all $u_1 \in [0, 1]$ it holds that the partial derivative of C with respect to u_2 $C_2 := \partial C(\boldsymbol{u})/\partial u_2$ exists for λ_1-almost every u_2. Furthermore, $0 \leq \partial C(\boldsymbol{u})/\partial u_2 \leq 1$. The same is true with (u_1, u_2) changed into (u_2, u_1).*

(iv) ***Invariance under increasing transformations.*** *If F_1 and F_2 are continuous and α, β are strictly increasing mappings, then the copula of $(\alpha \circ X_1, \beta \circ X_2)$ is C.*

(v) ***Kendall's-τ.*** *If F_1 and F_2 are continuous then the Kendall's τ is given by*

$$\tau = \tau_{X_1,X_2} = 4 \int_{[0,1]^2} C(\boldsymbol{u}) dC(\boldsymbol{u}) - 1.$$

For proofs of these results and more details regarding the general theory of copulas we refer the reader to the monographs Nelsen (2006); Joe (1997).

1.1.3 Empirical copula function

The empirical copula function is the most famous and easiest nonparametric estimator for the copula C of a random vector. Let $\mathbf{X}_i = (X_i(1), X_i(2))$, $i = 1, 2, ..., n$, be a sequence of independent identically distributed bivariate random vectors with cumulative distribution function (cdf) F, continuous marginal distribution functions F_1 and F_2 and copula C.

Let $\mathbb{1}_A$ denotes the indicator function of event A. The empirical copula as the simplest nonparametric estimator for C (going back to Deheuvels (1979)) simply replaces the unknown terms in equation (1.1.3) by their empirical counterparts, that is

$$\widetilde{C}_n(\mathbf{u}) := F_n(F_{n1}^{-1}(u_1), F_{n2}^{-1}(u_2)), \tag{1.1.4}$$

where

$$F_n(\mathbf{x}) = F_n(x_1, x_2) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i(1) \le x_1, X_i(2) \le x_2\}},$$

$$F_{nj}(x_j) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i(j) \le x_j\}}, \; j = 1, 2,$$

denote the corresponding empirical distribution functions. The functions F_{nj}^{-1}, for $j = 1, 2$, are defined via (1.1.1).

It is noteworthy that the literature provides several similar nonparametric estimators for the copula. For example, (see Genest et al (1995)) studied the rank-based estimator

$$\overline{C}_n(\mathbf{u}) = \overline{C}_n(u_1, u_2) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{F_{n1}(X_i(1))\leq u_1, F_{n2}(X_i(2))\leq u_2\}}. \tag{1.1.5}$$

In the latter expression the marginal empirical distribution functions F_{nj} are often replaced by their rescaled counterparts $\widehat{F}_{nj} = \frac{n}{n+1}F_{nj}$. Both modifications do not affect the asymptotic behavior of these empirical functions, see Proposition 2.2.2 (e).

1.1.4 Weak Convergence in metric spaces

Let $(\mathbb{D}, d)$ be a metric space and let $(P_n)_{n\in\mathbb{N}}$ and P be Borel probability measures on $(\mathbb{D}, \mathcal{D})$, where $\mathcal{D}$ denotes the Borel σ-field on $\mathbb{D}$. Weak convergence of P_n to P, which we write as $P_n \rightsquigarrow P$, is classically defined through the requirement that

$$\int_{\mathbb{D}} f\ dP_n \to \int_{\mathbb{D}} f\ dP \text{ for all } f \in C_b(\mathbb{D}),$$

where $C_b(\mathbb{D})$ denotes the set of all bounded, continuous and real-valued functions on $\mathbb{D}$ (see e.g. Billingsley (1968)). For $\mathbb{D}$-valued random variables $(X_n)_{n\in\mathbb{N}}$ and X, weak convergence is conveniently described in terms of their induced laws, so that $X_n \rightsquigarrow X$ if and only if

$$\mathbb{E}f(X_n) \to \mathbb{E}f(X) \text{ for all } f \in C_b(\mathbb{D}). \tag{1.1.6}$$

The classical theory of weak convergence requires that all random variables involved are Borel measurable. While this condition usually holds for separable metric spaces such as $\mathbb{R}^d$ or $C_b[0, 1]$, it becomes problematic when the metric spaces are nonseparable. A classical example is the càdlàg-space $D[0, 1]$, containing all functions on the unit interval which are right-continuous and possess left-hand limits, equipped with the metric induced by the supremum norm. For i.i.d. random variables $X_1, ..., X_n$ on $[0, 1]$ the empirical distribution function

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i\leq t\}},\ t \in [0, 1],$$

as well as the empirical process

$$\alpha_n(t) = \sqrt{n}(F_n(t) - F(t)),\ t \in [0,1],$$

seen as random variables in $D[0,1]$, are not Borel measurable if $D[0,1]$ is equipped with the supremum norm, see, e.g., Billingsley (1968).

During the last decades several approaches to overcome this difficulty have been suggested. In this section we will briefly summarize the most modern approach, essentially due to J. Hoffmann-Jørgensen, and extensively investigated in van der Vaart and Wellner (1996) and Kosorok (2008). The key idea is to drop the requirement of Borel measurability of each X_n, meanwhile upholding the requirement (1.1.6), where the expectations are replaced by outer expectations.

Definition 1.1.2 (Outer integral and outer probability) *Let T be an arbitrary map from a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ to the extended real line $\overline{\mathbb{R}}$. The outer integral of T with respect to $\mathbb{P}$ is defined as*

$$\mathbb{E}^*T = \inf\{\mathbb{E}U, U \geq T, U : \Omega \to \overline{\mathbb{R}} \textit{ mesurable and } \mathbb{E}U \textit{ exists } \}. \tag{1.1.7}$$

The outer probability of an arbitrary subset $B \subseteq \Omega$ is defined as

$$\mathbb{P}^*(B) = \inf\{\mathbb{P}(A), A \supset B, A \in \mathcal{A}\}. \tag{1.1.8}$$

Inner integrals and *inner probabilities* are defined by $\mathbb{E}_*T = -\mathbb{E}^*(-T)$ and $\mathbb{P}_*(B) = 1 - \mathbb{P}^*(\Omega \setminus B)$. The infima in the latter definitions are always achieved, see the following Lemma, which is proved in van der Vaart and Wellner (1996).

Lemma 1.1.2 (Measurable cover functions)

For any map $T : \Omega \to \overline{\mathbb{R}}$ there exists a measurable function $T^* : \Omega \to \overline{\mathbb{R}}$ with $T^* \geq T$ and with $T^* \leq U$ a.s. for every measurable $U : \Omega \to \overline{\mathbb{R}}$ with $U \geq T$ a.s. For every such T^* it holds $\mathbb{E}^*T = \mathbb{E}T^*$, provided that $\mathbb{E}T^*$ exists. The latter is certainly true if $\mathbb{E}^*T < \infty$.

With Definition 1.1.2 and Lemma 1.1.2 at hand, we can define weak convergence, outer almost sure convergence and convergence in outer probability for arbitrary nonmeasurable maps.

Definition 1.1.3 (Convergence: Weak, outer almost surely and in outer probability) *Let $X_n : \Omega_n \to \mathbb{D}$, $X : \Omega \to \mathbb{D}$ be arbitrary maps defined on some probability spaces $(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$, $(\Omega, \mathcal{A}, \mathbb{P})$.*

*(i) If X is Borel measurable we say that X_n **weakly converges** to X, written $X_n \rightsquigarrow X$,if and only if*

$$\mathbb{E}^* f(X_n) \to \mathbb{E} f(X) \text{ for all } f \in C_b(\mathbb{D}).$$

*(ii) If X_n, X are defined on a common probability space we say that X_n **converges outer almost surely** to X if $d(X_n, X)^* \to 0$ almost surely for some version of $d(X_n, X)^*$. This is denoted by $X_n \xrightarrow{as*} X$.*

*(iii) If X_n, X are defined on a common probability space we say that X_n **converges in outer probability** to X if $d(X_n, X)^* \to 0$ in probability. This is equivalent to $\mathbb{P}^*(d(X_n, X) > \varepsilon \to 0$ for every $\varepsilon > 0$ and is denoted by $X_n \xrightarrow{\mathbb{P}^*} X$.*

With this definition much of the theory for non-measurable maps parallels the classical theory, up to a remarkable degree. For example, the similarities include a Portmanteau Theorem, continuous mapping results, a Prohorov Theorem and the metrization of weak convergence to separable limits by the bounded Lipschitz-metric. The latter is developed in Section 1.12 in van der Vaart and Wellner (1996) and states that $X_n \rightsquigarrow X$, where X is Borel measurable and separable if and only if

$$\sup_{f \in BL_1(\mathbb{D})} |\mathbb{E}^* f(X_n) - \mathbb{E} f(X)| \to 0,$$

where $BL_1(\mathbb{D})$ denotes the set of all real functions on $\mathbb{D}$ which are bounded by 1 and satisfy the Lipschitz condition $|f(x) - f(y)| \leq d(x, y)$ for all $x, y \in \mathbb{D}$. Two further important properties for the investigation of stochastic convergence of nonmeasurable maps are summarized in the following definition.

Definition 1.1.4 *Let $(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$ be a sequence of probability spaces and let $X_n : \Omega_n \to D$ be arbitrary maps.*

*(i) $(X_n)_{n \in \mathbb{N}}$ is **asymptotically measurable** if $\mathbb{E}^* f(X_n) - \mathbb{E} f(X) \to 0$ for every $f \in \mathbb{C}_b(\mathbb{D})$.*

(ii) $(X_n)_{n\in\mathbb{N}}$ *is* ***asymptotically tight*** *if for every* $\varepsilon > 0$ *there exists a compact set* $K \subset \mathbb{D}$ *with* $\liminf_{n\to\infty} \mathbb{P}_*(X_n \in G) \geq 1-\varepsilon$ *for every open* $G \supseteq K$.

By Lemma 1.3.8 in van der Vaart and Wellner (1996) weak convergence of $(X_n)_n$ to a tight limit implies both asymptotic tightness and asymptotic measurability.

1.1.5 Empirical copula process

Let define the space $\ell^\infty([0,1]^2)$ of all uniformly bounded functions φ on $[0,1]^2$ into $\mathbb{R}$ equipped with the topology induced by sup-norm, namely, $|\varphi|_\infty := \sup_{\mathbf{u}\in[0,1]^2} |\varphi(\mathbf{u})|$.

The asymptotic behavior of empirical copula process, $\mathbb{C}_n := \sqrt{n}(\widetilde{C}_n - C)$ was studied in several papers, starting with Rüschendorf (1976) in the Skorohod space $D([0,1]^d)$ endowed with a particular metric (see, e.g., his Theorem 3.3). Deheuvels (1981) established this result in the framework of independent margins. In Gaenssler and Stute (1987),van der Vaart and Wellner (1996) and Fermanian et al (2004), the weak convergence in $\ell^\infty([0,1]^2)$ space of the empirical copula process is shown to hold under the assumption that the first-order partial derivatives of the copula exist and are continuous on certain closed subsets or on the whole unit hypercube.

The rates of convergence of certain remainder terms have been established in Tsukahara (2005) (see also Tsukahara (2011)) for copulas that are twice continuously differentiable on the closed hypercube.

Unfortunately, it turns out that for many (even most) popular copula families, even the first order partial derivatives of the copula fail to be continuous at some boundary points of the hypercube. We present below the examples of commonly-used families of copulas (e.g., Gaussian and Archimedean), for which the latter condition is not satisfied.

Example 1.1.1 (Gaussian copula) *The bivariate Gaussian (or Normal) copula with parameter* $\rho \in (-1,1)$ *is defined via an application of Sklar's theorem (see Theorem 1.1.1 on page 14 and Nelsen (2006)) by*

$$C^{Ga}(u_1,u_2) := \Phi_2(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho), \tag{1.1.9}$$

where Φ *is the cumulative distribution function (cdf) of the standard normal distribution and* Φ_2 *denotes the cdf of the bivariate standard normal distribution with correlation parameter* $\rho \in (1, 1)$. *Denote by* $C_1^{Ga}(u_1, u_2)$ *the first-order partial derivative of* C^{Ga} *with respect to* u_1*, namely,* $C_1^{Ga}(u_1, u_2) := \partial C^{Ga}(u_1, u_2)/\partial u_1$. *On the one hand, for* $\rho \in (0, 1)$*, we have that* $\lim_{u_1 \downarrow 0} C_1^{Ga}(u_1, u_2) = 1$ *for all* $u_2 \in (0, 1]$*, whereas on the other hand we have* $C^{Ga}(u_1, u_2) = 0$ *when* $u_2 = 0$. *As a consequence, the definition of* $C_1^{Ga}(u_1, u_2)$ *cannot be extended by continuity to the point* $(0, 0)$. *By similar arguments we can show that the definition of* $C_1^{Ga}(u_1, u_2)$ *cannot be extended to* $(1, 1)$ *(and to the points* $(0, 1)$ *and* $(1, 0)$ *if* $\rho \in (-1, 0)$*). This means that the first-order partial derivatives of* ***Gaussian copula*** *cannot be extended by continuity on the unit cube.*

Example 1.1.2 (Archimedean copulas) *Let* C^{Arch} *be a bivariate Archimedean copula, that is (see, e.g., Nelsen (2006) and Frees and Valdez (1998)):*

$$C^{Arch}(u_1, u_2) = \phi^{-1}(\phi(u_1), \phi(u_2)), \quad (u_1, u_2) \in [0, 1]^2, \tag{1.1.10}$$

where the function $\phi : [0, 1] \mapsto [0, \infty]$ *(called also generator) is convex, strictly decreasing, finite on* $(0, 1]$*, and vanishes at* 1*, whereas* $\phi^{-1} : [0, \infty) \mapsto [0, 1]$ *is its generalized inverse, defined via (1.1.1) on page 14. Suppose that* ϕ *is continuously differentiable on* $(0, 1]$ *and* $\phi'(0+) = -\infty$. *Then the first-order partial derivatives of* C^{Arch} *are given by*

$$C_j^{Arch}(u_1, u_2) = \frac{\phi'(u_j)}{\phi'(C^{Arch}(u_1, u_2))}, \; (u_1, u_2) \in [0, 1]^2, \; 0 < u_j < 1, \; j = 1, 2.$$

The partial derivative C_j^{Arch} *is likely to fail to be continuous at some boundary points. For instance, if* $\phi(1) = 0$*, then* C_j^{Arch} *cannot be extended by continuity to* $(1, 1)$. *When* ϕ^{-1} *is long-tailed, that is, if* $\lim_{x \to \infty} \phi^{-1}(x + y)/\phi^{-1}(x) = 1$ *for all* $y \in \mathbb{R}$*, then* $\lim_{u_1 \downarrow 0} C^{Arch}(u_1, u_2)/u_1 = 1$ *for all* $u_2 \in (0, 1]$*, whereas* $C_1^{Arch}(u_1, u_2) = 0$ *as soon as* $u_2 = 0$. *It follows that in this case* C_1^{Arch} *cannot be extended by continuity to the point* $(0, 0)$.

Segers (2012) provides a remedy to this situation by showing that the above cited results on the empirical copula process actually do hold under a much less restrictive assumption. The assumption is non-restrictive in the sense that it is needed anyway to ensure that the candidate limiting process exists and has continuous trajectories:

Condition 1. For each $j = 1, 2$, the $j - th$ first order derivative C_j exists and is continuous on the set

$$V_j := \{(u_1, u_2) \in [0, 1]^2 : 0 < u_j < 1\}. \tag{1.1.11}$$

We show below that for the copula families from the above mentioned examples, the **Condition 1** holds.

Example 1.1.3 (Archimedean copulas) *Recall from (1.1.10) on page 20 the definition of a bivariate Archimedean copula C^{Arch} with generator ϕ. We suppose that ϕ is continuously differentiable on $(0, 1]$ and $\phi'(0+) = -\infty$. The first-order partial derivatives of C^{Arch} are given by*

$$C_j^{Arch}(u_1, u_2) = \frac{\phi'(u_j)}{\phi'(C^{Arch}(u_1, u_2))}, \ (u_1, u_2) \in [0, 1]^2, \ 0 < u_j < 1, \ j = 1, 2.$$

*Without loss of generality we show that this condition holds for $j = 1$. The **Condition 1** is verified, for $j = 1$, if C_1^{Arch} exist and is continuous on the set $(0, 1) \times [0, 1]$. It is enough to prove this property on the set $(0, 1) \times \{0, 1\}$. If $(u_1, u_2) \in (0, 1) \times \{0\}$, then $C^{Arch}(u_1, u_2) = 0$ and $\phi'(C(u_1, u_2)) = -\infty$. This implies that $C_1^{Arch}(u_1, u_2) = 0$. By similar arguments we conclude that the same property holds on the set $(0, 1) \times \{1\}$. As a consequence, the bivariate Archimedean copulas fulfill **Condition 1**.*

Example 1.1.4 (Gaussian copula) *Recall the definition of the bivariate Gaussian (or Normal) copula given in Example 1.1.1 on page 19. Since,*

$$\begin{aligned}
\Phi_2(h, k; \rho) &:= \int_{-\infty}^{h} \int_{-\infty}^{k} \varphi_2(x, y; \rho) \ dx \ dy, \\
\varphi_2(x, y; \rho) &:= \frac{1}{2\pi\sqrt{1 - \rho^2}} \exp\left(-\frac{x^2 - 2\rho xy + y^2}{2(1 - \rho^2)}\right),
\end{aligned}$$

by direct computations and standard methods, we see that

$$\frac{\partial \Phi_2}{\partial h}(h, k) = \varphi(h) \cdot \Phi\left(\frac{k - \rho h}{\sqrt{(1 - \rho^2)}}\right),$$

where the function φ denotes the density of the standard normal $N(0,1)$ law. In view of (1.1.9), the first-order partial derivative of C^{Ga} with respect to u_1 is given by

$$C_1^{Ga}(u_1,u_2) := \frac{\partial C^{Ga}}{\partial u_1}(u_1,u_2) = \frac{\partial \Phi_2}{\partial h}(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho) \cdot \frac{1}{\Phi'(u_1)}.$$

By combining the previous facts we get

$$C_1^{Ga}(u_1,u_2) = \Phi\left(\frac{\Phi^{-1}(u_2) - \rho\Phi^{-1}(u_1)}{\sqrt{1-\rho^2}}\right).$$

If $(u_1,u_2) \in (0,1) \times \{0\}$, then $\Phi^{-1}(0) = -\infty$ and $C_1^{Ga}(u_1,u_2) \to 0$ as $u_2 \downarrow 0$, for all $u_1 \in (0,1)$. By similar arguments, if $(u_1,u_2) \in (0,1) \times \{1\}$, then $\Phi^{-1}(0) = \infty$ and $C_1^{Ga}(u_1,u_2) \to 0$ as $u_2 \uparrow 1$, for all $u_1 \in (0,1)$. This shows that C_1^{Ga} exists and is continuous on the set $(u_1,u_2) \in (0,1) \times \{0,1\}$. We conclude from these arguments that the bivariate normal copula with correlation parameter $\rho \in (1,1)$ verifies ***Condition 1****.*

It is noteworthy that a stronger condition then **Condition 1** about the smoothness of first derivatives was provided by Omelka et al (2009). These authors claimed that weak convergence of the empirical copula process still holds whenever the first-order partial derivatives are continuous at $[0,1]^2 \setminus \{(0,0),(0,1),(1,0),(1,1)\}$.

However, there exist examples of copulas which are not continuously differentiable on $[0,1]^2$ and for which **Condition 1** does not hold:

Example 1.1.5 (Fréchet-Hoeffding bounds) *Every bivariate copula function C is bounded by the so-called* Fréchet-Hoeffding bounds*,*

$$W_2(u_1,u_2) \le C(u_1,u_2) \le M_2(u_1,u_2),$$

for $(u_1,u_2) \in [0,1]^2$, where $W_2(u_1,u_2) := \max(u_1+u_2-1, 0)$ and $M_2(u_1,u_2) := \min(u_1,u_2)$ and both of them are copulas themselves. Note that the Fréchet-Hoeffding *lower bound is no longer d-increasing, for $d \ge 3$ (see, e.g., (Nelsen, 2006, p.47)). It is easy to see that the first-order partial derivatives of M_2 and W_2 do not exist on the sets $\{(u_1,u_2) \in [0,1]^2 : u_1 = u_2\}$ and $\{(u_1,u_2) \in [0,1]^2 : u_1 + u_2 = 1\}$ respectively. With these observations, one find that the* Fréchet-Hoeffding bounds *do not satisfy* ***Condition 1****.*

Example 1.1.6 (Extreme-value copulas-Cuadras-Augé family) *The Cuadras-Augé copula with parameters $\alpha, \beta \in [0,1]$ is defined by*

$$C_{\alpha,\beta}(u_1,u_2) := \min(u_1^{1-\alpha}u_2, u_1u_2^{1-\beta}),\ (u_1,u_2) \in [0,1]^2.$$

This copula is symmetric and has a singular component on the main diagonal. More specifically, $P(U=V) = \beta/(2\beta)$. The functions $C_{\alpha,\beta}$ are the survival copulas associated with the Marshall and Olkin (1967) bivariate exponential distribution. It is then readily shown that the first-order partial derivatives of $C_{\alpha,\beta}$ do not exist on the set $\{(u_1,u_2) \in [0,1]^2 : u_1 = u_2\}$. As a result, ***Condition 1*** *does not hold for Cuadras-Augé family of copulas.*

Example 1.1.7 (Checkerboard copula) *The checkerboard copula function is defined via its Lebesgue density $c(u_1,u_2) := 2 \cdot \mathbb{1}_{[0,1/2]^2 \cup [1/2,1]^2}(u_1,u_2)$ (see e.g., Urrleman et al (2000)). Then, from the relation $C(u_1,u_2) = \int_0^{u_1}\int_0^{u_2} c(s,t)dsdt$, we obtain that*

$$C(u_1,u_2) = \begin{cases} 2u_1u_2, & (u_1,u_2) \in [0,1/2) \times [0,1/2) \\ u_1, & (u_1,u_2) \in [0,1/2] \times [1/2,1] \\ u_2, & (u_1,u_2) \in [1/2,1] \times [0,1/2] \\ \frac{1}{2} + 2(u_1 - \frac{1}{2})(u_2 - \frac{1}{2}), & (u_1,u_2) \in [1/2,1] \times (1/2,1] \end{cases}$$

As a simple consequence of the arguments above, we conclude that the first-order derivatives of C, namely C_1 (with respect to u_1) and C_2 (with respect to u_2), do not exist on the sets $\{1/2\} \times (0,1)$ and $(0,1) \times \{1/2\}$ respectively. It means that the checkerboard copula C does not verify ***Condition 1****.*

1.1.6 Contribution to the weak convergence of empirical copula process.

Let μ be the finite Borel measure and $\mathcal{B}$ be the Borel σ-algebra on $[0,1]^2$. The space $L^p([0,1]^2,\mu) := L^p([0,1]^2,\mathcal{B},\mu)$ is equipped with the usual norm $|f|_p$, where $|f|_p^p = \int_S |f|^p \, d\mu$ and $1 \le p < \infty$.

For $p = \infty$, we refer to (2.2.1) on page 52 for definition of the space $L^\infty([0,1]^d,\mu)$.

In the previous section, we provided examples of copulas that do not verify the smoothness **Condition 1**. In the latter examples, that we cannot conclude to the weak convergence of associated empirical copula process with respect supremum norm, by a direct application of the results of Segers (2012). In our thesis, we present a solution to this problem by considering the weak convergence of empirical copula process in the $L^p([0,1]^2, \mu)$ $(1 \leq p < \infty)$ Banach space topology. This approach allows us to solve the problem of convergence in $\ell^\infty([0,1]^2)$ space under very general smoothness assumptions. Namely, the convergence with respect to the weaker topology L^p still allows us to construct goodness-of-fit tests based on empirical copula process. Our result requires a rather weak smoothness condition. We can establish the weak convergence provided that the copula function fulfills:

Condition 2. The copula function C belongs to the class of copulas $\mathcal{C}$, defined via

$$\mathcal{C} = \{C - \text{copula function} : \mu(\mathcal{D}_C) = 1\}, \tag{1.1.12}$$

where $\mathcal{D}_C$ is defined by

$$\mathcal{D}_C = \bigcap_{j=1}^{2} \left\{ \mathbf{u} \in [0,1]^2 : \ C_j \text{ is defined and continuous at } \mathbf{u} \ \right\},$$

and, for $j = 1, 2$, $C_j := \frac{\partial}{\partial u_j} C(u_1, u_2)$ (refer to (2.2.3) on page 53 for more precise definition of C_j).

Condition 2 is weaker than **Condition 1**. It easy to show that the copula functions considered in the Examples 1.1.5-1.1.7 verify **Condition 2**.

Finally we extend our convergence result to weighted empirical copula process for weight functions that are p^{th} power integrable with respect to the finite Borel measure μ .

1.1.7 Goodness-of-fit testing

The goodness of fit (GOF) tests measure the compatibility of a random sample with a theoretical probability distribution function. In other words, these tests show how well the selected distribution fits the data. Assessing the fit of a model (i.e., the discrepancy between a model and the data) is critical in applications, as inferences drawn on poorly fitting models may be misleading.

The general GOF procedure consists of defining a test statistic which is some function of the data measuring the distance between the hypothesis and the data, and then calculating the probability of obtaining data which have a still larger value of this test statistic than the value observed, assuming the hypothesis is true. This probability is called the confidence level.

Pearson proposed the first important test of goodness of fit in 1900, namely the χ^2 test. The subsequent research devoted to enhancements of this elementary goodness-of-fit procedure became a major source of motivation for the development of key areas in Probability and Statistics such as the theory of weak convergence in general spaces and the asymptotic theory of empirical processes.

In our thesis, we pay special attention to the application of the theory of empirical copula processes to the asymptotic theory of goodness-of-fit tests for copulas. Firstly, we present the brief overview about the techniques of goodness-of-fit theory for general distribution (univariate and multivariate case). The reason is that this introduction allows the better understanding of methods used in the case of the statistical tests for copulas.

1.1.7.1 Univariate case

Goodness-of-fit testing to a single distribution

One of the simplest goodness-of-fit problem consists of testing fit to a single fixed distribution, namely, given a random sample of real r.v.s $X_1, X_2, ..., X_n$ with common d.f. F, testing the null hypothesis $H_0 : F = F_0$ for a fixed d.f. F_0. While this procedure is usually of limited interest in applications, the solutions proposed to this problem provided the main idea in subsequent generalizations designed for testing fit to composite null hypotheses.

The Pearson chi-square test can be considered as the first conclusive approach to the problem of testing fit to a fixed distribution. The GOF test proposed by Pearson consists in dividing the real line into k disjoint categories or cells $K_1, ..., K_k$ into which observations fall, under the null hypothesis, with probabilities $p_1, ..., p_k$. That is, if H_0 were true, then $P(X_1 \in K_i) = p_i, i = 1, ..., k$.

However, as pointed out by many authors, this test has a major weakness in the case when F is continuous. Namely, consideration of only the cell frequencies produces a loss of information that results in lack of power (the χ^2 statistic will not distinguish two different distributions sharing the same cell probabilities).

One of possible way to improve Pearson's statistic, by using the complete information provided by the data, consists of employing a functional distance to measure the discrepancy between the hypothesized d.f. F_0 and the empirical d.f. F_n. The first representatives of this method were proposed in the late 20's and in the 30's. Cramér (1928) and, in a more general form, von Mises (1931), proposed the use of the statistic

$$\omega_n^2 = n \int_{-\infty}^{\infty} [F_n(x) - F_0(x)]^2 w(x)\ dx,$$

for some suitable weight function w as an adequate measure of discrepancy. However, the statistic ω_n^2 is in general not distribution free (depends on the law of the random variable X_j). In order to overcome this difficulty, we consider the following modified version of ω_n^2

$$W_n^2(\Psi) = n \int_{-\infty}^{\infty} \Psi(F_0(x))[F_n(x) - F_0(x)]^2\ dF_0(x),$$

which was proposed by Smirnov (1936, 1937). All the statistics of this type, which can be obtained for various choices of the weight function Ψ, are usually referred to as statistics of Cramer-von Mises type. Consideration of appropriate weight functions Ψ allows the statistician to put special emphasis on the detection of particular sets of alternatives.

The convenience of employing $W_n^2(\Psi)$ instead of the weighted versions of Kolmogorov's statistics for example, can be understood taking into account that the latter accounts only for the largest deviation between $F_n(t)$ and $F_0(t)$, while $W_n^2(\Psi)$ is a weighted average of all the deviations between $F_n(t)$ and $F_0(t)$. Two particular statistics have received special attention in the literature. When $\Psi(F_0) = 1$,

$$W_n^2 = n \int_{-\infty}^{\infty} [F_n(x) - F_0(x)]^2\ dF_0(x),$$

is called the usual Cramer-von Mises statistic, and when $\Psi(t) = (t \cdot (1-t))^{-1}$ then

$$A_n^2 = n \int_{-\infty}^{\infty} \frac{[F_n(x) - F_0(x)]^2}{F_0(x)(1 - F_0(x))}\ dF_0(x),$$

is referred to as the Anderson-Darling statistic. The A_n^2 statistics has the additional appeal of weighting the deviations according to their expected value, and this results in a more powerful statistic for testing fit to a fixed distribution (see, e.g., d'Agostino and Stephens (1986) and Deheuvels and Martynov (2003)).

For the use in practice of any of these appealing statistics we should be able to obtain the corresponding significance levels. Since one may find some difficulties to find the exact distributions of statistics of Cramer-von Mises type, the obtaining the asymptotic distribution of the test statistics became of special interest.

Smirnov (1944) derived the asymptotic distribution of statistics of Cramér-von Mises type. However his proof was not based on the asymptotic behaviour of uniform empirical process. It was Doob (1949) who first conjectured the convergence of the uniform empirical process to the Brownian bridge. A useful consequence of this fact would be that, under some (not explicited) hypotheses, the derivation of the asymptotic distribution of a functional of the uniform empirical process could be reduced to the derivation of the distribution of the same functional for the Brownian bridge. Therefore, the justification of Doob's conjecture is the key to provide a new simpler proof of Smirnov's result.

This justification was given by Donsker by introducing the notion of invariance principle (see Donsker (1951, 1952)). His results showed that, under weak general condition, the distribution of a continuous functional of the partial sum process (obtained from a sequence of i.i.d. r.v.s with finite second moment) converges to the distribution of the corresponding functional of a Brownian motion, and that, likewise, the distribution of a continuous functional of the uniform empirical process converges to the distribution of the corresponding functional of a Brownian bridge.

The space $C[0,1]$ (space of all continuous functions on $[0,1]$) was one of the first metric spaces for which this theory was developed, through the work of Prohorov (1956). The scheme consisting of proving the convergence of the finite dimensional distributions plus a tightness condition allowed to obtain distributional limit theorems for slight modifications of the partial sum and the uniform empirical processes, because both processes could be approximated by equivalent processes obtained from them by linear interpolation so that all the random objects considered in the limit theorems remained in $C[0,1]$. This last approximation is somehow artificial. In order to avoid it, a wider space had to be considered. A proper study of the weak convergence of the uniform empirical process could be attempted in the space $D[0,1]$ of all càdlàg functions on $[0,1]$. The fact that the empirical process is not measurable when the uniform norm is considered, led to the introduction of a more involved topology, namely the Skorohod topology that turns $D[0,1]$ into a separable and complete metric space in which the empirical process is measurable. In this setup the weak convergence of the empirical process could be analyzed in a better way (see, e.g., Billingsley (1968) p. 141):

Theorem 1.1.3 *Let α_n be the uniform empirical process: $\alpha_n(t) = \sqrt{n}(G_n(t) - t))$, $0 \leq t \leq 1$, where $G_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{U_i(1)\leq t\}}$ and U_i are i.i.d. uniform r.v.'s. If $B(t)$ is a Brownian bridge, then*

$$\alpha_n \rightsquigarrow B \text{ in } \mathcal{D}[0,1].$$

Theorem 1.1.3 can be used in derivation the asymptotic distribution of the Cramer-von Mises statistics. Since the functional $x \mapsto \int_0^1 x(t)^2\, dt$ is continuous for the Skorohod topology outside a set of B-measure zero, and by the continuous mapping theorem, we can obtain for W_n^2 that

$$W_n^2 \rightsquigarrow \int_0^1 B(t)^2\, dt \text{ in } \mathbb{R}.$$

In Anderson and Darling (1952), the author showed how to derive the asymptotic distribution of other statistics of Cramér-von Mises type. As a consequence of the Law of the Iterated Logarithm for the Brownian motion, and provided that

$$\int_0^\delta \Psi(t) t \log\log\frac{1}{t}\, dt \quad \text{and} \quad \int_\delta^1 \Psi(t)(1-t)\log\log\frac{1}{1-t}\, dt$$

are finite for some $\delta \in (0,1)$, they showed that the functional $x \mapsto \int_0^1 \Psi(t)x(t)^2\, dt$ is continuous,with respect to the Skorohod distance, outside a set of B-measure zero. Consequently,

$$W_n^2(\Psi) \rightsquigarrow \int_0^1 \Psi(t)B(t)^2\, dt \text{ in } \mathbb{R}.$$

This result covers the Anderson-Darling statistic A_n^2. It is important to be noted that there exists an alternative technique to derive the asymptotic distribution of statistics of Cramér-von Mises type. The method described above is based on on the weak convergence of the empirical process considered as a random element with values in the space of càdlàg functions, endowed with the Skorohod topology, plus the continuity of a suitable functional. There more natural way is to consider the uniform empirical process as a random element taking values in the separable Hilbert space $L^2((0,1),\Psi)$ of all real, Borel measurable functions f on $(0,1)$ such that $\int_0^1 \Psi(t)f(t)^2\, dt$ is finite, where we consider the norm given by

$$\|f\|_{2,\Psi}^2 = \int_0^1 \Psi(t)f(t)^2\, dt.$$

In this setup, from the fact that $W_n^2(\Psi) = \|\alpha_n\|_{2,\Psi}^2$ and from the Central Limit Theorem (CLT) in the Hilbert Space $L^2(0,1)$ (see, e.g., Araujo and Giné (1980) p. 205, ex. 1) turned the problem of studying the asymptotic distribution of $W_n^2(\Psi)$ into an easier task.

Goodness-of-fit testing to a family of distributions

We consider in this section the problem of testing whether the underlying d.f. of the sample, F, belongs to a given family of distribution functions, $\mathcal{F}_0$. We will assume that $\mathcal{F}_0$ is a parametric family, i.e., $\mathcal{F}_0 := \{F_\theta : \theta \in \Delta\}$, and that Δ is an open subset of $\mathbb{R}^p$. In the previous section, we considered the GOF procedures based on the measurement of distances between an empirical distribution obtained from the sample and a fixed distribution. A way to adapt this idea for the new setup consists of choosing some adequate estimator $\widehat{\theta}_n$ of θ (assuming the null hypothesis is true) and, then, replacing the fixed distribution by $F_{\widehat{\theta}_n}$.

The use of quadratic statistics based on the empirical d.f. with parameters estimated from the data could provide more powerful tests, just as in the fixed distribution setup. The adaptation of $W_n^2(\Psi)$ to this situation can be easily carried out. Let θ_n be some estimator of θ. Using $\widehat{\theta}_n$, the estimator of θ, we can define the statistics

$$\widehat{W}_n^2(\Psi) = n\int_{-\infty}^{\infty} \Psi(F_{\widehat{\theta}_n}(x))[F_n(x) - F_{\widehat{\theta}_n}(x)]^2\, dF_{\widehat{\theta}_n}(x),$$

and use them as statistical tests, rejecting the null hypothesis when large values of $\widehat{W}_n^2(\Psi)$ are observed. Though, it took a long time until these statistics were considered as serious competitors to the χ^2-test, little was known about these versions of Cramér-von-Mises or Kolmogorov-Smirnov tests until the 50's (see, e.g., Cochran (1952)).

The property exhibited by $W_n^2(\Psi)$ of being distribution free does not carry over to $\widehat{W}_n^2(\Psi)$. If we set $Z_i = F_{\widehat{\theta}_n}(X_i)$, and with $\widehat{G}_n$ denoting the empirical d.f. associated to $Z_1, ..., Z_n$, then, obviously,

$$\widehat{W}_n^2(\Psi) = n \int_{-\infty}^{\infty} \Psi(t)[\widehat{G}_n(t) - t]^2 \, dt,$$

but, unlike in the fixed distribution case, $Z_1, ..., Z_n$ are not i.i.d. uniform r.v.'s.

The first attempt to derive the asymptotic distribution of any statistic of $\widehat{W}_n^2(\Psi)$ type is due to Darling (1955). His study concerns the case where $\Psi \equiv 1$, i.e.,

$$\widehat{W}_n^2 = n \int_{-\infty}^{\infty} [F_n(x) - F_{\widehat{\theta}_n}(x)]^2 \, dF_{\widehat{\theta}_n}(x) = n \int_{-\infty}^{\infty} \Psi(t)[\widehat{G}_n(t) - t]^2 \, dt,$$

assuming that θ was one-dimensional. Let us define

$$\begin{aligned} \widetilde{W}_n^2 &= n \int_{-\infty}^{\infty} \left(F_n(x) - F_\theta(x) - (\widehat{\theta}_n - \theta) \frac{\partial}{\partial \theta} F_\theta(x) \right)^2 \, dF_\theta(x) \\ &= \int_0^1 \left(\sqrt{n}(G_n(t) - t) - T_n g(t) \right)^2 \, dt, \end{aligned}$$

where $T_n = \sqrt{n}(\widehat{\theta}_n - \theta)$ and

$$g(t) = g(t, \theta) = \frac{\partial}{\partial \theta} F_\theta(x)|_{x = F_\theta^{-1}(t)}.$$

Darling's approach was based on showing that, when the underlying distribution of the sample is F_θ, and $\mathcal{F}_0$ and $\widehat{\theta}_n$ satisfying some adequate regularity conditions, then

$$\widehat{W}_n^2 - \widetilde{W}_n^2 = o_{\mathbb{P}}(1).$$

Thus, the asymptotic distribution of $\widehat{W}_n^2$ can be studied through that of $\widetilde{W}_n^2$. Darling showed that the finite-dimensional distributions of $\sqrt{n}(G_n(t) - t) - T_n g(t)$ converge weakly to those of a Gaussian process $Y(t)$ with covariance function $K(s,t) = s \wedge t - st - \phi(t)\phi(s)$, where $\phi(t) = \sigma g(t)$ and σ^2 is the asymptotic variance of T_n. He showed, further, that, under some additional assumptions on $\widehat{\theta}_n$, Donsker's invariance principle could be applied to conclude that

$$\widehat{W}_n^2 \rightsquigarrow \int_0^1 (Y(t))^2 \; dt,$$

and, as in the fixed distribution case, a Karhunen-Loève expansion for $\int_0^1 (Y(t))^2 \; dt$ can provide a good way to tabulate the limiting distribution of $\widehat{W}_n^2$. Sukhatme (1972) extended Darling's result to a multidimensional framework and gave very valuable informations for the Karhunen-Loève expansion of the limiting Gaussian process. Instead of considering the process $\{\sqrt{n}(G_n(t) - t) - T_n g(t)\}_t$, a direct study of the estimated empirical process, $\{\sqrt{n}(\widehat{G}_n(t) - t)\}_t$, could yield the asymptotic distribution of general $\widehat{W}_n^2(\Psi)$.
The empirical process with estimated parameters is

$$\widehat{\alpha}_n^{\widehat{\theta}_n}(x) = \sqrt{n}(F_n(x) - F_{\widehat{\theta}_n}(x)), \; x \in \mathbb{R},$$

where $\widehat{\theta}_n$ is a sequence of estimators. We will assume this sequence to be efficient in the sense that, as $n \to \infty$,

$$\sqrt{n}(\widehat{\theta}_n - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l(X_i, \theta) + o_{\mathbb{P}}(1),$$

where $l(X_1, \theta)$ is centered and has finite second moments. To obtain the null asymptotic distribution of $\widehat{\alpha}_n^{\widehat{\theta}_n}$ we assume that $F = F_\theta$ and write

$$\begin{aligned}
\widehat{\alpha}_n^{\widehat{\theta}_n}(x) &= \sqrt{n}(F_n(x) - F_\theta(x)) - \sqrt{n}(F_{\widehat{\theta}_n}(x) - F_\theta(x)) \\
&= \widehat{\alpha}_n^{F_\theta}(x) + \dot{F}_\theta(x)^T \sqrt{n}(\widehat{\theta}_n - \theta) + o_{\mathbb{P}}(1) \\
&= \widehat{\alpha}_n^{F_\theta}(x) + \dot{F}_\theta(x)^T \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l(X_i, \theta) + o_{\mathbb{P}}(1) \\
&= \widehat{\alpha}_n^{F_\theta}(x) + \dot{F}_\theta(x)^T \int_{\mathbb{R}} l(x, \theta) \; d\alpha_n^{F_\theta}(x) + o_{\mathbb{P}}(1) \\
&= \alpha_n(F_\theta(x)) - H(F_\theta(x), \theta)^T \int_0^1 L(t, \theta) \; d\alpha_n(t) + o_{\mathbb{P}}(1) \\
&= \widehat{\alpha}_n(F_\theta(x)) + o_{\mathbb{P}}(1),
\end{aligned}$$

where α_n is the uniform empirical process, $H(t, \theta) = \dot{F}_\theta(F_\theta^{-1}(t), \theta)$, $L(t, \theta) = l(F_\theta^{-1}(t), \theta)$, $\dot{F}(x, \theta) = \left(\frac{\partial}{\partial \theta_1} F_\theta(x), ..., \frac{\partial}{\partial \theta_p} F_\theta(x)\right)^T$ and

$$\widehat{\alpha}_n(t) = \alpha_n(t) - H(t,\theta)^T \int_0^1 L(s,\theta)\ d\alpha_n(s),\ \ 0 < t < 1,$$

Theorem 1.1.4 *Provided that $H(t,\theta)$ is continuous on $[0,1]$ and that $L(t,\theta)$ is continuous and of bounded variation on $[0,1]$, we can define α_n, together with a sequence of Brownian bridges B_n, such that*

$$\|\widehat{\alpha}_n - \widehat{B}_n\|_\infty = O\left(\frac{\log n}{\sqrt{n}}\right)\ \ a.s.,$$

where $\widehat{B}_n(t) = B_n(t) - H(t,\theta)\int_0^1 L(t,\theta)\ dB_n(t)$. $\widehat{B}_n$ is a centered Gaussian process with covariance

$$\begin{aligned}\widehat{K}(s,t) &= s\wedge t - st - H(t,\theta)\int_0^s L(x,\theta)\ dx - H(s,\theta)\int_0^t L(x,\theta)\ dx \\ &+ H(s,\theta)^T\int_0^t L(x,\theta)L(x,\theta)^T\ dxdH(t,\theta)\end{aligned}$$

Theorem 1.1.4 provides among others, as an easy corollary, the asymptotic distribution of a Cramér-von-Mises $\widehat{W_n^2}$ and Anderson-Darling $\widehat{A}_n^2$ statistics under the null hypothesis. Also, as in the fixed distribution case, quadratic statistics exhibit in general, better power properties than Kolmogorov-Smirnov -type statistics, with $\widehat{A}_n^2$ outperforming $\widehat{W_n^2}$. We recall that A-D statistics are given by

$$\widehat{A}_n^2 = n\int_{-\infty}^{\infty}\frac{\left(F_n(x) - F_{\widehat{\theta}_n}(x)\right)^2}{F_{\widehat{\theta}_n}(x)(1 - F_{\widehat{\theta}_n}(x))}\ dF_{\widehat{\theta}_n}(x)$$

We refer to del Barrio et al (2007) for more details about the univariate goodness-of-fit testing problem. Notice that an alternative GOF procedure is due to Kolmogorov and Smirnov and based on the supremum distance between the real and empirical distribution functions. However, this topic is not in the range of the present thesis.

1.1.7.2 Multivariate case

As in the univariate case, denoting by F the cumulative distribution function (cdf), we want to test

$$H_0 : F = F_0, \text{against } H_1 : F \neq F_0,$$

for a given cdf F_0, or, alternatively,

$$H_0 : F \in \mathcal{F}_0, \text{against } H_1 : F \notin \mathcal{F}_0,$$

where $\mathcal{F}_0 := \{F_\theta : \theta \in \Delta\}$ is a given family of distribution and Δ is an open subset of $\mathbb{R}^p$. One can define the same GOF procedures based on the transformations of the underlying empirical process. We recall that we are concerned here with discussion of Cramér-von-Mises-type statistics. These testing procedures are *universal* (or *omnibus*), in the sense they can be applied independently of the underlying distribution. In other terms, they do not depend on some particular properties of F_0 or of the assumed family $\mathcal{F}_0$. In general, such tests are of primary interest for statistical modelling.

As in the univariate case, we obtain readily multivariate analogues of Theorem 1.1.3 and Theorem 1.1.4.

1.1.7.3 The Copula case

As mentioned earlier copula-based modeling of multivariate distributions is finding extensive applications in fields such as finance and actuarial science. The corresponding important issue of statistical applications of copulas, that is currently drawing a lot of attention, is whether the unknown copula C actually belongs to the chosen parametric copula family or not. More formally, we consider here semiparametric copula models, where the unknown copula C associated to F belongs to a parametric class $\mathcal{C}_0 = \{C_\theta : \theta \in \Delta\}$, where Δ is an open subset of $\mathbb{R}^p$.
Namely, we want to test,

$$H_0 : C \in \mathcal{C}_0, \text{against } H_1 : C \notin \mathcal{C}_0.$$

We treat this problem within a semiparametric framework for copula models because the marginal distributions F_1 and F_2 are are treated as (infinite-dimensional) nuisance parameters (see Genest et al (2009)). One of the most used in practice GOF statistics are based on the weak convergence of the empirical copula process.

The empirical copula $\overline{C}_n$ defined in (1.1.5) (see also (2.2.10) on page 56) is a consistent estimator of the unknown copula C, whether H_0 is true or not. Hence, as suggested in Fermanian (2005), Quessy (2005), and Genest and Remillard (2008), a natural goodness-of-fit test consists of comparing $\bar{C}_n$ with an estimation C_{θ_n} of C obtained assuming that $C \in \mathcal{C}_0$ holds. More precisely, these authors propose to base a test of goodness-of-fit on the empirical process

$$\mathbb{C}_{n,\theta} := n^{1/2}(\overline{C}_n - C_{\theta_n}),$$

According to the large scale simulations carried out in Genest et al. (2009), the most powerful version of this procedure is based on the Cramér-von-Mises statistic

$$T_n := \int_{[0,1]^2} \mathbb{C}^2_{n,\theta}(u_1, u_2) d\overline{C}_n(u_1, u_2).$$

The method of proving the weak convergence of test statistics T_n is based on the similar technique as described in the previous sections. Namely, first, under some regularity conditions, we prove the weak convergence in space $l^\infty([0,1]^2)$ of underlying copula process $\mathbb{C}_{n,\theta}$. In the second place, from the continuity of the functional $x \mapsto \int_0^1 x(t)^2\, dt$ and using the continuous mapping theorem we derive the asymptotic result for T_n.

1.1.8 Contribution to goodness of fit testing of copulas models.

The described GOF procedure based on T_n is valid only for the copula functions which verify **Condition 1** defined via (1.1.11) on page 21. For the copulas which fulfill the weaker **Condition 2** (refer to (1.1.12) on page 24) we propose to base goodness of fit tests on the following statistics

$$S_{1,n} := \int_{[0,1]^2} \mathbb{C}^2_{n,\theta}(u_1, u_2) du_1 du_2.$$

The weak convergence of $S_{1,n}$ is provided by the weak convergence in space $L^2([0,1]^2)$ of the copula process $\mathbb{C}_{n,\theta}$ (see Proposition 2.4.1 on page 62) and by the continuity of functional $L^2([0,1]^2) \ni x \mapsto \int_0^1 x(t)^2\, dt$.

We study also the weighted version of previous test. We introduce the following weighted copula process

$$\mathbb{C}^{w_n}_{n;\theta}(\mathbf{u}) = \sqrt{n}\{\overline{C}_n(\mathbf{u}) - C_{\theta_n}(\mathbf{u})\} \cdot w_{\theta_n}(\mathbf{u}), \quad \mathbf{u} = (u_1, u_2) \in [0,1]^2,$$

where $w_{\theta_n} = c_{\theta_n}$ and c_{θ_n} represents the density function of copula C_{θ_n}. The corresponding GOF statistics are given by

$$S_{2,n} := \int_{[0,1]^2} \left\{ \mathbb{C}_{n;\theta}^{w_n}(u_1, u_2) \right\}^2 du_1 du_2.$$

Under general regularity conditions, the weak convergence of $S_{2,n}$ is entailed by the weak convergence in space $L^2([0,1]^2)$ of the copula process $\mathbb{C}_{n;\theta}^{w_n}$ (see Proposition 2.4.2 on page 64) and by the continuity of functional $L^2([0,1]^2) \ni x \mapsto \int_0^1 x(t)^2 \, dt$.

An approximate p-value for the test based on the above statistic $S_{2,n}$ may be obtained by means of a parametric bootstrap, as follows from Genest and Rémillard (2008). As the sample size increases, the application of parametric bootstrap-based goodness-of- fit tests becomes somewhat prohibitive. In order to circumvent this high computational cost, we propose a fast large-sample testing procedure based on multiplier central limit theorems. The validity of this technique has been proved in Kojadinovic et al (2010). The accuracy of applying the multiplier central limit theorem in the case of statistics $S_{1,n}$ nd $S_{2,n}$ is given by Theorem 2.6.1 and Theorem 2.6.2 on page 80 and 83 respectively.

1.2 General Introduction to Paper II

1.2.1 Introduction to the reserve risk under Solvency I (ultimate view)

In the present section, we introduce the necessary background for the reserve risk under the current regulatory regime - Solvency I.

Solvency I.
The regulatory solvency framework for insurance companies, known under the name of *Solvency I*, was introduced in the early 1970's. The corresponding solvency requirements (which dealt only with a minimum harmonisation directive) were primarily focussed on prudential standards for insurers and did not include requirements for risk management and governance within firms. Solvency I was only completed in the early 1990's with the third generation Insurance Directives. The current European Union (EU) Solvency I regime is considered to have simplistic capital requirements which are not fully representative of the underlying risks faced by the insurers. The corresponding requirements are based on relatively simple factor-based expressions, dealing mainly with premiums and reserves which are used to determine the sufficient level of capital needed to cover risks. This approach potentially ignores large components of overall risk. For instance risks related to investment assets as well as potentially excessive exposures to catastrophe risk or to other heavy tailed risks are not fully taken into account.

The following general introduction is essentially inspired by the work of Merz and Wüthrich (2008b).

General insurance
In the present study we consider the insurance branch known under the name of non-life insurance (Continental Europe), general insurance (Great Britain), and property and casualty insurance (North America). In non-life insurance, claims reserves are often the largest liability and the most volatile item on company's balance sheet (see Figure 1.1). Therefore, given the available information about the past, the prediction of an adequate amount to face the responsibilities assumed by a non-life insurance company as well as the quantification of the uncertainties for reserves are major issues in actuarial practice and science.

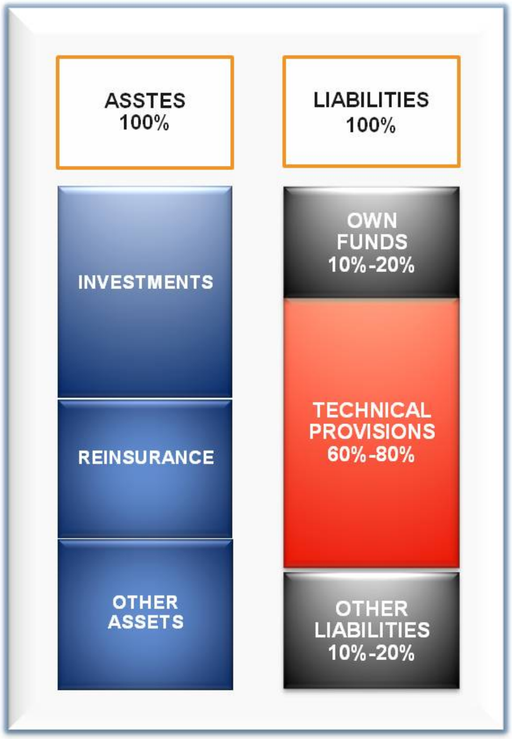

Figure 1.1: Balance Sheet of Non Life insurance company under Solvency I

Claims Settlement Process

A non-life insurance policy is a contract between two parties, namely the insurer and the insured. It provides to the insurer a specific payment (called premium), to the insured a financial coverage against the hazards of well-specified events (or at least a promise that he will get a well-defined financial coverage in case such an event happens). The right of the insured to these amounts (in case the event happens) constitutes a claim by the insured on the insurer. The amount which the insurer is obliged to pay in respect of a claim is known as claim amount or loss amount. Depending on the type of policy, the determination of the proper claim amount can often be very difficult and time consuming. Factors such as reporting delay, recovery process time required for the insurer to obtain all necessary details surrounding the claim, and new developments that can reopen closed claims contribute to this difficulty.

Claims Reserves
Claims Reserve is defined as a type technical reserve or accounting provision of an insurance company, and is established to provide for the future liability for claims which have occurred but which have not yet been settled.

The delay between event and settlement dates means that the insurer must set up "reserves" in respect of those claims still to be settled. The reserves required at any time are the resources needed to meet the costs, as they arise, of all claims not finally settled at that time. The insurer must be able to quantify this liability if it is to assess its financial position correctly, both for statutory and for internal purposes.

There are two different types of claims reserves for past exposures:

- **IBNyR** reserves (incurred but not yet reported): We need to build claims reserves for claims which have occurred before the valuation date, but which have not been reported by the end of the year (i.e. the reporting delay laps into the next accounting years).
- **IBNeR** reserves (incurred but not enough reported): We need to build claims reserves for claims which have been reported before the valuation date, but which have not been settled yet, i.e. we still expect payments in the future, which need to be financed by the already earned premium.

Claims reserves is often referred to the provision for outstanding claims which is one of the main components of technical provisions of insurance company's liabilities (see Figure 1.2).

Uncertainty of reserves
In fact, the actual future loss payments may deviate - sometimes substantially - from the amount that was estimated. Senior managers, shareholders, rating agencies, and regulators all have an interest in knowing the magnitude of these potential deviations since companies with large potential deviations need more capital or reinsurance than other firms with smaller potential deviations. These deviations are often called in literature the loss reserve uncertainty which is a measure of the magnitude of this potential difference between forecast and actual loss payments. Actuarial journals provide several proposed procedures for measuring loss reserve uncertainty. One of the most known is the approach developed by Thomas Mack (1993) based on the deterministic reserving method called Chain Ladder.

Figure 1.2: Technical Provisions in the Solvency I balance sheet

1.2.2 Claims Reserving Notation

1.2.2.1 Data

Large insurance companies often have quite extensive data bases with historical information on incurred claims. Such information can include the numbers of claims reported and settled, the origin year of the events, the paid amounts, the year of the payments and case estimates. The actuary can regularly analyze the data in order to predict the outstanding claims and, hence, the claims reserve. The analysis is typically done in the following way. To begin with, the actuary separates the data into risk homogenous groups such as lines of business, e.g. Motor vehicle liability, Fire and other damage, General liability. A finer segmentation can be applied if the groups or the subgroups contain a sufficient number of observations. The actuary might also choose to divide some group according to the severity of the claims. The large claims can then be reserved according to case estimates while the subgroup consisting of smaller, but frequently occurring, claims can be reserved by some statistical method. When the risk classification is established the actuary usually aggregates the data within the groups into development trapezoids. We now consider such a cumulative trapezoid of paid claims $\{C_{i,j} : (i,j) \in \triangle_{sup}\}$, where $\triangle_{sup} = \{(i,j) : i \in \{1,...,I\}, j \in \{1,...,J\}, i+j \leq I+1\}$. As an example and for the sake of brevity we consider here the claims development triangle, i.e., $I = J = 5$, see Table 1.1. The suffixes i and j of the paid claims refer to the origin year and the payment year, respectively. In addition, the suffix $k = i + j$ is used for the calendar years, i.e., the diagonals of development triangle. If we assume that the claims are settled within the $I = 5$ observed years the purpose of a claims reserving exercise is to predict the last column of unobserved future triangle $\{C_{i,j} : (i,j) \in \triangle_{inf}\}$. , where $\triangle_{inf} = \{(i,j) : i \in \{1,...,I\}, j \in \{1,...,J\}, i+j > I+1\}$.

Accident	Development Year j				
Year i	1	2	3	4	5
1	$C_{1,1}$	$C_{1,2}$	$C_{1,3}$	$C_{1,4}$	$C_{1,5}$
2	$C_{2,1}$	$C_{2,2}$	$C_{2,3}$	$C_{2,4}$	
3	$C_{3,1}$	$C_{3,2}$	$C_{3,3}$		
4	$C_{4,1}$	$C_{4,2}$			
5	$C_{5,1}$				

Table 1.1: Run-off triangle ($I = J = 5$)

Outstanding Loss Liabilities
Let R_i and R denote the *outstanding loss liabilities* for accident year i at time I,

$$R_i = C_{i,J} - C_{i,I-i+1},\ i \in \{1,...,I\},$$

and the total outstanding loss liabilities for all accident years,

$$R = \sum_{i=1}^{I} R_i,$$

respectively.

The prediction of the outstanding loss liabilities R_i and R, as well as quantifying the uncertainty in this prediction, is the classical actuarial claims reserving problem studied at every non-life insurance company. We use the term claims reserves to mean the prediction of the outstanding loss liabilities.

Hence, let $\widehat{R}_i$ and $\widehat{R}$ denote the claims reserves for accident year i at time I,

$$\widehat{R}_i = \widehat{C}_{i,J} - C_{i,I-i+1},\ i \in \{1, ..., I\},$$

and the total claims reserves for aggregated accident years,

$$\widehat{R} = \sum_{i=1}^{I} \widehat{R}_i,$$

respectively, where $\widehat{C}_{i,j}$ is a predictor for $C_{i,j}$.

Why insurance companies use the aggregate data

Due to the inherent random nature of a non-life insurance company's liabilities, the task of setting aside proper provisions to limit the chance of ruin is difficult. The task can be approached on a per claim level. For each reported claim a reserve amount is estimated by claims handlers that is deemed appropriate to cover the remaining payments of the policy. This amount is called the case reserves. For an aggregate portfolio, the case reserves of all reported claims are added to provide an estimate of future liabilities on reported claims. The result is called the claims incurred. However, this omits consideration for claims that have already occurred but have yet to be reported, or so-called incurred but not yet reported (**IBNyR**) claim amounts. One could independently set aside an amount for the **IBNyR**, but in practice the inclusion of the **IBNyR** is done through modelling the aggregate claim amounts. By modelling the aggregate claim amounts, one is able to capture the emerging **IBNyR** pattern.

Prediction Uncertainty

As mentioned above, finding suitable claims reserves, i.e., $\widehat{R}_i$ and $\widehat{R}$, is rather the beginning of the process of reserving and the insurers companies need to assess the variability of these amounts. The important challenge for actuaries is to quantify not only the claims reserves but also the uncertainty of the resulting predictors. In the present work, as in the work of Taylor and Ashe (1983), we quantify the prediction uncertainty with the aid of the most popular such measure, the so-called mean-square error of prediction (MSEP). For predictor $\widehat{C}_{i,I}$ of the ultimate claim amount $C_{i,I}$ of accident year i, the conditional MSEP is defined as

$$msep_{\widehat{C}_{i,I}|D_I}(C_{i,I}) := E\left[\left(\widehat{C}_{i,I} - C_{i,I}\right)^2 |D_I\right].$$

Note that with regards to the conditional MSEP, it does not matter whether one considers the predictor $\widehat{C}_{i,I}$ of the ultimate claim amount or the predictor $\widehat{R}_i$ of the claims reserves of accident year i. Both yield the same result. We adopt the convention of using the predictor of the ultimate claim amount. If the predictor $\widehat{C}_{i,I}$ is D_I-measurable, the conditional MSEP decouples as follows:

$$msep_{\widehat{C}_{iI}|D_I}(C_{iI}) = Var(C_{iI}|D_I) + (E(C_{iI}|D_I) - \widehat{C}_{iI})^2,$$

The first term on the right-hand side of the above equation is called the **conditional process variance** It represents the inherent uncertainty of the underlying model chosen for the observed data. The second term on the right-hand side is called the **conditional estimation error**, it represents the uncertainty in the estimation of the unknown model parameters.

1.2.2.2 The Chain Ladder Method

The chain-ladder method is probably the most popular reserving technique in practice. According to Taylor (2000) its lineage can be traced to the mid-60's. In the meantime, some authors claim that this technique was mentioned in France in the 30's (see Charpentier and Denuit (2005)). At the center of the method are the so-called age to age factors that develop the cumulative claim amounts one period. The name of this method should refer to the chaining of a sequence of age-to-age development factors into a ladder of factors by which one can climb from the observations to date to the predicted ultimate claim cost. The chain-ladder was originally deterministic, but in order to assess the variability of the estimate it has been developed into a stochastic method. This claims reserving method is arguably the most widely used in practice and also goes by the name *the loss development triangle method*.

Chain Ladder method(CL) - model assumptions

The classical actuarial literature often explains the CL claims reserving method as a pure computational algorithm to estimate claims reserves. A distribution-free stochastic model underlying the CL algorithm was proposed by Mack (1993). It was based on the following model assumptions:

(1) The accident years $(C_{i,1},..,C_{i,J})_{1\leq i\leq I}$ are independent

(2) There exist deterministic development factors $f_1, f_2, .., f_{I-1}$ such that

$$E(C_{i,k+1}|C_{i,1},..,C_{i,k}) = f_k C_{i,k}.$$

(3) There exist constants $\sigma_k^2 > 0$ such that for all $1 \le i \le I$ and $1 \le k \le J-1$ we have

$$Var(C_{i,k+1}|C_{i,1},..,C_{i,k}) = \sigma_k^2 C_{i,k}.$$

From assumption (2), a routine calculus on conditional expectation shows that

$$E[C_{i,I}|D_I] = f_{I-1}E[C_{i,I-1}|D_I] = ... = C_{i,I-i+1}\prod_{j=I-i+1}^{I-1} f_j,$$

for all $i \in \{1,...,I\}$, where the factors f_j are called the CL factors and $D_I = \{C_{i,j} : (i,j) \in \triangle_{sup}\}$. Given the observations D_I and CL factors f_j, the above equation gives a recursive algorithm for predicting the ultimate claim amount $C_{i,I}$ by $E[C_{i,I}|DI]$. However, in most practical applications the CL factors f_j are not known and have to be estimated from the data D_I.

Chain Ladder method(CL) - model estimators

- Given the information D_I, the factors f_k are estimated by

$$\widehat{f}_k = \frac{\sum_{i=1}^{I-k} C_{i,k+1}}{\sum_{i=1}^{I-k} C_{i,k}} = \frac{\sum_{i=1}^{I-k} C_{i,k}F_{i,k}}{\sum_{i=1}^{I-k} C_{i,k}}, \text{ for } 1 \le k \le I-1,$$

  where $F_{i,k} := C_{i,k+1}/C_{i,k}$ are often called in actuarial literature the *age-to-age factors*, *individual chain ladders factors* or *link ratios.*

- The variance parameters σ_k^2 are estimated by

$$\widehat{\sigma}_k^2 = \frac{1}{I-k-1}\sum_{i=}^{I-k} C_{i,k}(F_{i,k} - \widehat{f}_k)^2, \text{ for } 1 \le k \le I-2,$$

  and because of lack of data, the last parameter σ_{I-1}^2 is estimated by

$$\widehat{\sigma}_{I-1}^2 = \min(\widehat{\sigma}_{I-2}^4/\widehat{\sigma}_{I-3}^2, \min(\widehat{\sigma}_{I-3}^2, \widehat{\sigma}_{I-2}^2)).$$

1.2.3 Contribution to the reserve risk in Solvency I framework (ultimate view)

The claims reserving estimation is an important process in every general (non-life) insurance company. In the present thesis we provide a general framework for the stochastic chain ladder model. Our general approach can be used to solve many of the practical actuarial problems. Namely, many of actuaries often proceed in the following way: they compute the best estimate of provision (estimation of parameters f_k) by excluding (if needed) some of the age-to-age factors (link ratios $F_{i,k}$), then they use the complete data (all observed link ratios) in the estimation of dispersion of this estimate from its theoretical value (estimation of parameters σ_k^2). They justify the suppression of some of the data from the estimation of claims reserves by the need of monitoring the quality of the observations. The loss of information caused by the selection of data is compensated by the fact that all information (all age-to-age factors $F_{i,k}$) are taking into account in the estimation of its volatility. We provided the technical background for this type of actuarial practice. The results we derived allow to estimate the mean-square error of prediction for claims reserves in the case of using two different sets of age to age factors for estimation of the chain ladder coefficients f_k and the volatility parameters σ_k^2. To the best of our knowledge these results are new and bridge the gap between the theoretical research and actuarial practice. Recall that Thomas Mack derived the similar results (see, e.g., Mack (1999)) in the case where these two sets of coefficients, used in estimation of parameters f_k and σ_k^2 respectively, are the same. Our thesis provided though the important contribution for his result from the practical point of view.

In particular, we derived the estimators for the mean-square error of prediction (MSEP) of the claims reserves in the case when the actuary wants to exclude some of individuals age-to-age factors (link ratios $F_{i,k}$) in the estimation of the chain ladder factors f_k, and to keep all of them in the estimation of the variance parameters σ_k^2.

1.3 General Introduction to Paper III

1.3.1 Introduction to the reserve risk under Solvency II (one-year view)

The fact that several European countries has gone down the route to develop their own solvency regimes can be seen as a strong indicator of a consensus that current regulatory rules (Solvency I) not being sufficient to reflect the capital need of insurance companies.

This is one of the key drivers, together with the lack of risk-based principles within Solvency I , to create a more harmonized and risk-based solvency regulation within the EU. As a response to the above, the new solvency framework has been developed, named Solvency II.

Solvency II.
Solvency II is a fundamental review of the capital adequacy regime for European insurers and reinsurers, planned to take effect progressively from the beginning of January 2016. The main intension of the new directive is to have a harmonized regulation across EU.

Solvency II is based on economic principles for the measurement of assets and liabilities, risk-based capital requirements based on market consistent scenarios, i.e., scenarios under which the valuation of assets and liabilities can be directly verified from the observable market prices. It will be a risk-based system as risk will be measured on consistent principles and capital requirements will depend directly on this. It will bring also the harmonization of asset and liabilities valuation techniques across EU.

The proposed Solvency II framework has three main areas (pillars):

- **Pillar 1**: considers key quantitative requirements including own funds, the calculation of technical provisions and the rules relating to the calculation of Solvency II capital requirements (Solvency capital Requirement-SCR, and Minimum Capital requirement MCR, see Figure 1.3). SCR can be calculated either through an approved (by the local supervisor) full or partial internal model or through European standard formula approach with an option of Undertaking Specific Parameters (USPs). One of the main idea of Pillar I is the fact that the technical provisions should now include a Risk Margin.

- **Pillar 2**: sets out requirements for the governance and risk management of insurers, as well as for the effective supervision of insurers. It does include ORSA (Own Risk Solvency Assessment) which may involve quantitative analysis on a different basis then the Pillar I assessment.
- **Pillar 3**: focuses on disclosure and transparency requirements.

The main purposes of Solvency II.
The insurance industry strongly supports a Solvency II framework which aims to achieve the following:

- Give an incentive to the supervised institutions to measure and properly manage their risks
- Contribute to a better managed and more competitive insurance industry that can better perform its key function of accepting and spreading risk
- Encourage a single European market for financial services
- Enable an institution to absorb significant unforeseen losses and gives reasonable assurance to policyholders

Reserve risk in Solvency II.

As previously mentioned, one of the major risks of non-life insurance company is the reserve risk. Reserve risk concerns the liabilities for insurance policies covering historical years, often simply referred to as the risk in the claims reserve, i.e., the provision for outstanding claims. More precisely, the reserve risk corresponds to the risk that technical provisions set up for claims already occurred at the valuation date will be insufficient to cover these claims. Reserve risk stems from two sources: on the one hand, the absolute level of the claims provisions may be mis-estimated. On the other hand, because of the stochastic nature of future claims payments, the actual claims will fluctuate around their statistical mean value.

As in the Solvency II framework, the time horizon is one year, the reserve risk is only the risk of the technical provisions (in the Solvency II balance sheet) for existing claims needing to be increased within a twelve-month period. Following Ohlsson and Lauzeningks (2009), we define the reserve risk over one-year time horizon as the risk in the one-year run-off result.

If C_1 denotes amount paid during next year, R_0 the opening reserve at the beginning, and R_1 the closing reserve estimate at the end of the year, then the technical run-off result is

$$T = R_0 - C_1 - R_1.$$

The one-year reserve risk is captured by the probability distribution of T, conditioned on the observations by time 0. Notice that T is also the difference between the estimate of the ultimate cost at time 0 (beginning of the accounting year) and at time 1 (end of the accounting year). This is in contrast to the ultimo or full run-off risk, which is described as the risk in $R_0 - C^{\infty}$, where C^{∞} states for the payments over the entire run-off period. In the setup of run-off triangles the random variable T is often called *claims development result* (CDR). In the following, this term will be used to measure the variability of one year reserve risk.

1.3.2 Contribution to the reserve risk in Solvency II framework (one-year view)

In our thesis, we consider the problem of claims reserving in the setup of run-off triangles. This problem is motivated by the need of monitoring the randomness of claims development up to the time when the ultimate claim is finally settled. This aspect of claims reserving relies, typically, on a long-term point of view. This is in contrast with the short term horizon inherent to models describing total risk for an insurance company, such as the one-year risk perspective used in the Solvency II project. The challenge of bridging the gap between these two viewpoints gave rise to some innovative research in the study of reserving process. One of the first papers dealing with the one-year reserve risk was that of Merz and Wüthrich (2008a) (MW). In the special case of a pure Chain-ladder estimate, they provided analytic formulae for the mean-squared error of predictions of the run-off result, referred to as the *claims development result* (CDR). Their methods rely on an extension of the well-known Mack (1993) model.

In the present thesis, we intend to provide a general methodology for measuring the uncertainty of CDR. Our approach largely extends that of MW and differs from it mainly in the assumption on conditional variance of individual development factors $F_{i,k}$ (defined as ratio $C_{i,k+1}/C_{i,k}$). Our extension for measuring the uncertainty of CDR is three-fold. Firstly, we derive several possible estimators of loss development factors f_k. Notice that in the original paper of Merz and Wüthrich (2008a) only the classical chain-ladder estimates of f_k were considered.

Secondly, we allow selecting the link ratios $F_{i,k}$ manually by employing experience, judgment, benchmarks, etc.

Thirdly, we provide the possibility to select different loss ratios for estimation of loss development factors f_k and for estimation of the variance parameters σ_k^2.

Moreover, our general framework can be applied in the case of incomplete run-off triangles. Finally, the presented approach for measuring one-year volatility of reserve risk can be applied within the Solvency II framework to computation of SCR (Solvency Capital Requirement) and Risk Margin (see Figure 1.3) by standard formula, USPs (Undertaking-Specific Parameters) methods or (full/partial) internal models.

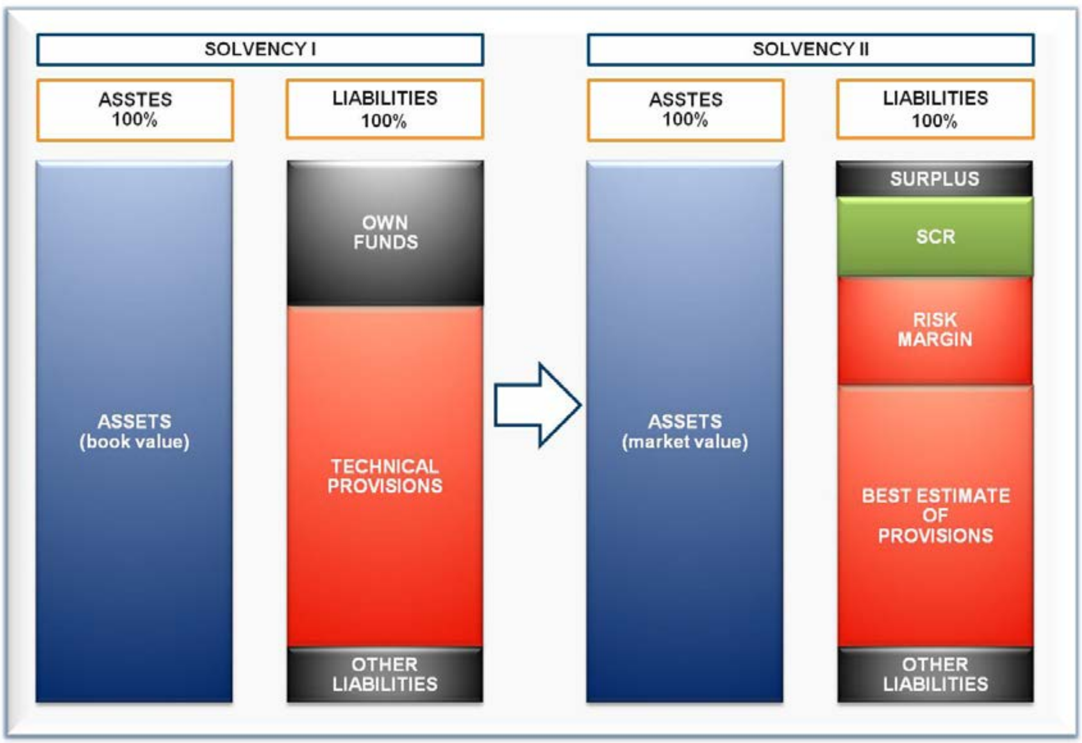

Figure 1.3: Technical Provisions in the Solvency II balance sheet

Chapter 2

Paper I - Weak convergence of weighted empirical copula process in L^p spaces

2.1 Motivation

The weak convergence of the empirical copula processes has been extensively investigated in the literature starting with Rüschendorf (1976)(for convergence in the Skorohod space $D([0,1]^d)$ endowed with a particular metric), and Deheuvels (1981) (in the setup of independent marginals). We refer to Gaenssler and Stute (1987), van der Vaart and Wellner (1996) and Fermanian et al (2004) for extensions of these results in $D([0,1]^d)$, $\ell^\infty([a,b]^d)$ $(0 < a, b < 1)$ and $\ell^\infty([0,1]^d)$, respectively. Their results impose some regularity conditions such as the continuity on $[0,1]^d$ of the first-order partial derivatives of the copula function C. Unfortunately, for a number of commonly used copulas, the condition of continuity of the first-partial derivatives on the unit cube is not satisfied. In the latter, the copula process does not converge weakly with respect to the topology induced by the sup-norm.

Recently, Segers (2012) has shown how to cope with the situation where the first-order partial derivatives fail to be continuous at some boundary points of the hypercube. However, as pointed out in his paper, his arguments do not apply to some useful examples of copulas, such as the Archimedean and extreme-value copulas.

One possible way to address this difficulty is to consider the weak convergence of the empirical copula processes with respect to some weaker metric than considered by Segers (2012). In this paper, we establish a convergence result in the topology induced by the L^p-norm. We work under nonrestrictive regularity conditions which reduce to assuming existence and continuity of the first-partial derivatives almost everywhere on the unit cube. We generalize this property to weighted empirical copula processes with weight functions in L^p. By the continuous mapping theorem we obtain, as a consequence of our results, the weak convergence of weighted Cramér-von-Mises-type statistics.

Organization of the paper. In Section 2.2 we present our notation and introduce the weighted empirical copula process. Our main result is given in Section 2.3. In Section 2.4 we provide some statistical applications based on our main result. Proofs are postponed to Section 2.5. The auxiliary results are presented in Section 2.6. Finally, some indications of future research directions are given in Appendix A.

2.2 Introduction

2.2.1 Copula Function

Let $\mathbf{X} = (X(1), ..., X(d))$ be a random vector in $\mathbb{R}^d$ with d-variate (right-continuous) cumulative distribution function [cdf] F. Denote by $F_1, ..., F_d$ the margins (right-continuous cdf's) of $\mathbf{X}$. For each $j = 1, ..., d$ denote by Q_j the (left-continuous) quantile function [qf] pertaining to F_j , and extended by continuity to $[0, 1]$. For an arbitrary (right-continuous) cdf K, the corresponding (left-continuous) qf is defined by

$$K^{-1}(u) := \begin{cases} \inf\{\, x \in \mathbb{R} \,|\, K(x) \geq u \,\}, & 0 < u \leq 1, \\ \sup\{\, x \in \mathbb{R} \,|\, K(x) = u \,\}, & u = 0, \end{cases}$$

where $\inf \emptyset := \infty$ and $\sup \emptyset := -\infty$.

Consider a sequence $\mathbf{X}_i = (X_i(1), ..., X_i(d))$, $i = 1, 2, ..., n$, of independent random copies of $\mathbf{X}$ with common cdf F.
Let μ be the finite Borel measure and $\mathcal{B}$ be the Borel σ-algebra on $[0, 1]^d$. The space $L^p([0, 1]^d, \mu) := L^p([0, 1]^d, \mathcal{B}, \mu)$ is equipped with the usual norm $|f|_p$, where $|f|_p^p = \int_S |f|^p \, d\mu$ and $1 \leq p < \infty$. For $p = \infty$, the space $L^\infty([0, 1]^d, \mu)$ is defined as follows. We start with the set of all measurable functions from $[0, 1]^d$ to $\mathbb{R}$ which are **essentially** bounded, i.e. bounded up to a set of measure zero. Again two such functions are identified if they are equal μ-almost everywhere. Denote this set by $L^\infty([0, 1]^d, \mu)$. For f in $L^\infty([0, 1]^d, \mu)$, its essential supremum serves as an appropriate norm:

$$\|f\|_\infty := \inf\{M \geq 0 : |f(x)| \leq M \text{ for } \mu\text{-almost every } x\}. \tag{2.2.1}$$

The function C is called a *copula function* if C is a cdf on $[0, 1]^d$ with uniformly distributed margins. The copula function C is associated to F if C satisfies, for all $\mathbf{x} = (x_1, ..., x_d) \in \mathbb{R}^d$, the so-called *fundamental identity* (see, e.g., (Deheuvels, 2009, p.124) and the references therein)

$$F(\mathbf{x}) = F(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d)). \tag{2.2.2}$$

In general, the copula function C of F is not unique. A necessary and sufficient condition for uniqueness of C fulfilling (2.2.2) is the continuity of the margins $F_1, ..., F_d$. This condition is assumed throughout in the sequel.

Copulas have many useful properties, such as uniform continuity (see, e.g., Nelsen (2006)). Moreover, it can be shown that every copula function C is bounded by the so-called *Fréchet-Hoeffding bounds*,

$$W_d(u_1, ..., u_d) \leq C(u_1, ..., u_d) \leq M_d(u_1, ..., u_d), \quad \text{for} \quad \mathbf{u} = (u_1, ..., u_d) \in [0,1]^d,$$

where $W_d := \max(u_1 + ... + u_d - d + 1, 0)$ and $M_d := \min(u_1, ..., u_d)$. For $d = 2$, the upper and lower *Fréchet-Hoeffding bounds* W_2 and M_2 are copulas. On the other hand, for $d \geq 3$, M_d is a copula, but not W_d (see, e.g., (Nelsen, 2006, p.47)).

For $j \in 1, ..., d$, let $\mathbf{e_j}$ be the j-th coordinate vector in $\mathbb{R}^d$. For $\mathbf{u} = (u_1, ..., u_d) \in [0,1]^d$ such that $0 < u_j < 1$, let

$$C_j(\mathbf{u}) := \lim_{h \to 0} \frac{C(\mathbf{u} + h\mathbf{e_j}) - C(\mathbf{u})}{h}, \tag{2.2.3}$$

be the j-th first-order partial derivative of C, whenever this limit exists. We extend this definition to $[0,1]^d$ by setting

$$C_j(\mathbf{u}) = \begin{cases} \limsup_{h \downarrow 0} \dfrac{C(\mathbf{u} + h\mathbf{e_j})}{h} & \text{if } \mathbf{u} \in [0,1]^d, u_j = 0, \\ \limsup_{h \downarrow 0} \dfrac{C(\mathbf{u}) - C(\mathbf{u} - h\mathbf{e_j})}{h} & \text{if } \mathbf{u} \in [0,1]^d, u_j = 1. \end{cases}$$

Note that, the above definition entails that, C_j is defined everywhere on $[0,1]^d$, with $C_j(\mathbf{u}) = 0$ whenever $u_i = 0$ for some $i \neq j$.
For any copula function C, letting $\mathbf{u} = (u_1, ..., u_d) \in [0,1]^d$, define $\mathcal{D}_C \subset [0,1]^d$ as follows:

$$\mathcal{D}_C = \bigcap_{j=1}^{d} \left\{ \mathbf{u} \in [0,1]^d : \ C_j(\mathbf{u}) \text{ is defined and continuous at } \mathbf{u} \ \right\}. \tag{2.2.4}$$

Consider the set of copulas defined by

$$\mathcal{C} := \{C \text{ is copula function such that } \mu(\mathcal{D}_C) = 1\}. \tag{2.2.5}$$

The assumption that $\mu(\mathcal{D}_C) = 1$ turns out not to be restrictive in practice. However there exist copulas that do not satisfy this condition. In particular, we can construct an example of copula function such that, for every $\varepsilon > 0$, its first-partial derivatives are not continuous on a set of Lebesgue measure $1 - \varepsilon$ (see Section 2.6.6).

Example 2.2.1 (d=2) *It is readily checked that:* $\mathcal{D}_{M_2} = [0,1]^2 \setminus \{(u,u) : u \in [0,1]\}$ *and* $\mathcal{D}_{W_2} = [0,1]^2 \setminus \{(u, 1-u) : u \in [0,1]\}$.

2.2.2 Empirical Copula Functions

Consider a sequence $\mathbf{X}_i = (X_i(1), ..., X_i(d))$, $i = 1, 2, ..., n$, of independent random copies of $\mathbf{X}$ with common cdf F. Let $\mathbf{U}_i = (U_i(1), ..., U_i(d)) = (F_1(X_i(1)), ..., F_d(X_i(d)))$, $i = 1, 2, ..., n$, be a sequence of independent random copies of $\mathbf{U} = (U(1), ..., U(d)) = (F_1(X(1)), ..., F_d(X(d)))$. It is well known that the cdf of $\mathbf{U}$ is the copula function C associated to F. Here and elsewhere, order relations on vectors are to be interpreted componentwise, namely, we set $\mathbf{x} \leq \mathbf{y}$, when $\mathbf{x} = (x_1, ..., x_d)$ and $\mathbf{y} = (y_1, ..., y_d)$ fulfill $x_j \leq y_j$, for all $j = 1, .., d$. Setting $\mathbb{1}_A$ for th indicator function of A, we define the following empirical distribution functions, for $\mathbf{x} = (x_1, ..., x_d) \in \mathbb{R}^d$ and $\mathbf{u} = (u_1, ..., u_d) \in (0,1)^d$,

$$F_n(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{\mathbf{X}_i \leq \mathbf{x}\}}, \quad F_{nj}(x_j) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i(j) \leq x_j\}}. \tag{2.2.6}$$

$$G_n(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{\mathbf{U}_i \leq \mathbf{u}\}}, \quad G_{nj}(u_j) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{U_i(j) \leq u_j\}}. \tag{2.2.7}$$

By continuity of $F_1, ..., F_p$ for each $j = 1, .., d$, the order statistics

$$X_{1,n}(j) < ... < X_{n,n}(j) \quad \text{and} \quad U_{1,n}(j) < ... < U_{n,n}(j),$$

pertaining to $X_1(j), ..., X_n(j)$ and $U_1(j), ..., U_n(j)$, respectively, are distinct with probability 1. The marginal empirical quantile functions associated to F_{nj} and G_{nj} are given by

$$F_{nj}^{-1}(u_j) = \begin{cases} X_{k,n}(j) & \text{if}(k-1)/n < u_j \leq k/n, \\ -\infty & \text{if} \quad u = 0, \end{cases}$$

and

$$G_{nj}^{-1}(u_j) = \begin{cases} U_{k,n}(j) & \text{if} \quad (k-1)/n < u_j \leq k/n, \\ 0 & \text{if} \quad u = 0. \end{cases}$$

In view of the characterization (2.2.2), we define an *empirical copula function* associated to F_n, as any copula function C_n (namely a cdf on $[0,1]^d$ with uniformly distributed margins) fulfilling the *fundamental identity*

$$F_n(\mathbf{x}) = F_n(x_1, ..., x_d) = C_n(F_{n1}(x_1), ..., F_{nd}(x_d)) \quad \text{for} \quad \mathbf{x} = (x_1, ..., x_d) \in \mathbb{R}^d. \tag{2.2.8}$$

Let $\mathbf{r}_n(j) = (r_{1,n}(j), ..., r_{n,n}(j))$, $j = 1, ..., d$, be the *marginal rank vector* which is the uniquely defined permutation of $\{1, ..., n\}$, such that

$$X_i(j) = X_{r_{i,n}(j),n}(j) \quad \text{for} \quad i = 1, ..., n, j = 1, ..., d.$$

The following proposition, stated in (Deheuvels, 2009, p.129) as Proposition 1.5, (see also Proposition (2.1) in (Deheuvels, 1979, p.276)) gives a constructive characterization of all copula function C_n fulfilling (2.2.8).

Proposition 2.2.1 *A copula function C_n defines, via (2.2.8), an empirical copula function associated to F_n if and only if, for each integer vector $\mathbf{k} = (k_1, ..., k_d) \in \{0, ..., n\}^d$,*

$$C_n\Big(\frac{k_1}{n}, ..., \frac{k_d}{n}\Big) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{r_{i,n}(1)\leq k_1, ..., r_{i,n}(d)\leq k_1 d\}}.$$

As follows from Proposition 2.2.1 that the empirical copula function C_n associated to F_n is not unique.This is a straightforward consequence of discontinuity of the marginal cdf's $F_{n1}, ..., F_{nd}$.

Denote by $\stackrel{\mathcal{L}}{=}$ equality in distribution. For an arbitrary $d \geq 1$, let $\ell^\infty([0,1]^d)$ denotes the space of all uniformly bounded functions φ on $[0,1]^d$ into $\mathbb{R}$ equipped with the topology induced by sup-norm, namely, $|\varphi|_\infty := \sup_{\mathbf{u}\in[0,1]^d} |\varphi(\mathbf{u})|$.

The following remark states a useful properties of empirical copula functions.

Remark 2.2.1

(a) *The law of any empirical copula function C_n depends upon F only through the associated copula function C. In particular, a copula function C_n is an empirical copula function associated to F_n in (2.2.6) if and only if C_n*

is an empirical copula function associated to G_n in (2.2.7). Therefore, in the study of asymptotic behavior of C_n, in the sense of weak convergence for example, we may assume that $F = C$. It means that, without loss of generality, we can work directly on a $\mathbf{U}_1, ..., \mathbf{U}_n$ with cdf C.

(b) Let $C_n^{(1)}$ and $C_n^{(2)}$ be any two empirical copula function associated to F_n. Then

$$|C_n^{(1)} - C_n^{(2)}|_\infty \leq \frac{d-1}{dn}.$$

Proof.
(a). See, e.g., the proof of Corollary 1.1 in (Deheuvels, 2009, p.131) and Lemma 2.1 in (Deheuvels, 1979, p.277)
(b). See, e.g., the proof of Proposition 1.7 in (Deheuvels, 2009, p.132-133).
□

Consider the following empirical distribution functions, for all $\mathbf{u} = (u_1, ..., u_d) \in \mathbb{R}^d$

$$\widetilde{C}_n(\mathbf{u}) = F_n(F_{n1}^{-1}(u_1), ..., F_{nd}^{-1}(u_d)), \tag{2.2.9}$$

$$\overline{C}_n(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{F_{n1}(X_i(1)) \leq u_1, ..., F_{nd}(X_i(d)) \leq u_d\}}. \tag{2.2.10}$$

The forthcoming proposition gives some useful properties of $\widetilde{C}_n$ and $\overline{C}_n$.

Proposition 2.2.2

(a) Neither $\widetilde{C}_n$ nor $\overline{C}_n$ is a copula function.

(b) Both $\widetilde{C}_n$ and $\overline{C}_n$ fulfill the fundamental identity (2.2.8).

(c) $\overline{C}_n$ is a cdf on $[0,1]^d$.

(d) For any empirical copula function C_n associated to F_n, $|\widetilde{C}_n - C_n|_\infty = \frac{1}{n}$.

(e) We have $|\widetilde{C}_n - \overline{C}_n|_\infty \leq \frac{d}{n}$.

(f) Set $a_n = \sqrt{\frac{n}{\log\log n}}$. *If* $b_n = o(a_n)$, *then, for any empirical copula function* C_n *associated to* F_n

$$\lim_{n\to\infty} b_n |C_n - C|_\infty = 0 \quad a.s.$$

Proof.
For (a) we observe that the marginals $\widetilde{C}_n$ and $\overline{C}_n$ are not uniform. The proof of (b) and (c) are archived by straightforward computations. For (d) we invoke of Proposition 1.8 in (Deheuvels, 2009, p.134). We have, for (e)

$$|\widetilde{C}_n - C_n|_\infty \leq \max_{1\leq i_1,\ldots,i_d\leq n} \left|\widetilde{C}_n\Big(\frac{i_1}{n},\ldots,\frac{i_d}{n}\Big) - \widetilde{C}_n\Big(\frac{i_1-1}{n},\ldots,\frac{i_d-1}{n}\Big)\right| \leq \frac{d}{n}.$$

Finally, for (f), we use the proof of Theorem (3.1) in (Deheuvels, 1979, p.277). □

Remark 2.2.2 *The law of* $\widetilde{C}_n$ *and* $\overline{C}_n$ *depends upon* F *only through the associated copula function* C *(see,e.g., (Tsukahara, 2005, p.358)). Hence* $\widetilde{C}_n \stackrel{\mathcal{L}}{=} G_n(G_{n1}^{-1}(u_1),\ldots,G_{nd}^{-1}(u_d))$ *and* $\overline{C}_n \stackrel{\mathcal{L}}{=} \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{G_{n1}(U_i(1))\leq u_1,\ldots,G_{nd}(U_i(d))\leq u_d\}}$.

2.2.3 Empirical Copula Processes

We define the *empirical copula process* by setting

$$\mathbb{C}_n(\mathbf{u}) := n^{1/2}(\widetilde{C}_n(\mathbf{u}) - C(\mathbf{u})), \quad \text{for} \quad \mathbf{u} \in [0,1]^d, \tag{2.2.11}$$

where the empirical function $\widetilde{C}_n$ is defined by (2.2.9). The above definition of *empirical copula process* is used by many authors (see, e.g., Tsukahara (2005), Deheuvels (2009) and Segers (2012)). Note that, for each of the possible copula functions C_n associated to F_n, the process $\{n^{1/2}(C_n(\mathbf{u}) - C(\mathbf{u})) : \mathbf{u} \in [0,1]^d\}$ could also be used to define the *empirical copula process*. Unfortunately, this process(see, e.g., Proposition 2.2.1) is not uniquely defined. On the other hand, the empirical function $\widetilde{C}$ is a "good" approximation (refer to Proposition 2.4.1 (d), (f)) of any empirical copula function C_n in the sense that $|\mathbb{C}_n - n^{1/2}(C_n - C)|_\infty \leq n^{-1/2}$ (see Proposition 2.2.2 (d)).

In view of Remark 2.2.2 we have that $\mathbb{C}_n \stackrel{\mathcal{L}}{=} n^{1/2}(G_n(G_{n1}^{-1},...,G_{nd}^{-1})-C)$ and we may define the following empirical processes, associated to the empirical distribution functions G_n and G_{nj} , by

$$\alpha_n(\mathbf{u}) := n^{1/2}(G_n(\mathbf{u}) - C(\mathbf{u})), \quad \alpha_{nj}(u_j) := n^{1/2}(G_{nj}(u_j) - u_j), \quad (2.2.12)$$

for $j \in \{1, ..., d\}$, $\mathbf{u} \in [0,1]^d$ and $u_j \in [0,1]$. Recall that $\rightsquigarrow$ denotes weak convergence, with the meaning of Definition 1.3.3 from page 17 in van der Vaart and Wellner (1996)(see also Definition 1.1.3 on page 18). Now we invoke the following well known result on the weak convergence of the empirical process α_n in $\ell^\infty([0,1]^d)$. In view of multivariate version of Donsker's theorem (readily checked from Theorem 1.1.3 on page 28), we have that

$$\alpha_n \rightsquigarrow \mathbb{B} \quad \text{as} \quad n \to \infty,$$

where the limit process $\mathbb{B}$ is a *C-Brownian bridge*, i.e. a (tight) centered Gaussian process, with covariance function,

$$Cov[\mathbb{B}(\mathbf{u}), \mathbb{B}(\mathbf{u}')] = C(\mathbf{u} \wedge \mathbf{u}') - C(\mathbf{u})C(\mathbf{u}'), \quad \text{for} \quad \mathbf{u}, \mathbf{u}' \in [0,1]^d, \quad (2.2.13)$$

where $\mathbf{z} \wedge \mathbf{v} = (\min(z_1, v_1), ..., \min(z_d, v_d)) \in [0,1]^d$.
If we assume that the first-order partial derivatives C_j exist and are continuous on $[0,1]^d$, then $\mathbb{C}_n \rightsquigarrow \mathbb{B}^*$ in $\ell^\infty([0,1]^d)$, as $n \to \infty$ (see, e.g., Fermanian et al (2004) and Tsukahara (2005)), where the limit process is a centered Gaussian process defined by

$$\mathbb{B}^*(\mathbf{u}) := \mathbb{B}(\mathbf{u}) - \sum_{j=1}^{d} C_j(\mathbf{u})\mathbb{B}(\mathbf{1}, u_j, \mathbf{1}), \quad \text{for} \quad \mathbf{u} \in [0,1]^d, \quad (2.2.14)$$

where $\mathbb{B}(\mathbf{1}, u_j, \mathbf{1}) := \mathbb{B}(1, ..., 1, u_j, 1, ..., 1)$, the variable u_j appearing at the j-th entry.

2.2.4 Weighted Empirical Copula Process

Let $w : [0,1]^d \to \mathbb{R}$ be any measurable function. For any function $f \in L^p\left([0,1]^d, \mu\right)$, define by wf the product of functions w and f. Making use of this notation, we may define the weighted processes $w\mathbb{C}_n$ and $w\mathbb{B}^*$.

Remark 2.2.3 *For $C \in \mathcal{C}$, the Gaussian process $\mathbb{B}^*$ in (2.2.14) is properly defined on the set $\mathcal{D}_C$.*

Proposition 2.2.3 *If the weight function w is such that*

$$\int_{[0,1]^d} \left(\mathbb{E}\left[w(\mathbf{u})\mathbb{B}^*(\mathbf{u})\right]^2\right)^{p/2} d\mu(\mathbf{u}) < \infty, \tag{2.2.15}$$

then $w\mathbb{B}^*$ is almost surely in $L^p\left([0,1]^d, \mu\right)$.

Proof. See, e.g., Deheuvels (2005), and the references therein. □

2.3 Main Result

We seek conditions on the weight function w, which ensure weak convergence of the process $w\mathbb{C}_n$ to a limiting Gaussian process $w\mathbb{B}^*$ (this property is denoted by $w\mathbb{C}_n \rightsquigarrow w\mathbb{B}^*$). The proof of the following result is based on a decomposition of the empirical copula process (see, e.g., Gaenssler and Stute (1987) and Tsukahara (2005)). This result can also be proven by functional delta-methods by the arguments in section 3.9.4.4 in (van der Vaart and Wellner, 1996, p.389).

Theorem 2.3.1 *We assume that $C \in \mathcal{C}$ and F have continuous marginal distribution functions. If, for $1 \leq p < \infty$,*

$$w \in L^p\left([0,1]^d, \mu\right), \tag{2.3.1}$$

then the weighted empirical copula process $w\mathbb{C}_n$, converges weakly in $L^p\left([0,1]^d, \mu\right)$ to the Gaussian process $w\mathbb{B}^*$.

Remark 2.3.1 *We note that the assumption (2.3.1) of Theorem 2.3.1 implies (2.2.15).*

Remark 2.3.2 *For both of the* Fréchet-Hoeffding bounds *W_2 and M_2, the limiting process $\mathbb{B}^*$ in Theorem 2.3.1 is such that*

$$P\{\mathbb{B}^*(u,v) = 0\} = 1 \ \ a.e. \ \ on \ [0,1]^2.$$

This follows from Cauchy-Schwarz inequality when applied to the covariance function of $\mathbb{B}^*$, and the readily established fact that $Var(\mathbb{B}^*(u,v)) = \mathbb{E}\left\{[\mathbb{B}^*(u,v)]^2\right\} = 0$ a.e. on $[0,1]^2$.

In view of Remark 2.3.2 we may ask the question whether the limiting process $\mathbb{B}^*$ is always trivial whenever the support of copula function is of Lebesgue measure zero. The following example gives the negative answer to that question.

Example 2.3.1 (d=2) *Set $\theta \in (0,1)$ and let consider a copula function $C(\cdot,\cdot,\theta)$, defined as a convex combination of the* Fréchet-Hoeffding bounds*, W_2 and M_2, namely*

$$C(u,v,\theta) = \theta M_2(u,v) + (1-\theta)W_2(u,v).$$

From (2.6.8) it is readily to show that, for $u \le v$ and $v \le 1-u$, $Var(\mathbb{B}^*(u,v)) = \theta(1-\theta)u$.

Remark 2.3.3 *It is natural to ask whether the condition (2.3.1) in Theorem 2.3.1 can be weakened or not. We conjecture that the techniques of Mason (1984) (see Theorem 1, Corollary 2 and Corollary 3) should be helpful in this setup. Observe that, Proposition 2.2.3, condition (2.2.15) is the minimal possible assumption which can be imposed upon w.*

2.4 Statistical Applications

In the sequel of statistical applications, we consider the estimator $\overline{C}_n$ instead of $\widetilde{C}_n$ in definition of the GOF statistics (refer to (2.2.9) and (2.2.10) respectively). The reason is that $\overline{C}_n$ is easier to compute than $\widetilde{C}_n$ and the difference between these two empirical functions is of order $1/n$ (see Proposition 2.2.2 on page 56). The empirical function $\overline{C}_n$ is often called the *càdlàg* version of the empirical copula function (see e.g., (Genest et al, 2009, p.201) and (Fermanian et al, 2004, p.854)).

Recall that $\widetilde{C}_n$ is used in definition of empirical copula process $\mathbb{C}_n$ (see 2.2.11). In the present section, the weak convergence is considered in the L^2 space with respect to the Lebesgue measure on $[0,1]^d$. This means that we apply our main result (Theorem 2.3.1 on page 59) with $p = 2$.

2.4.1 Goodness-of-fit test to a specific copula

One of the simplest goodness-of-fit problem consists of testing fit to a single copula function, namely, testing the null hypothesis $H_0 : C = C_0$ against

$H_1 : C \neq C_0$. We define the Cramér-von-Mises statistic

$$S_{0,n} := \int\limits_{[0,1]^d} \overline{\mathbb{C}}_n^2(\mathbf{u})\, dC_0(\mathbf{u}),$$

$$\overline{\mathbb{C}}_n(\mathbf{u}) := n^{1/2}(\overline{C}_n(\mathbf{u}) - C(\mathbf{u})), \quad \text{for} \quad \mathbf{u} \in [0,1]^d. \tag{2.4.1}$$

Under assumption that copula function C_0 admits a density c_0, this GOF statistic is expressed by

$$S_{0,n} := \int\limits_{[0,1]^d} \overline{\mathbb{C}}_n^2(\mathbf{u})\, dC_0(\mathbf{u}) = \int\limits_{[0,1]^d} \overline{\mathbb{C}}_n^2(\mathbf{u}) \cdot c_0(\mathbf{u})\; d\mathbf{u},$$

and converges weakly to a Gaussian quadratic functional. This follows from the fact that $|\overline{\mathbb{C}}_n - \mathbb{C}_n|_\infty = O(1/\sqrt{n})$ (see Proposition 2.2.2 on page 56) and from Theorem 2.3.1 when applied with the weight function $w = c_0^{1/2}$, and making use of continuous mapping theorem applied to the map $L^2\left([0,1]^d\right) \ni f \mapsto |f|_2^2$, which obviously is continuous.

2.4.2 Goodness-of-fit tests for a family of copulas

We consider here semiparametric copula models, where the unknown copula C associated to F belongs to a parametric class $\mathcal{C}_0 = \{C_\theta : \theta \in \Delta\}$, where Δ is an open subset of $\mathbb{R}^p$. In the literature (see, e.g., Genest et al (2009) and Kojadinovic et al (2010)), the following procedure of testing $H_0 : C \in \mathcal{C}_0$, against $H_1 : C \notin \mathcal{C}_0$, is commonly used. The process

$$\mathbb{C}_{n,\theta} := n^{1/2}(\overline{C}_n - C_{\theta_n}), \tag{2.4.2}$$

is used as a goodness-of-fit process, where θ_n is a consistent estimator of θ and $\overline{C}_n$ is defined via (2.2.9).

Two of the most popular rank-based estimation methods involve the inversion of a consistent estimator of a moment of the copula. The two best-known moments are Spearman's rho and Kendall's tau. As an example, for a bivariate copula C_θ, these are respectively given by

$$\rho_\theta = 12 \int_{[0,1]^2} C_\theta(u,v) dudv - 3 \text{ and } \tau_\theta = 4 \int_{[0,1]^2} C_\theta(u,v) dudv - 1$$

We assume that the bivariate copula family $\mathcal{C}_0$ is such that the functions ρ et τ are one-to-one. Consistent estimators θ_n of θ are then given by $\theta_{n,\rho} = \rho^{-1}(\rho_n)$ and $\theta_{n,\tau} = \tau^{-1}(\tau_n)$, where ρ_n and τ_n are the sample versions of Spearman's rho and Kendall's tau, respectively.
We define a Cramér-von-Mises type statistic by setting

$$S_{1,n} := \int_{[0,1]^d} \{\mathbb{C}_{n,\theta}(\mathbf{u})\}^2 d\mathbf{u}.$$

The convergence of $S_{1,n}$ is established by the following Proposition 2.4.1. Throughout, we assume that $w \in L^2\left([0,1]^d\right)$ and

H.1 For all $\theta \in \Delta$, $w \cdot \{n^{1/2}(\overline{C}_n - C_\theta)\}$ and $\Theta_n := n^{1/2}(\theta_n - \theta)$ jointly weakly converge to $(w\mathbb{B}^*, \Theta)$ in $L^2\left([0,1]^d\right) \otimes \mathbb{R}^p$.

H.2 For all $\theta \in \Delta$ and as $\epsilon \downarrow 0$,

$$\sup_{\|\theta^* - \theta\| < \varepsilon} \left\{ \sup_{\mathbf{u} \in [0,1]^d} \left| \dot{C}_{\theta^*}(\mathbf{u}) - \dot{C}_\theta(\mathbf{u}) \right| \right\} \to 0,$$

where

$$\dot{C}_\theta = \nabla_\theta C_\theta := \left(\frac{\partial}{\partial \theta_1} C_\theta, ..., \frac{\partial}{\partial \theta_p} C_\theta \right)^T. \qquad (2.4.3)$$

Proposition 2.4.1 *Under H.1, H.2 and under the assumptions of Theorem 2.3.1 , the goodness-of-fit process $w\mathbb{C}_{n,\theta}$ converges weakly in $L^2\left([0,1]^d\right)$ to the centered Gaussian process $w\mathbb{B}^{**}$ defined via*

$$w\mathbb{B}^{**} := w(\mathbb{B}^* - \dot{C}_\theta^T \Theta).$$

Proof. See Section 2.5.2.□
Assumptions H.1 and H.2 are very similar to the hypotheses A.2 and A.3 used by (Kojadinovic et al, 2010, p.4) (see also Quessy (2005) and Berg and Quessy (2009)).

We infer from Proposition 2.4.1 taken with the weight function $w \equiv 1$ and from the continuous mapping theorem as in Section 2.4.1 that the statistic $S_{1,n}$ converges weakly to the corresponding quadratic functional of the Gaussian process.

Remark 2.4.1 *From the point of view of statistical applications the statistic*

$$T_n := \int_{[0,1]^d} \{\mathbb{C}_{n;\theta}(\mathbf{u})\}^2 \, d\overline{C}_n(\mathbf{u}) = \sum_{i=1}^{n} \{\mathbb{C}_{n;\theta}(F_{n1}(X_i(1)), ..., F_{nd}(X_i(d)))\}^2, \tag{2.4.4}$$

appears as an interesting alternative to $S_{1,n}$*.*
It is well known fact that, under nonrestrictive smoothness condition (see ***Condition 1*** *defined via (1.1.11) on page 21 and Segers (2012)), the statistic* T_n *converge weakly to the corresponding quadratic functional of the Gaussian process.*
However, the topology induced by the L^2 *norm seems to be to weak to infer the convergence of the statistic* T_n *from the continuous mapping theorem and our Proposition 2.4.1.*

2.4.3 Goodness-of-fit tests based on random weighted copula processes

We consider here semiparametric copula models, where the unknown copula C associated to F belongs to a parametric class $\mathcal{C}_0 = \{C_\theta : \theta \in \Delta\}$, where Δ is an open subset of $\mathbb{R}^p$.
The process

$$\mathbb{C}^{w_n}_{n;\theta}(\mathbf{u}) := \sqrt{n}\{\overline{C}_n(\mathbf{u}) - C_{\theta_n}(\mathbf{u})\} \cdot w_{\theta_n}(\mathbf{u}), \quad \mathbf{u} \in [0,1]^d, \tag{2.4.5}$$

may be used as a goodness-of-fit process, where θ_n is a consistent estimator of θ, w_{θ_n} is a random weight function and $\overline{C}_n$ is defined via (2.2.10). The weak convergence of $\mathbb{C}^{w_n}_{n;\theta}$ to a centered Gaussian process in $L^2\left([0,1]^d\right)$ is provided by Proposition 2.4.3 below. Throughout, we assume that

H.1 For all $\theta \in \Delta$, $w_\theta \cdot \{\sqrt{n}(\overline{C}_n - C_\theta)\}$ and $\sqrt{n}(\theta_n - \theta)$ jointly weakly converge to $(w_\theta \mathbb{B}^*, \Theta)$ in $L^2\left([0,1]^d\right) \otimes \mathbb{R}^p$.

H.2 For each $\theta \in \Delta$, as $\epsilon \downarrow 0$,

$$\sup_{\|\theta^* - \theta\| < \varepsilon} \sup_{\mathbf{u} \in [0,1]^d} \left| \dot{C}_{\theta^*}(\mathbf{u}) - \dot{C}_\theta(\mathbf{u}) \right| \to 0,$$

where $\dot{C}_\theta$ is defined in (2.4.3).

H.3 As $n \to \infty$,$w_{\theta_n} \to w_\theta$ in probability in $L^2\left([0,1]^d\right)$,

H.4 For all $\theta \in \Delta$, $|w_\theta|_2 < \infty$.

Proposition 2.4.2 *Under H.1-H.4 and under the assumptions of Theorem 2.3.1 for all copula C_θ, the goodness-of-fit process $\mathbb{C}_{n;\theta}^{w_n}$ from (2.4.5) converges weakly in $L^2([0,1]^2)$ to the centered Gaussian process $w_\theta\mathbb{B}^{**}$ defined as*

$$w_\theta\mathbb{B}^{**}(\mathbf{u}) := w_\theta(\mathbf{u})\left[\mathbb{B}^*(\mathbf{u}) - \dot{C}_\theta^T(\mathbf{u})\Theta\right], \ \mathbf{u} \in [0,1]^d.$$

Proof. See Section 2.5.3.□
Assumptions H.1-H.2 are similar to the hypotheses A.2-A.3 in Kojadinovic et al (2010) p.4. Assumption H.3 is necessary to apply the Slutsky's theorem and H.4 is requested by Theorem 2.3.1.

We will show in Section 2.6.2 that the hypothesis H.2 holds for the Clayton, Frank, Gumbel-Hougaard, Gaussian and Student copulas.

2.4.3.1 Copula goodness-of-fit statistic

The asymptotic result obtained in Proposition 2.4.2 motivates the use of goodness-of-fit statistics based on continuous functionals based upon $\mathbb{C}_{n;\theta}^{w_n}$ in virtue of the continuous mapping theorem. An omnibus statistic which turns out to have good power properties in general is the Cramér-von-Mises statistic

$$S_{2,n} := \int\limits_{[0,1]^d} \{\mathbb{C}_{n;\theta}^{w_n}(\mathbf{u})\}^2 \, d\mathbf{u}. \tag{2.4.6}$$

2.4.3.2 P-value's approximation

In practice, the limiting distribution, under the composite null hypothesis, of the goodness-of-fit process $\mathbb{C}_{n;\theta}^{w_n}$ from (2.4.5) depends upon the family of copulas $\{C_\theta : \theta \in \Delta\}$ and on the unknown parameter value θ (see Proposition 2.4.1). As a result, the asymptotic distribution of the test statistic $S_{2,n}^w$ cannot be tabulated easily and in practice approximate p-values can only be obtained via specially adapted simulation methods. We expose below such simulation algorithm based on the multiplier central limit theorem (see Kojadinovic et al (2010)).
The validity of this algorithm is provided by Theorem 2.6.1 and Theorem

2.6.2 in Section 2.6.1.

Goodness of fit procedure: In the present procedure we limit ourselves to the bivariate case ($d = 2$).

1. Given a random sample $(X_1, Y_1), ..., (X_n, Y_n)$ from the c.d.f F, compute $\overline{C}_n$ using (2.2.10). Compute an estimate θ using an estimator θ_n from class $\mathcal{R}_1$ or $\mathcal{R}_2$ defined in Section 2.6.1.

2. Compute the Cramér-von Mises statistic

$$S_{2,n} := \int\limits_{[0,1]^2} n \left[\left(\overline{C}_n(u,v) - C_{\theta_n}(u,v) \right) \cdot w_{\theta_n}(u,v) \right]^2 du\, dv.$$

   By using the numerical approximation based on $m > 0$ (large integer) uniformly spaced points on $(0,1)^2$ denoted $(u_1^*, v_1^*), ..., (u_m^*, v_m^*)$ evaluate,

$$S_{2,n} \approx \frac{1}{m} \sum_{i=1}^{m} n \left[\left(\overline{C}_n(u_i^*, v_i^*) - C_{\theta_n}(u_i^*, v_i^*) \right) \cdot w_{\theta_n}(u_i^*, v_i^*) \right]^2.$$

3. Then, for some large integer N, repeat the following steps for every $k \in \{1, .., N\}$:

   (a) Generate n i.i.d. random variates $Z_1^{(k)}, ..., Z_n^{(k)}$ with expectation 0 and variance 1.

   (b) Form an approximate realization of the test statistic under H_0 by

$$S_{2,n}^{(k)} := \int\limits_{[0,1]^2} n \left[\left(\mathbb{C}_n^{(k)}(u,v) - \Theta_n^{(k)} \dot{C}_{\theta_n}(u,v) \right) \cdot w_{\theta_n}(u,v) \right]^2 du\, dv$$
$$\approx \frac{1}{m} \sum_{i=1}^{m} n \left[\left(\mathbb{C}_n^{(k)}(u_i^*, v_i^*) - \Theta_n^{(k)} \dot{C}_{\theta_n}(u_i^*, v_i^*) \right) \cdot w_{\theta_n}(u_i^*, v_i^*) \right]^2.$$

   where

$$\begin{aligned}
\mathbb{C}_n^{(k)}(u,v) :=& \alpha_n^{(k)}(u,v) - \frac{\partial C_\theta^{(n)}}{\partial u}(u,v)\alpha_n^{(k)}(u,1) - \frac{\partial C_\theta^{(n)}}{\partial v}(u,v)\alpha_n^{(k)}(1,v),\\
\alpha_n^{(k)}(u,v) :=& \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i^{(k)}\left[\mathbb{1}_{\{U_{i,n}\le u, V_{i,n}\le v\}} - \overline{C}_n(u,v)\right],\\
\Theta_n^{(k)} :=& \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i^{(k)}\mathcal{J}_{\theta_n,i,n},
\end{aligned}$$

$\mathcal{J}_{\theta_n,i,n}$ is defined in Section 2.6.1 and different for class $\mathcal{R}_1$ and $\mathcal{R}_2$ (refer to Theorem 2.6.1 and Theorem 2.6.2).

The functions $\partial C_\theta^{(n)}/\partial u$ and $\partial C_\theta^{(n)}/\partial v$ are consistent estimators of the partial derivatives $\partial C_\theta/\partial u$ and $\partial C_\theta/\partial v$ respectively (see definition on page 176).

From the proof of Theorem 2.1 in Rémillard and Scaillet (2009), it follows that the empirical process $\mathbb{C}_n^{(k)}$ can be regarded as approximate independent copies of the weak limit $\mathbb{B}^*$ defined in (2.2.14).
The random vectors $(U_{i,n}, V_{i,n})$ are *pseudo-observations* from C computed from the data

$$U_{i,n} = \frac{1}{n+1}\sum_{j=1}^{n}\mathbb{1}_{\{X_j\le X_i\}},\ V_{i,n} = \frac{1}{n+1}\sum_{j=1}^{n}\mathbb{1}_{\{Y_j\le Y_i\}},\ i=1,...,n. \tag{2.4.7}$$

4. An approximate *p-value* for the test is then given by $N^{-1}\sum_{k=1}^{N}\mathbb{1}_{\{S_{2,n}\ge S_{2,n}^{(k)}\}}$.

Note that some authors (see e.g., Kojadinovic et al (2010)) base their computations on the scaled version of *càdlàg* empirical copula function, i.e., $\frac{n}{n+1}\overline{C}_n$.

2.4.3.3 Choice of weight functions

We consider here an example of the random weight function. For $\mathbf{u}\in[0,1]^d$, we set

$$w_{\theta_n}(\mathbf{u}) = [c_{\theta_n}(\mathbf{u})]^{1/2},$$

where c_{θ_n} represents the density function of C_{θ_n}.

The hypothesis H.3 of Proposition 2.4.2 (see page 64) for the weight function $c_{\theta_n}^{1/2}$ is fulfilled via Proposition 2.4.3 below. Throughout, we assume that

H.5 For all $\theta \in \Delta$, the copula C_θ admits a density c_θ which is continuous on $\Delta \times (0,1)^d$, and such that

$$c_\theta(\mathbf{u}) = \mathrm{O}\left(\frac{1}{\sqrt{\prod_{j=1}^d u_j(1-u_j)}}\right),$$

for $\mathbf{u} = (u_1, ..., u_d) \in \mathbb{R}^d$.

Proposition 2.4.3 *Under H.1, H.2 and H.5 , the function* $c_{\theta_n}^{1/2}$ *fulfills the condition H.3 in Proposition 2.4.1.*

Proof. See Section 2.5.4.□

We will show in Section 2.6.3 that the hypothesis **H.5** holds for the Clayton, Frank, Gumbel-Hougaard, Gaussian and Student copulas.

2.4.3.4 Power study-discussion

The finite sample performance of the goodness-of-fit test based on statistic $S_{2,n}$ can be carried out in a large scale simulation study. The validity of proposed simulation algorithm relies on the multiplier central theorem. We have in hand the necessary methods (see algorithm in Section 2.4.3.2 and Theorem 2.6.1 and Theorem 2.6.2 in Appendix 2.6.1) to implement simulations for statistic $S_{2,n}$ to measure the power of this GOF statistic.

However, writing source codes demands high-level programming skills and is extremely time-consuming. Therefore, this part of our study is omitted and will be left for future research to complete the present results in this thesis.

2.5 Proofs.

In this section we give details on the proofs of our results.

2.5.1 Proof of Theorem 2.3.1

For $i = 1, ..., d$, denote by Π_i the projection on coordinate i, i.e., $\Pi_i(\mathbf{u}) = (1, .., u_i, ..., 1)$. Recall that $\alpha_{nj} = \alpha_n \circ \Pi_j$. Set

$$\mathbb{C}_{n,0} = \alpha_n - \sum_{j=1}^{d} C_j \alpha_{nj}.$$

Set $R_n = \mathbb{C}_n - \mathbb{C}_{n,0}$. Our first lemma describes the limiting behavior of $w\mathbb{C}_{n,0}$.

Lemma 2.5.1 *If w is in $L^p([0,1]^d, \mu)$, then $w\mathbb{C}_{n,0}$ converges weakly, as $n \to \infty$, to $w\mathbb{B}^*$ in $L^p([0,1]^d, \mu)$.*

Proof. Define Y_i to be w times

$$\mathbb{1}_{[\mathbf{U}_i,1]} - C - \sum_{j=1}^{d} C_j \cdot (\mathbb{1}_{[\mathbf{U}_i,1]}(\Pi_j) - C \circ \Pi_j),$$

so that $w\mathbb{C}_{n,0} = n^{-1/2} \sum_{1 \le i \le n} Y_i$. The random variables Y_i are in $L^p([0,1]^d, \mu)$, and obey to the central limit theorem in $L^p([0,1]^d, \mu)$ (see Section 2.6).

□

Lemma 2.5.2 *If w is in $L^p([0,1]^d, \mu)$, then $|wR_n|_p = o_P(1)$ as $n \to \infty$.*

Proof. Let I denote the identity function on $[0, 1]$. Consider, for $j \in \{1, ..., d\}$, the quantile processes associated to the marginals G_{nj} ,

$$\beta_{nj} = n^{1/2}(G_{nj}^{-1} - I).$$

We have the following decomposition of the empirical copula process

$$\mathbb{C}_n = \alpha_n + n^{1/2}[C(F_{n1}^{-1}, ..., F_{nd}^{-1}) - C] + \alpha_n(F_{n1}^{-1}, ..., F_{nd}^{-1}) - \alpha_n.$$

From the above decomposition of $\mathbb{C}_n$ we can decompose R_n into the sum of

$$\begin{aligned} R_{1,n} &= \alpha_n(G_{n1}^{-1}, ..., G_{nd}^{-1}) - \alpha_n, \\ R_{2,n} &= \sum_{j=1}^{d} C_j(\alpha_{nj} + \beta_{nj}), \\ R_{3,n} &= \sqrt{n}[C(G_{n1}^{-1}, ..., G_{nd}^{-1}) - C] - \sum_{j=1}^{d} C_j \beta_{nj}, \end{aligned}$$

By using the triangle inequality, it suffices to show for $i = 1, 2, 3$, $|wR_{i,n}|_p = o_P(1)$ as $n \to \infty$. The proof of Proposition 1 in Tsukahara (2005) implies that $R_{1,n}$, $R_{2,n} \to 0$ uniformly on $[0,1]^d$, almost surely, as $n \to \infty$. Since $|wR_{i,n}|_p \leq |R_{i,n}|_\infty |w|_p$, we conclude that $|wR_{i,n}|_p = o_P(1)$ for $i = 1, 2$.
To show that $|wR_{3,n}|_p \to 0$ we use the following extended continuous mapping theorem

Theorem 2.5.3 (Extended continuous mapping) *Let $\mathbb{D}_n \subset \mathbb{D}$ and $g_n : \mathbb{D}_n \to \mathbb{E}$ satisfy the following statements:if $x_n \to x$ with $x_n \in \mathbb{D}_n$ for every n and $x \in \mathbb{D}_0$, then $g_n(x_n) \to g(x)$, where $\mathbb{D}_0 \subset \mathbb{D}$ and $g : \mathbb{D}_0 \mapsto \mathbb{E}$. Let X_n be maps with values in $\mathbb{D}_n$, let X be Borel measurable and separable, and takes values in $\mathbb{D}_0$. Then*

(i) $X_n \rightsquigarrow X$ *implies that* $g_n(X_n) \rightsquigarrow g(X)$

(ii) $X_n \xrightarrow{\mathbb{P}*} X$ *implies that* $g_n(X_n) \xrightarrow{\mathbb{P}*} g(X)$

(iii) $X_n \xrightarrow{as*} X$ *implies that* $g_n(X_n) \xrightarrow{as*} g(X)$

Proof. See Theorem 1.11.1 in van der Vaart and Wellner (1996). □

Introduce the spaces: $\mathbb{D} := \ell^\infty([0,1])^{\bigotimes d}$, $\mathbb{D}_n := \{h \in \mathbb{D} : u + \sqrt{n}h(u) \in [0,1]^d\}$, $\mathbb{D}_0 := \mathcal{C}([0,1])^{\bigotimes d}$ and $\mathbb{E} := L^p([0,1]^d, \mu)$. Set, for $x_n = (x_{n1}, ..., x_{nd}) \in \mathbb{D}_n$ and $x = (x_1, ..., x_d) \in \mathbb{D}_0$, $g_n(x_n) := \sqrt{n}[C(I + n^{-1/2}x_n) - C] - \sum_{j=1}^d C_j x_{nj}$ and $g(x) := 0$. Let define $\xi_n(s) := I + sn^{-1/2}x_n$ and $f_n(s) := C(\xi_n(s))$. Obviously, $\sqrt{n}[C(I + n^{-1/2}x_n) - C] = f_n(1) - f_n(0)$. Since $\partial \xi_n / \partial s = n^{-1/2}x_n$, we obtain readily that (see Theorem A.7.1 on page 181)

$$\sqrt{n}[C(I + n^{-1/2}x_n) - C] = \sqrt{n}[g_n(1) - g_n(0)] = \int_0^1 g_n^{'}(s)\ ds,$$

where $f_n'(s) = \sum_{j=1}^d C_j(\xi_n(s))n^{-1/2}x_{nj}$. Hence

$$\begin{aligned}
\sqrt{n}\int_0^1 f_n'(s)\ ds &= \sqrt{n}\int_0^1 \sum_{j=1}^d C_j(\xi_n(s))\ ds \cdot n^{-1/2}x_{nj} \\
&= \sum_{j=1}^d \int_0^1 C_j(\xi_n(s))\ ds \cdot x_{nj}, \\
g_n(x_n) - g(x) &= \sum_{j=1}^d \left(\int_0^1 C_j(\xi_n(s))\ ds - C_j\right) x_{nj} \\
&= \sum_{j=1}^d \int_0^1 (C_j(\xi_n(s)) - C_j)\ \ ds\ x_{nj}.
\end{aligned}$$

By the triangle inequality, to show that $|g_n(x_n) - g(x)|_p$ tends to 0 as $n \to \infty$, it is sufficient to show that, for all $j \in \{1, ..., d\}$, $|I_j|_p$ tends to 0 as $n \to \infty$, where $I_j := [\int_0^1 (C_j(\xi_n(s)) - C_j)\ \ ds]x_{nj}$. Furthermore, $|I_j|_p^p := \int\limits_{[0,1]^d} I_j(\mathbf{u})^p d\mu(\mathbf{u}) \leq |x_{nj}^2|_\infty \cdot \int\limits_{[0,1]^d} \left[\int_0^1 (C_j(\xi_n(s)) - C_j)\ \ ds\right]^p d\mu(\mathbf{u})$. Since $x_n \to x$ in $\mathbb{D}$, this implies that $|x_{nj}^2|_\infty$ tends to $|x_j^2|_\infty$, as $n \to \infty$. It remains to show that $\int\limits_{[0,1]^d} \delta_n(\mathbf{u})^p d\mu(\mathbf{u})$ tends to 0 as $n \to \infty$, where $\delta_n(\mathbf{u}) := \int_0^1 \eta_n(s)\ ds$ and $\eta_n(s) := \int_0^1 (C_j(\xi_n(s)) - C_j)\ \ ds$. We start by showing that $\delta_n(\mathbf{u}) \to 0$, μ-almost everywhere, as $n \to \infty$.

Note that from the continuity of C_j on the set D_C and from the fact that, for all s, $\xi_n(s) \to 0$, $\eta(s) \to 0$, λ_1-almost everywhere, as $n \to \infty$. Furthermore, from the bounds $0 \leq C_j \leq 1$, the sequence η_n is uniformly bounded λ_1-almost everywhere, i.e., $|\eta_n(s)| \leq M$, for some positive constant M. By the dominated convergence theorem, as $n \to \infty$

$$\delta_n(\mathbf{u}) = \int_0^1 \eta_n(s)\ ds \to 0.$$

This implies that $\delta_n(\mathbf{u})^2 \to 0$ μ-almost everywhere, $n \to \infty$. Again, from the fact $0 \leq C_j \leq 1$, we conclude that δ_n^2 is uniformly bounded on $[0,1]^d$,

μ-almost everywhere. By the dominated convergence theorem, as $n \to \infty$ $|g_n(x_n) - g(x)|_p$ tends to 0, which establishes the assertion that $g_n(x_n) \to g(x)$ in the space $\mathbb{E}$. To finish our proof , we apply the extended continuous mapping theorem to the sequence $X_n := \frac{1}{\sqrt{n}}\beta_n$. Since $\beta_n = O_P(1)$ in $\ell^\infty([0,1]^d)$, X_n converges weakly to 0, as $n \to \infty$, in ℓ^∞ and in L^p as well. Since the weak convergence to a constant entails convergence in probability, the proof of Lemma 2.5.2 is completed. □

Theorem 2.3.1 follows readily from Lemmas 2.5.1 and 2.5.2. □

2.5.2 Proof of Proposition 2.4.1

Set

$$B_{1,n} = n^{1/2}(\overline{C}_n - C_\theta) \quad \text{and} \quad B_{2,n} = n^{1/2}(C_{\theta_n} - C_\theta),$$

so that $\mathbb{C}_{n,\theta} = wB_{1n} - wB_{2n}$. Since $|C_n - \overline{C}_n|_\infty \leq d/n$ (see Proposition 2.2.2, (e)), Theorem 2.3.1 implies that $wB_{1,n}$ converges weakly to $w\mathbb{B}^*$ as $n \to \infty$ By the inequality

$$|wB_{2n} - w\Theta_n\dot{C}_\theta|_p \leq |w|_p|B_{2n} - \Theta_n\dot{C}_\theta|_\infty,$$

our proof follows from the observation that $|B_{2n} - \Theta_n\dot{C}_\theta|_\infty \to 0$ in probability. This is a consequence of the mean value theorem and of the assumptions H.1 and H.2 (see e.g. (Quessy, 2005, p.71-73) and (Berg and Quessy, 2009, p.2)).□

2.5.3 Proof of Proposition 2.4.2

We have, $\mathbb{C}^{w_n}_{n;\theta} := \sqrt{n}\{\overline{C}_n - C_{\theta_n}\} \cdot w_{\theta_n} = \sqrt{n}\{\overline{C}_n - C_{\theta_n}\} \cdot w_\theta + \sqrt{n}\{\overline{C}_n - C_{\theta_n}\} \cdot (w_{\theta_n} - w_\theta)$.

From Hypothesis H.3 the term $w_{\theta_n} - w_\theta$ converges in L^p to 0 in probability. Slutsky's theorem implies that $\sqrt{n}\{\overline{C}_n - C_{\theta_n}\} \cdot (w_{\theta_n} - w_\theta)$ converges weakly in L^p to 0. Since the weak convergence to a constant is equivalent to the convergence in probability, the last term converges in L^p to 0 in probability. From Proposition 2.4.1 the process $\sqrt{n}\{\overline{C}_n - C_{\theta_n}\} \cdot w_\theta$ converges weakly in L^p to the limit $w_\theta\mathbb{B}^{**}$. The second application of Slutsky's theorem completes our proof.□

2.5.4 Proof of Proposition 2.4.3

We assume, without loss of generality, that the density function c_θ is unbounded in the neighborhood of $\mathbf{0} = (0, ..., 0) \in [0,1]^d$. A similar argument can be used when the density function c_θ is unbounded in the neighborhood of other corner points of $[0,1]^d$.

To show H.3, i.e., that $c_{\theta_n}^{1/2} \overset{\mathbb{P}}{\to} c_\theta^{1/2}$ in $L^2([0,1]^d)$, as $n \to \infty$, it is enough to prove that $c_{\theta_n} \overset{\mathbb{P}}{\to} c_\theta$ in $L^1([0,1]^d)$, as $n \to \infty$. This follows from the continuous mapping theorem, when applied to convergence in probability and to the mapping $g : L^1 \ni f \mapsto f^{1/2} \in L^2$. In order to establish the convergence $c_{\theta_n} \overset{\mathbb{P}}{\to} c_\theta$ in $L^1([0,1]^d)$ it is enough to show that $|c_{\theta_n} - c_\theta|_1 \overset{\mathbb{P}}{\to} 0$ in $\mathbb{R}$. We have, for $\mathbf{t}^0 = (t_1^0, ..., t_d^0)$ and $A = [0, t_1^0] \times \ldots \times [0, t_d^0]$,

$$
\begin{aligned}
|c_{\theta_n} - c_\theta|_1 &= \int\limits_{[0,1]^d} |c_{\theta_n}(\mathbf{u}) - c_\theta(\mathbf{u})|\, d\mathbf{u} \\
&= \int_{[0,1]^d \setminus A} |c_{\theta_n}(\mathbf{u}) - c_\theta(\mathbf{u})|\, d\mathbf{u} + \int_A |c_{\theta_n}(\mathbf{u}) - c_\theta(\mathbf{u})|\, d\mathbf{u}.
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
\int_A |c_{\theta_n}(\mathbf{u}) - c_\theta(\mathbf{u})|\, d\mathbf{u} &\leq \int_A |c_{\theta_n}(\mathbf{u})|\, d\mathbf{u} + \int_A |c_\theta(\mathbf{u})|\, d\mathbf{u} \\
&= \int_A c_{\theta_n}(\mathbf{u})\, d\mathbf{u} + \int_A c_\theta(\mathbf{u})\, d\mathbf{u} \\
&\leq C_{\theta_n}(\mathbf{t^0}) + C_\theta(\mathbf{t^0}) \\
&\leq |C_{\theta_n}(\mathbf{t^0}) - C_\theta(\mathbf{t^0})| + 2C_\theta(\mathbf{t^0})
\end{aligned}
$$

Thus, we get

$$
|c_{\theta_n} - c_\theta|_1 \leq \int_{[0,1]^d \setminus A} |c_{\theta_n}(\mathbf{u}) - c_\theta(\mathbf{u})|\, d\mathbf{u} + |C_{\theta_n}(\mathbf{t^0}) - C_\theta(\mathbf{t^0})| + 2C_\theta(\mathbf{t^0}).
$$

Note that $C_\theta(\mathbf{t^0}) \to 0$ as $\mathbf{t^0} \to \mathbf{0}$. We show next that $|C_{\theta_n}(\mathbf{t^0}) - C_\theta(\mathbf{t^0})|$ converges to 0 in probability. By the mean value theorem there exists a θ_n^* such that $\|\theta_n^* - \theta\| \leq \|\theta_n - \theta\|$ and

$$
|C_{\theta_n}(\mathbf{t^0}) - C_\theta(\mathbf{t^0})| \leq |\dot{C}_{\theta_n^*}(\mathbf{t^0}) - \dot{C}_\theta(\mathbf{t^0})| \|\theta_n - \theta\| + |\dot{C}_\theta(\mathbf{t^0})| \|\theta_n - \theta\|.
$$

Since H.1 entails that $\Theta_n := \sqrt{n}(\theta_n - \theta)$ converges in distribution to Θ, the second term converges to 0 in probability via Slutsky's theorem. Concerning the first term, note that for all $\delta > 0$ there exist an $M = M_\delta$ such that $P(\|\Theta_n\| \geq M) < \delta$ (this is due to the tightness of Θ_n). Furthermore, we have

$$\begin{aligned}
&P(|\dot{C}_{\theta_n^*}(\mathbf{t^0}) - \dot{C}_\theta(\mathbf{t^0})|\|\theta_n - \theta\| \geq \varepsilon) \\
&= P(|\dot{C}_{\theta_n^*}(\mathbf{t^0}) - \dot{C}_\theta(\mathbf{t^0})|\|\theta_n - \theta\| \geq \varepsilon, \|\Theta_n\| \leq M) \\
&+ P(|\dot{C}_{\theta_n^*}(\mathbf{t^0}) - \dot{C}_\theta(\mathbf{t^0})|\|\theta_n - \theta\| \geq \varepsilon, \|\Theta_n\| \geq M) \\
&\leq P(\sup_{\mathbf{t}\in[0,1]^d} |\dot{C}_{\theta_n^*}(\mathbf{t}) - \dot{C}_\theta(\mathbf{t})|\|\theta_n - \theta\| \geq \varepsilon, \|\Theta_n\| \leq M) + P(\|\Theta_n\| \geq M) \\
&\leq P\left(\sup_{\|\theta^*-\theta\|\leq M/\sqrt{n}} \sup_{\mathbf{t}\in[0,1]^d} |\dot{C}_{\theta_n^*}(\mathbf{t}) - \dot{C}_\theta(\mathbf{t})| \geq \frac{\varepsilon}{M}\right) + \delta,
\end{aligned}$$

where the last inequality is due to inclusion of corresponding events and from the appropriate choice of the constant M depending on δ.
As $\delta > 0$ can be chosen arbitrarily small, and in view of the hypothesis H.2 the result follows.
Finally, we show that $\int_{[0,1]^d\setminus A} |c_{\theta_n}(\mathbf{u}) - c_\theta(\mathbf{u})|\, d\mathbf{u}$ converges to 0 in probability, as $n \to \infty$.
Set, for $\mathbf{t} \in [0,1]^d$

$$g_n(\mathbf{t}) := \sup_{\|\theta^*-\theta\|\leq M/\sqrt{n}} |c_{\theta^*}(\mathbf{t}) - c_\theta(\mathbf{t})|.$$

By similar arguments as above we obtain for the first term, the upper bound

$$P\left(\int_{[0,1]^d\setminus A} |c_{\theta_n}(\mathbf{u}) - c_\theta(\mathbf{u})|\, d\mathbf{u} \geq \varepsilon\right) \leq P\left(\int_{[0,1]^d\setminus A} g_n(\mathbf{t})\, d\mathbf{t} \geq \varepsilon\right). \quad (2.5.1)$$

We infer from the continuity of the function $\theta \mapsto c_\theta$, for all $\mathbf{t} \in [0,1]^d$, that $g_n(\mathbf{t}) \to 0$, as $n \to \infty$. The assumption H.5 implies that $|g_n(\mathbf{t})| = O(h(\mathbf{t}))$, as $\mathbf{t} \to 0$, where $h(\mathbf{t}) = \frac{1}{\sqrt{\prod_{j=1}^d u_j \times \prod_{j=1}^d (1-u_j)}}$. This is implied by the continuity of $\theta \mapsto c_\theta$ on compact sets. We have, therefore

$$g_n(t) = |c_{\theta_n^*}(\mathbf{t}) - c_\theta(\mathbf{t})|,$$

for some θ_n^* such that $\|\theta_n^* - \theta\| \le M/\sqrt{n}$. Since the function h is integrable this implies that the sequence g_n is uniformly integrable. By dominating convergence theorem and from (2.5.1) we obtain that $\int_{[0,1]^d\setminus A} |c_{\theta_n}(\mathbf{u}) - c_\theta(\mathbf{u})|\, d\mathbf{u}$ converges to 0 in probability, as $n \to \infty$.□

2.5.5 An Alternative Proof to Theorem 2.3.1

The main goal of the second proof of Theorem 2.3.1 is to introduce the techniques of functional delta method. We show that this approach can be useful in proving the weak-convergence-type results of empirical copula processes.
We give proof for $d = 2$. It will become obvious later on that our arguments may be adapted to $d \ge 3$ by routine modifications.

2.5.5.1 Preliminaries

Let $\mathcal{X}$ be a set and $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ be a normed space. Denote by $\ell^\infty(\mathcal{X}, \mathcal{Y})$ the space of bounded functions from $\mathcal{X}$ to $\mathcal{Y}$ defined by

$$\ell^\infty(\mathcal{X}, \mathcal{Y}) := \{f : \mathcal{X} \mapsto \mathcal{Y} : \exists M > 0,\ \forall x \in \mathcal{X} : \|f(x)\|_{\mathcal{Y}} \le M\},$$

where $|f|_\infty := \sup_{x\in\mathcal{X}} \|f(x)\|_{\mathcal{Y}}$. Set

$$S_1 := \ell^\infty([0,1]^2, [0,1]^2),\ S_2 := \ell^\infty([0,1]^2, [0,1]),$$
$$S_3 := C([0,1]^2, [0,1]) := \{\beta : [0,1]^2 \mapsto [0,1] : \beta \ \text{ continuous on } [0,1]^2\}.$$

We endow S_1, S_2 and S_3 respectively, with the norms

$$\begin{aligned}\|\alpha\|_{S_1} &:= \max\left\{|\alpha^{(1)}|_\infty; |\alpha^{(2)}|_\infty\right\}\\ &= \max\left\{\sup_{(u,v)\in[0,1]^2} |\alpha^{(1)}(u,v)|;\ \sup_{(u,v)\in[0,1]^2} |\alpha^{(2)}(u,v)|\right\},\\ \|\beta\|_{S_2} &:= |\beta|_\infty,\ \text{ and } \|\beta\|_{S_3} := \|\beta\|_{S_2},\end{aligned}$$

where $\alpha := (\alpha^{(1)}, \alpha^{(2)})$ and $\alpha^{(1)}, \alpha^{(2)} \in S_2$.

2.5.5.2 A Proof of Theorem 2.3.1 based on the Delta-method

Consider the functional $\Phi : S_2 \mapsto L^p\left([0,1]^2, \mu\right)$ defined by

$$\Phi(H) = w \cdot F \circ \left(F_1^{-1}, F_2^{-1}\right) = w \cdot C, \tag{2.5.2}$$

where C is a copula function corresponding to F via (2.2.2). We know that $\alpha_n \rightsquigarrow \mathbb{B}$, where $\alpha_n \in S_1$ and $\mathbb{B}$ are defined in (2.2.12) and (2.2.13) respectively. The next Lemma 2.5.4 shows that the map Φ is Hadamard-differentiable at C.

Lemma 2.5.4 *The map Φ defined via (2.5.2) is Hadamard-differentiable at C tangentially to the set S_3. The corresponding derivative is given by*

$$\Phi_C^{'}(\beta) = w(u,v) \cdot \left\{\beta(u,v) - C_u(u,v) \cdot \beta(u,1) - C_v(u,v) \cdot \beta(1,v)\right\}, \tag{2.5.3}$$

where $\beta \in S_3,\ (u,v) \in \mathcal{D}_C$.

By an application of the Delta-method (see Theorem 3.9.4 in van der Vaart and Wellner (1996)) to the process α_n we see that the process $w\mathbb{C}_n = \Phi(\alpha_n)$ converges weakly in $L^p([0,1]^2, \mu)$ to $\Phi_C^{'}(\mathbb{B})$. We infer from (2.5.3) in Lemma 2.5.4 that $\Phi_C^{'}(\mathbb{B}) = w\mathbb{B}^*$. This completes the proof of Theorem 2.3.1.□

Denote by I the identity function on $[0,1]$. Set $\mathbf{I} := (I, I) \in S_1$ and $\theta_0 = (\mathbf{I}, C) \in S_1 \times S_2$.

2.5.5.3 Proof of Lemma 2.5.4

The functional Φ can be decomposed into

$$\Phi = \phi \circ \xi \circ \zeta,$$

where $\zeta : S_2 \mapsto S_1 \times S_2$, $\xi : S_1 \times S_2 \mapsto S_1 \times S_2$ and $\phi : S_1 \times S_2 \mapsto L^p([0,1]^2, \mu)$ are defined by

$$\zeta(F) := ((F_1, F_2), F), \xi((F_1, F_2), F) := ((F_1^{-1}, F_2^{-1}), F),$$

and

$$\phi((F_1^{-1}, F_2^{-1}), F) := w \cdot F \circ (F_1^{-1}, F_2^{-1}),$$

where $(F_1, F_2), (F_1^{-1}, F_2^{-1}) \in S_1, F \in S_2$ and $w \in L^p([0,1]^2, \mu)$.
The map ζ is linear and continuous, hence Hadamard-differentiable tangentially to the space S_3, and such that

$$\zeta_C^{'}(\beta) = ((\beta(\cdot, 1), \beta(1, \cdot)), \beta) \in S_1 \times S_2.$$

The map ξ is Hadamard-differentiable at θ_0 by Lemma 3.9.23, p.386, of van der Vaart and Wellner (1996) (see also the proof of Theorem 2., p.4, in Fermanian et al (2004)). The coresponding derivative $\xi_C^{'}$ is given by

$$\xi_C^{'}(\alpha, \beta) = (-\alpha, \beta).$$

The map ϕ is Hadamard-differentiable at θ_0, tangentially to the set $S_1 \times S_3$. This is due to the following lemma

Lemma 2.5.5 *The functional ϕ is Hadamard-differentiable at θ_0 tangentially to the set $S_1 \times S_3$. The corresponding derivative is given by*

$$\phi_{\theta_0}^{'}(\alpha, \beta) = w \cdot \beta \circ \mathbf{I} + w \cdot C_{\mathbf{I}}^{'} \circ \alpha, \ (\alpha, \beta) \in S_1 \times S_3, \tag{2.5.4}$$

where $C_{(s,t)}^{'}$, for (s, t) denotes the Fréchet derivative of the copula function C (refer to Example 3.9.2 in van der Vaart and Wellner (1996) for details).

Remark 2.5.1 *The Fréchet derivative $C_{(s,t)}^{'} : [0,1]^2 \to [0,1]$ of C, for every $(s,t) \in [0,1]^2$ is given by (refer to Example 3.9.2 in van der Vaart and Wellner (1996))*

$$C_{(s,t)}^{'}(u, v) = C_1(s, t) \cdot u + C_2(s, t) \cdot v.$$

By applying the chain rule (see e.g., Lemma 3.9.3 in van der Vaart and Wellner (1996)), we get

$$\begin{aligned} \Phi_C^{'}(\beta) &= \phi_{\xi\circ\zeta(C)}^{'} \circ \xi_{\zeta(C)}^{'} \circ \zeta_C^{'}(\beta) = \phi_{\theta_0}^{'}\left(\xi_{\theta_0}^{'}\left(\zeta_C^{'}(\beta)\right)\right) = \phi_{\theta_0}^{'}\left(\xi_{\theta_0}^{'}(\alpha, \beta)\right), \\ &= \phi_{\theta_0}^{'}((-\alpha, \beta)) = w \cdot \left\{\beta \circ \mathbf{I} + C_{\mathbf{I}}^{'} \circ (-\alpha)\right\} = w \cdot \left\{\beta \circ \mathbf{I} - C_{\mathbf{I}}^{'} \circ \alpha\right\}, \end{aligned}$$

where $\alpha = (\beta(\cdot, 1), \beta(1, \cdot))$. In view of Remark 2.5.1 we have

$$\Phi_C^{'}(\beta)(u, v) = w(u, v) \cdot \{\beta(u, v) - C_1(u, v)\beta(u, 1) - C_2(u, v)\beta(1, v)\}, \ (u, v) \in \mathcal{D}_C.$$

□

2.5.5.4 Proof of Lemma 2.5.5

First we show that the relation

$$\frac{\phi(\theta_0 + t_n h_n) - \phi(\theta_0)}{t_n} \to \phi'_{\theta_0}(h) \text{ in } L^p([0,1]^d, \mu) \text{ as } n \to \infty, \tag{2.5.5}$$

holds (2.5.5) for all sequences $t_n \to 0$, and $h_n \to h$, with $h \in S_3$, and $h_n, \theta_0 + t_n h_n \in S_1 \times S_2$ for all n. Let $h_n = (\alpha_n, \beta_n)$ and $h = (\alpha, \beta)$. The convergence $h_n \to h$ in $S_1 \times S_3$ entails that $\alpha_n \to \alpha$ in S_1 and $\beta_n \to \beta$ in S_3. We can therefore rewrite (2.5.5) into

$$\left| \frac{\phi(\theta_0 + t_n(\alpha_n, \beta_n)) - \phi(\theta_0)}{t_n} - \phi'_{\theta_0}((\alpha, \beta)) \right|_p \to 0 \text{ in } \mathbb{R} \text{ as } n \to \infty. \tag{2.5.6}$$

From the definitions of ϕ, θ_0, and recalling the expression of ϕ'_{θ_0} given in (2.5.4), we get

$$\begin{aligned}
&\frac{\phi(\theta_0 + t_n(\alpha_n, \beta_n)) - \phi(\theta_0)}{t_n} - \phi'_{\theta_0}\big((\alpha, \beta)\big) = \\
&= \frac{\phi\Big(\big(\mathbf{I} + t_n\alpha_n, C + t_n\beta_n\big)\Big) - \phi\Big(\big(\mathbf{I}, C\big)\Big)}{t_n} - \phi'_{\theta_0}((\alpha, \beta)), \\
&= \frac{(w \cdot C + w \cdot t_n\beta_n) \circ \big(\mathbf{I} + t_n\alpha_n\big) - w \cdot C \circ \mathbf{I}}{t_n} - w \cdot \beta \circ \mathbf{I} - w \cdot C'_{\mathbf{I}} \circ \alpha, \\
&= \frac{w \cdot C \circ \big(\mathbf{I} + t_n\alpha_n\big) - w \cdot C \circ \mathbf{I}}{t_n} + w \cdot \beta_n \circ \big(\mathbf{I} + t_n\alpha_n\big) - w \cdot \beta \circ \mathbf{I} - w \cdot C'_{\mathbf{I}} \circ \alpha.
\end{aligned}$$

Furthermore, by straightforward computations we write

$$\frac{\phi(\theta_0 + t_n(\alpha_n, \beta_n)) - \phi(\theta_0)}{t_n} - \phi'_{\theta_0}\big((\alpha, \beta)\big) = A_{1n} + A_{2n} + A_{3n} + A_{4n},$$

where

$$\begin{aligned}
A_{1n} &= w \cdot (\beta_n - \beta) \circ \big(\mathbf{I} + t_n\alpha_n\big), \\
A_{2n} &= w \cdot (\beta \circ \big(\mathbf{I} + t_n\alpha_n\big) - \beta \circ \mathbf{I}), \\
A_{3n} &= \frac{w \cdot C \circ \big(\mathbf{I} + t_n\alpha_n\big) - w \cdot C \circ \mathbf{I}}{t_n} - w \cdot C'_{\mathbf{I}} \circ \alpha_n, \\
A_{4n} &= w \cdot C'_{\mathbf{I}} \circ (\alpha_n - \alpha).
\end{aligned}$$

In view of the triangle inequality, to establish (2.5.6) it is enough to prove that $|A_{in}|_p = o_P(1)$, as $n \to \infty$, for $i = 1, .., 4$. We now show that $|A_{1n}|_p = o_P(1)$, as $n \to \infty$. For this, we observe that

$$\begin{aligned} |A_{1n}|_p^p &:= \int_{[0,1]^2} \left\{ w(u,v)(\beta_n - \beta)\big((u,v) + t_n\alpha_n(u,v)\big)\right\}^p \, d\mu(u,v) \\ &\leq |w|_p^p \cdot \sup_{(s,t)\in[0,1]^2} \{\beta_n(s,t) - \beta(s,t)\}^p . \end{aligned}$$

Hence, we see that

$$|A_{1n}|_p \leq |w|_p \cdot \|\beta_n - \beta\|_{S_2} \to 0,$$

where the convergence is implied by the fact that $\beta_n \to \beta$ in S_2. We next show that $|A_{2n}|_p = o_P(1)$, as $n \to \infty$. We have the inequality

$$\begin{aligned} |A_{2n}|_p^p &:= \int_{[0,1]^2} \left\{ w(u,v)\beta\big((u,v) + t_n\alpha_n(u,v)\big) - \beta(u,v)\right\}^p \, d\mu(u,v) \\ &\leq |w|_p^p \sup_{(u,v)\in[0,1]^2} \left\{ \beta\big((u,v) + t_n\alpha_n(u,v)\big) - \beta(u,v)\right\}^p \to 0. \end{aligned}$$

This last convergence follows from the fact that $\beta \in S_3$, with β uniformly continuous on $[0,1]^2$, and from the fact that $(u,v) + t_n\alpha_n(u,v) \to 0$ as $n \to \infty$.

The convergence to 0 of the term $|A_{3n}|_p$ is established by the arguments used to show the convergence of R_{3n} (see the first proof of Theorem 2.3.1). We conclude by showing that $|A_{4n}|_p = o_P(1)$, as $n \to \infty$. In view of the inequality $|f+g|_p \leq |f|_p + |g|_p$, and making use of Remark 2.5.1 we obtain that

$$
\begin{aligned}
|A_{4n}|_p^p &= \int\limits_{[0,1]^2} \left\{ w(u,v) \cdot C'_{(u,v)}(\alpha_n(u,v) - \alpha(u,v)) \right\}^p d\mu(u,v), \\
&\le \left| (\alpha_n^{(1)} - \alpha^{(1)})^p \right|_\infty \int\limits_{[0,1]^2} w^p(u,v) \cdot C_1^p(u,v)\, d\mu(u,v) \\
&+ \left| (\alpha_n^{(2)} - \alpha^{(2)})^p \right|_\infty \int\limits_{[0,1]^2} w^p(u,v) \cdot C_2^p(u,v)\, d\mu(u,v), \\
&\le |w|_p^p \left\{ \left| (\alpha_n^{(1)} - \alpha^{(1)})^p \right|_\infty + \left| (\alpha_n^{(2)} - \alpha^{(2)})^p \right|_\infty \right\},
\end{aligned}
$$

where this last inequality is due to Proposition 2.6.3. The convergence $\alpha_n \to \alpha$ in S_1 readily implies that $|wR_{4,n}|_p = o_{\mathbb{P}}(1)$, as $n \to \infty$.
To complete the proof of Lemma 2.5.5, we show the continuity of the derivative ϕ'_{θ_0}. We get, namely

$$
\begin{aligned}
\left| \phi'_{\theta_0}(\alpha,\beta) - \phi'_{\theta_0}(\alpha_0,\beta_0) \right|_p^p &\le |w|_p^p \left\{ \|\beta - \beta_0\|_{S_3}^2 + \left\| C'_{\mathbf{I}} \circ (\alpha - \alpha_0) \right\|_{S_2}^2 \right\}, \\
&\le |w|_p^p \left\{ \|\beta - \beta_0\|_{S_3}^2 + \left\| C'_{\mathbf{I}} \right\|_{S_2}^2 \cdot \|\alpha - \alpha_0\|_{S_1}^2 \right\}, \\
&\le |w|_p^p \left\{ \|\beta - \beta_0\|_{S_3}^2 + \|\alpha - \alpha_0\|_{S_1}^2 \right\}.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\left| \phi'_{\theta_0}(\alpha,\beta) - \phi'_{\theta_0}(\alpha_0,\beta_0) \right|_p^p &\le |w|_p^2 \left\{ \|\beta - \beta_0\|_{S_3} + \|\alpha - \alpha_0\|_{S_1} \right\}, \\
&\le 2\, |w|_p^p \cdot \max\left\{ \|\beta - \beta_0\|_{S_3} ; \|\alpha - \alpha_0\|_{S_1} \right\}, \\
&\le \text{const} \cdot \|(\alpha,\beta) - (\alpha_0,\beta_0)\|_{S_1 \times S_3}.
\end{aligned}
$$

□

2.6 Auxiliary Results

2.6.1 Multiplier Central Limit Theorem

Before stating the key result, we define a first class of rank-based estimators of the copula parameter θ. Without loss of generality we assume that $\theta \in \Delta \subset \mathbb{R}$. Set, hereafter, $\Theta_n = \sqrt{n}(\theta_n - \theta)$.

Definition 2.6.1 *A rank-based estimator θ_n of θ is said to belong to the class $\mathcal{R}_1$ if*

$$\Theta_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} J_\theta(U_{i,n}, V_{i,n}) + o_P(1),$$

where $J_\theta : [0,1]^2 \to \mathbb{R}$ is a score function that satisfies the following regularity conditions:

(a) for all $\theta \in \Delta$, J_θ is bounded on $[0,1]^2$ and centered, i.e., $\int_{[0,1]^2} J_\theta(u,v) dC_\theta(u,v) = 0$,

(b) for all $\theta \in \Delta$, the partial derivatives $J_\theta^{[1]} = \partial J_\theta / \partial u$, $J_\theta^{[2]} = \partial J_\theta / \partial v$, $J_\theta^{[1,1]} = \partial^2 J_\theta / \partial u^2$, $J_\theta^{[2,2]} = \partial^2 J_\theta / \partial v^2$, $J_\theta^{[1,2]} = \partial^2 J_\theta / \partial u \partial v$, $J_\theta^{[2,1]} = \partial^2 J_\theta / \partial v \partial u$ exist and are bounded on $[0,1]^2$,

(c) for all $\theta \in \Delta$, the partial derivatives $\dot{J}_\theta = \partial \dot{J}_\theta / \partial \theta$, $\dot{J}_\theta^{[1]} = \partial \dot{J}_\theta^{[1]} / \partial \theta$ and $\dot{J}_\theta^{[2]} = \partial \dot{J}_\theta^{[2]} / \partial \theta^2$ exist and are bounded on $[0,1]^2$,

(d) for all $\theta \in \Delta$, there exist $c_\theta > 0$ and $M_\theta > 0$ such that, if $|\theta' - \theta| < c_\theta$, then $|\dot{J}_{\theta'}| \le M_\theta$, $|\dot{J}_{\theta'}^{[1]}| \le M_\theta$ and $|\dot{J}_{\theta'}^{[2]}| \le M_\theta$.

It is well known (refer to Kojadinovic et al (2010) or Berg and Quessy (2009)) that the estimator $\theta_{\rho,n}$ based on the one of the most popular measure of association - Spearman's rho, belongs to class $\mathcal{R}_1$. Its score function is given by

$$J_{\theta,\rho}(u,v) = \frac{1}{\rho'(\theta)}(12uv - 3 - \rho(\theta)), \quad (u,v) \in [0,1]^2.$$

The following result is instrumental for verifying the asymptotic validity of the fast goodness-of-fit procedure proposed in Section 2.4.3.2 for the estimators of θ belonging to the class $\mathcal{R}_1$.

Theorem 2.6.1 *Let θ_n be an estimator of θ belonging to class $\mathcal{R}_1$ and, for any $k \in \{1, ..., N\}$, let*

$$\Theta_n^{(k)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i^{(k)} \mathcal{J}_{\theta_n,i,n},$$

where

$$\mathcal{J}_{\theta_n,i,n} := J_{\theta_n}(U_{i,n},V_{i,n}) + \frac{1}{n}\sum_{j=1}^{n} J_{\theta_n}^{[1]}(U_{i,n},V_{i,n})[\mathbb{1}_{\{U_{i,n}\le U_{j,n}\}} - U_{j,n}] + \frac{1}{n}\sum_{j=1}^{n} J_{\theta_n}^{[2]}(U_{i,n},V_{i,n})[\mathbb{1}_{\{V_{i,n}\le V_{j,n}\}} - V_{j,n}]. \tag{2.6.1}$$

Then, under Assumptions H.1-H.4,

$$w_n\mathbb{\Gamma}_n := \Big([\sqrt{n}(\overline{C}_n - C_{\theta_n})]w_{\theta_n}, [\mathbb{C}_n^{(1)} - \Theta_n^{(1)}\dot{C}_{\theta_n}]w_{\theta_n}, ..., [\mathbb{C}_n^{(N)} - \Theta_n^{(N)}\dot{C}_{\theta_n}]w_{\theta_n}\Big)$$

converges weakly to

$$w\mathbb{\Gamma} := \Big([\mathbb{B}^* - \Theta\dot{C}_\theta]w_\theta, [\mathbb{B}^{*(1)} - \Theta^{(1)}\dot{C}_\theta]w_\theta, ..., [\mathbb{B}^{*(N)} - \Theta^{(N)}\dot{C}_\theta]w_\theta\Big)$$

in $L^2\left([0,1]^d\right)^{\otimes(N+1)}$*, where* Θ *is the weak limit of* $\Theta_n = \sqrt{n}(\theta_n - \theta)$ *and* $(\mathbb{B}^{*(1)},\Theta^{(1)}),...,(\mathbb{B}^{*(N)},\Theta^{(N)})$ *are independent copies of* $(\mathbb{B}^*,\Theta)$*.*

Proof. Recall that the map Π_i denotes projection on the coordinate i, $i = 1, 2$.
From the proof of Theorem 2 in Kojadinovic et al (2010), the process $\mathbb{\Psi}_n := (\Psi_n, \Psi_n^{(1)}, ..., \Psi_n^{(N)})$ defined, for $j \in \{1,..,N\}$, by

$$\begin{aligned}
\Psi_n &= \alpha_n - \frac{\partial C_\theta}{\partial u}\alpha_n \circ \Pi_1 - \frac{\partial C_\theta}{\partial v}\alpha_n \circ \Pi_2 - \Theta_n \dot{C}_\theta,\\
\Psi_n^{(j)} &= \alpha_n^{(j)} - \frac{\partial C_\theta}{\partial u}\alpha_n^{(j)} \circ \Pi_1 - \frac{\partial C_\theta}{\partial v}\alpha_n^{(j)} \circ \Pi_2 - \Theta_n^{(j)} \dot{C}_\theta,
\end{aligned}$$

converges weakly to $\mathbb{\Gamma}$ in $\ell^\infty([0,1]^2)^{\otimes(N+1)}$. Then, from the continuous mapping theorem, $w\mathbb{\Psi}_n := (\Psi_n w_\theta, \Psi_n^{(1)} w_\theta, ..., \Psi_n^{(N)} w_\theta)$ converges weakly to $w\mathbb{\Gamma}$ in $L^2([0,1]^2)^{\otimes(N+1)}$. We have, for $j \in \{1,..,N\}$

$$\begin{aligned}
(\sqrt{n}(\overline{C}_n - C_{\theta_n}))w_{\theta_n} - w_\theta\Psi_n &= R_n + Q_n + T_n,\\
(\mathbb{C}_n^{(j)} - \Theta_n^{(j)}\dot{C}_{\theta_n})w_{\theta_n} - w_\theta\Psi_n^{(j)} &= Q_n^{(j)} + T_n^{(j)},\\
Q_n &:= (w_{\theta_n} - w_\theta)\sqrt{n}(\overline{C}_n - C_{\theta_n})),\\
T_n &:= w_\theta[\sqrt{n}(C_{\theta_n} - C) - \Theta_n\dot{C}_\theta],\\
Q_n^{(j)} &:= (w_{\theta_n} - w_\theta)(\mathbb{C}_n^{(j)} - \Theta_n^{(j)}\dot{C}_{\theta_n}),\\
T_n^{(j)} &:= w_\theta[\alpha_n^{(j)} \circ \Pi_1(\partial C_\theta/\partial u - \partial C_\theta^{(n)}/\partial u)]\\
&+ w_\theta[\alpha_n^{(j)} \circ \Pi_2(\partial C_\theta/\partial v - \partial C_\theta^{(n)}/\partial v)]\\
&+ w_\theta[\Theta_n^{(j)}(\dot{C}_\theta - \dot{C}_{\theta_n})],
\end{aligned}$$

where R_n is defined in the proof of Theorem 2.3.1. The hypothesis H.3 implies that Q_n and $Q_n^{(j)}$ tend to 0 in L^2 in probability. In view of Proposition 2.4.1 T_n tends to 0 in L^2 in probability. By using the fact that $\partial C_\theta^{(n)}/\partial u$ and $\partial C_\theta^{(n)}/\partial v$ converge uniformly in probability to $\partial C_\theta/\partial u$ and $\partial C_\theta/\partial v$ respectively (Rémillard and Scaillet (2009), Prop. A.2.), and making use of the fact following, from Assumption H.2, that $\dot{C}_{\theta_n}$ converges uniformly in probability to $\dot{C}_\theta$, we obtain that $w\Psi_n - w_n\Gamma_n$ tends to 0 in L^2 in probability. Since $w\Psi_n$ converges weakly in L^2 to $w\Gamma$, this competes the proof of the theorem. □

We define next the second important class, denoted by $\mathcal{R}_2$, of rank-based estimators of θ. We state and prove as well the analogue of Theorem 2.6.1 for the class $\mathcal{R}_2$.

Definition 2.6.2 *A rank-based estimator θ_n of θ is said to belong to the class $\mathcal{R}_2$ if*

$$\Theta_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} J_\theta(U_i, V_i) + o_P(1),$$

where $(U_i, V_i) := (F_1(X_i), F_2(Y_i))$ for all $i \in \{1, ..., n\}$ and $J_\theta : [0,1]^2 \to \mathbb{R}$ is a score function that satisfies the following regularity conditions:

(a) for all $\theta \in \Delta$, J_θ is bounded on $[0,1]^2$ and centered, i.e., $\int_{[0,1]^2} J_\theta(u,v)dC_\theta(u,v) = 0$,

(b) for all $\theta \in \Delta$, the partial derivatives $J_\theta^{[1]} = \partial J_\theta/\partial u$, $J_\theta^{[2]} = \partial J_\theta/\partial v$, exist and are bounded on $[0,1]^2$,

(c) for all $\theta \in \Delta$, the partial derivative $\dot{J}_\theta = \partial \dot{J}_\theta/\partial\theta$, exists and is bounded on $[0,1]^2$,

(d) for all $\theta \in \Delta$, there exist $c_\theta > 0$ and $M_\theta > 0$ such that, if $|\theta' - \theta| < c_\theta$, then $|\dot{J}_{\theta'}| \leq M_\theta$.

It is well known fact that (see e.g. Berg and Quessy (2009) and the references therein) that the estimator $\theta_{\tau,n}$ based on the one of the most popular measure of association - Kendall's tau, belongs to the class $\mathcal{R}_2$. Its score function is given by

$$J_{\theta,\tau}(u,v) = \frac{4}{\tau'(\theta)}\left(2C_\theta(u,v) - u - v + \frac{1-\tau(\theta)}{2}\right), \quad (u,v) \in [0,1]^2.$$

The following result, which is the analogue of Theorem 2.6.1 for class $\mathcal{R}_2$, is instrumental for verifying the asymptotic validity of the fast goodness-of-fit procedure proposed in Section 2.4.3.2 for the estimators of θ belonging to the class $\mathcal{R}_2$. Its proof is omitted as it is very similar to and simpler than that of Theorem 2.6.1.

Theorem 2.6.2 *Let θ_n be an estimator of θ belonging to class $\mathcal{R}_2$ and, for any $k \in \{1, ..., N\}$, let*

$$\Theta_n^{(k)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i^{(k)} \mathcal{J}_{\theta_n,i,n},$$

where $\mathcal{J}_{\theta_n,i,n} := J_{\theta_n}(U_{i,n}, V_{i,n})$. Then, under Assumptions H.1-H.4,

$$w_n \Gamma_n := \left([\sqrt{n}(\overline{C}_n - C_{\theta_n})] w_{\theta_n}, [\mathbb{C}_n^{(1)} - \Theta_n^{(1)} \dot{C}_{\theta_n}] w_{\theta_n}, ..., [\mathbb{C}_n^{(N)} - \Theta_n^{(N)} \dot{C}_{\theta_n}] w_{\theta_n} \right)$$

converges weakly to

$$w\Gamma := \left([\mathbb{B}^* - \Theta \dot{C}_\theta] w_\theta, [\mathbb{B}^{*(1)} - \Theta^{(1)} \dot{C}_\theta] w_\theta, ..., [\mathbb{B}^{*(N)} - \Theta^{(N)} \dot{C}_\theta] w_\theta \right)$$

in $L^2\left([0,1]^d\right)^{\otimes(N+1)}$, where Θ is the weak limit of $\Theta_n = \sqrt{n}(\theta_n - \theta)$ and $(\mathbb{B}^{*(1)}, \Theta^{(1)}), ..., (\mathbb{B}^{*(N)}, \Theta^{(N)})$ are independent copies of $(\mathbb{B}^*, \Theta)$.

2.6.2 Verification of hypothesis H.2 for various copula models

2.6.2.1 Preliminaries

We consider the parametric family of copulas $\{C_\theta : \theta \in \Delta\}$, where Δ is an open subset of $\mathbb{R}^p$.

Recall from (2.4.3) on page 62 definition of $\dot{C}_\theta$. The following two conditions can be used to verify hypothesis H.2 :

$$\text{The function } \dot{C}(\theta, \mathbf{u}) := \dot{C}_\theta(\mathbf{u}) \text{ is continuous on } \Delta \times [0,1]^d, \tag{2.6.2}$$

$$\sup_{\|\theta^* - \theta\| < \varepsilon} \left\{ \sup_{\mathbf{u} \in [0,1]^d} \left| \ddot{C}_\theta(\mathbf{u}) - \ddot{C}_\theta(\mathbf{u}) \right| \right\} < \infty, \text{ for some } \varepsilon > 0, \tag{2.6.3}$$

where $\ddot{C}_\theta(\mathbf{u}) := \nabla_\theta \dot{C}_\theta(\mathbf{u}) = \left(\frac{\partial}{\partial \theta_1} \dot{C}_\theta(\mathbf{u}), .., \frac{\partial}{\partial \theta_p} \dot{C}_\theta(\mathbf{u}) \right)^T.$

Proposition 2.6.1 *If the family of copulas* $\{C_\theta : \theta \in \Delta\}$ *fulfil condition (2.6.2) or condition (2.6.3) then it satisfies hypothesis H.2.*

Proof. Let first assume condition (2.6.2). From the continuity of function $\dot{C}_\theta$ on $[0,1]^d$ there exist $\mathbf{u}^*$ (depending on θ^*) such that

$$\left|\dot{C}_{\theta^*}(\mathbf{u}^*) - \dot{C}_\theta(\mathbf{u}^*)\right| = \sup_{\mathbf{u}\in[0,1]^d} \left|\dot{C}_{\theta^*}(\mathbf{u}) - \dot{C}_\theta(\mathbf{u})\right|$$

The continuity of function $\dot{C}_\theta$ on Δ implies that, for $\delta > 0$ and every sequence ε_n ($\varepsilon_n \to 0$) there exist N such that for $n \geq N$, $\|\theta^* - \theta\| < \varepsilon_n$ and

$$\sup_{\|\theta^*-\theta\|<\varepsilon_n} \left|\dot{C}_{\theta^*}(\mathbf{u}^*) - \dot{C}_\theta(\mathbf{u}^*)\right| < \delta.$$

Condition (2.6.3) implies the hypothesis H.5 by mean value theorem. □

Note that similar derivations included in this section can be partially found in : (Quessy, 2005, p. 74-75), Appendix B in (Genest et al, 2006, p. 23-28), Appendix B in (Berg and Quessy, 2009, p. 27-29).

2.6.2.2 Multivariate Archimedean copulas

The Archimedean representation allows us to reduce the study of a multivariate copula to a single univariate function. The next theorem generalizes the concepts of bivariate Archimedean copula (see (1.1.10) on page 20) for the d-dimensional case ($d \geq 3$).

Definition 2.6.3 *A function f is completely monotonic in an interval* $[a, b]$ *if for* $t \in [a, b]$ *and* $k \in N$ *it satisfies*

$$(-1)^k \frac{d^k}{dt^k} f(t) \geq 0.$$

Assume that ϕ is a convex, strictly decreasing function with domain $(0, 1]$ and range $[0, \infty]$ such that $\phi(0) = 1$. Use ϕ^{-1} for the generalized inverse function of ϕ defined via (1.1.1). We call ϕ a generator of an Archimedean copula.

Theorem 2.6.3 *The function*

$$C^{d-Arch}(\mathbf{u}) = \phi^{-1}\left(\sum_{j=1}^{d} \phi(u_j)\right), \ \textit{for } \mathbf{u} = (u_1, ..., u_d), \tag{2.6.4}$$

is a copula for all $d \geq 2$ *if and only if* ϕ^{-1} *is completely monotonic in* $[0, 1)$.

Proof. See Proof of Theorem 2.2. in McNeil and Nešlehová (2009). □

For more information and properties of Archimedean copulas, see for example Joe (1997).

Proposition 2.6.2 *For Archimedean copulas, we have*

$$\dot{C}_\theta^{d-Arch}(\mathbf{u}) = \frac{\sum_{j=1}^{d} \dot{\phi}_\theta(u_j) - \dot{\phi}_\theta(C_\theta^{d-Arch}(\mathbf{u}))}{\phi'_\theta(C_\theta^{d-Arch}(\mathbf{u}))}.$$

Proof. From (2.6.4) we may write

$$\phi_\theta(C_\theta^{d-Arch}(\mathbf{u})) = \sum_{j=1}^{d} \phi_\theta(u_j).$$

Applying the chain rule for a function of several variables finishes the proof of Proposition 2.6.2. □

In the sequel, to simplify our exposition, we use the common notation C_θ for different families of Archimedean copulas (Clayton, Frank and Gumbel-Hougaard).

- **Clayton** family of copulas. For this family $\theta > 0$, therefore $\Delta = (0, \infty)$. It is well known that (see for example Quessy (2005), p. 75)

$$\begin{aligned}
\phi_\theta(t) &= \frac{1}{\theta} t^{-\theta} - 1 \\
\phi_\theta^{-1}(t) &= (1 + \theta s)^{-1/\theta} \\
C_\theta(\mathbf{u}) &= \left(\sum_{j=1}^{d} u_j - d + 1\right)^{-1/\theta} \\
\phi'_\theta(t) &= -t^{-\theta-1} \\
\dot{\phi}_\theta(t) &= \frac{1 - t^{-\theta} - \theta t^{-\theta} \ln t}{\theta^2}
\end{aligned}$$

From Proposition 2.6.1, the function

$$\dot{C}_\theta(\mathbf{u}) = \frac{C_\theta(\mathbf{u})}{\theta}\left\{\frac{\sum_{j=1}^d u_j^{-\theta}\log u_j}{\sum_{j=1}^d u_j^{-\theta} - d + 1} - \log C_\theta(\mathbf{u})\right\}$$

is continuous on $(0,\infty)\times[0,1]^d = \Delta\times[0,1]^d$. Hence from Proposition 2.6.2, for Clayton family, hypothesis H.2 holds for $\Delta = (0,\infty)$.

- **Frank** family of copulas. For $\theta \in \mathbb{R}\setminus\{0\}$ we have

$$\begin{aligned}
\phi_\theta(t) &= \ln\frac{e^{\theta t}-1}{e^\theta - 1},\\
\phi_\theta^{-1}(s) &= \theta^{-1}\ln\left[1 + e^s(e^\theta - 1)\right]\\
C_\theta(\mathbf{u}) &= \theta^{-1}\ln\left[1 + \frac{\prod_{j=1}^d (e^{\theta u_j}-1)}{(e^\theta-1)^{d-1}}\right]\\
\phi_\theta^{'}(t) &= \theta\frac{(e^{\theta t}-1)}{(e^{\theta t}-1)}\\
\dot{\phi}_\theta(t) &= \frac{te^{\theta t}}{(e^{\theta t}-1)} - \frac{e^{\theta t}-1}{e^\theta-1}
\end{aligned}$$

We use Proposition 2.6.1 to compute $\dot{C}_\theta$. The function

$$\dot{C}_\theta(\mathbf{u}) = \frac{1}{\theta}\left[\frac{\psi_1(\theta,\mathbf{u}) - (d-1)\frac{e^\theta}{e^\theta-1}}{\psi_2(\theta,\mathbf{u})} - C_\theta(\mathbf{u})\right],$$

where $\psi_1(\theta,\mathbf{u}) = \sum_{j=1}^d u_j \frac{e^{\theta u_j}}{e^{\theta u_j}-1}$ and

$$\psi_2(\theta,\mathbf{u}) = \frac{1 + \frac{\prod_{j=1}^d (e^{\theta u_j}-1)}{(e^\theta-1)^{d-1}}}{\frac{\prod_{j=1}^d (e^{\theta u_j}-1)}{(e^\theta-1)^{d-1}}} = 1 + \frac{(e^\theta-1)^{d-1}}{\prod_{j=1}^d (e^{\theta u_j}-1)},$$

is continuous on $\mathbb{R}\setminus\{0\}\times[0,1]^d = \Delta\times[0,1]^d$. Hence from Proposition 2.6.2, for Frank family, hypothesis H.2 holds for $\Delta = \mathbb{R}\setminus\{0\}$.

- **Gumbel-Hougaard** family of copulas. For $\theta \geq 1$, we have

$$\begin{aligned}
\phi_\theta(t) &= (-\ln t)^\theta, \\
\phi_\theta^{-1}(s) &= e^{-s^{1/\theta}}, \\
\psi_\theta^G(\mathbf{u}) &:= \sum_{j=1}^{d} (-\ln u_j)^\theta \\
C_\theta(\mathbf{u}) &= \exp\left[-\psi_\theta^G(\mathbf{u})^{1/\theta}\right] \\
\phi_\theta^{'}(t) &= -\frac{\theta}{t}(-\ln t)^{\theta-1} \\
\dot{\phi}_\theta(t) &= (-\ln t)^\theta \ln(-\ln t).
\end{aligned}$$

  We use Proposition 2.6.1 to compute

$$\dot{C}_\theta(\mathbf{u}) = \frac{C_\theta(\mathbf{u})\left[\varphi_\theta^G(\mathbf{u}) - \psi_\theta^G(\mathbf{u}) \times \ln[\psi_\theta^G(\mathbf{u})^{1/\theta}]\right]}{-\theta[\psi_\theta^G(\mathbf{u})]^{(\theta-1)/\theta}}, \tag{2.6.5}$$

  where $\varphi_\theta^G(\mathbf{u}) = \sum_{j=1}^{d}(-\ln u_j)^\theta \times \ln(-\ln u_j)$. The function $\dot{C}_\theta$ is continuous on $[1,\infty) \times [0,1]^2$. Hence from Proposition 2.6.2, for Gumbel-Hougaard family, hypothesis H.2 holds for $\Delta = [1,\infty)$.

For simplicity of the presentation, we restrict ourselves here to the case $d = 2$ for two last examples of copulas. Following Appendix B.2 of Kojadinovic et al (2010) (p.26) we have:

2.6.2.3 Gaussian (Normal) copula

Recall the definition of the bivariate Gaussian (or Normal) copula given in Example 1.1.1 on page 19. The expression of $\dot{C}_\rho^{Ga}(u,v) := \frac{\partial}{\partial\rho} C^{Ga}(u,v)$ follows from the so-called Plackett formula (see Plackett (1954)):

$$\frac{\partial}{\partial\rho}\Phi_2(h,k,\rho) = \frac{\exp\left(\frac{h^2-2\rho hk+k^2}{2(1-\rho^2)}\right)}{2\pi\sqrt{1-\rho^2}},$$

where $\Phi_2(\cdot,\cdot,\theta)$ is the bivariate standard normal cumulative distribution function (cdf) with correlation $\rho \in (-1,1)$.
Since $\frac{\partial}{\partial\rho} C^{Ga}(u,v) = \frac{\partial}{\partial\rho}\Phi_2(\Phi^{-1}(u),\Phi^{-1}(v);\rho)$ and from the continuity of $\frac{\partial}{\partial\rho}\Phi_2$ we obtain that the function $\dot{C}_\rho^{Ga}$ is continuous on $(-1,1) \times [0,1]^2$. In view of Proposition 2.6.1 on page 84 this function verifies the hypothesis H.5 with $\Delta = (-1,1)$.

2.6.2.4 Student copula

The bivariate *Student* copula with m degrees of freedom is defined via an application of Sklar's theorem (see Theorem 1.1.1 on page 14 and Nelsen (2006)) by

$$C^t(u,v) := \int_{-\infty}^{t_m^{-1}(u)} \int_{-\infty}^{t_m^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{s^2 - 2\rho st + t^2}{m(1-\rho^2)}\right)^{-(m+2)/2} ds\ dt, \tag{2.6.6}$$

where $t_m^{-1}(\cdot)$ is the quantile function of the Student distribution with m degrees of freedom.

The expression of $\dot{C}^t_\rho(u,v) := \frac{\partial}{\partial\rho} C^t(u,v)$ follows from generalization of the Plackett formula given in Genz (2004):

$$\frac{\partial}{\partial\rho} t_2(h,k,\rho,m) = \frac{\left(1 + \frac{h^2 - 2\rho hk + k^2}{m(1-\rho^2)}\right)}{2\pi\sqrt{1-\rho^2}},$$

where $t_2(\cdot,\cdot,\rho,m)$ is the bivariate standard student cdf with m degrees of freedom and correlation $\rho \in (-1,1)$.

Since $\frac{\partial}{\partial\rho} C^t(u,v) = \frac{\partial}{\partial\rho} t_2(t_m^{-1}(u), t_m^{-1}(v); \rho)$ and from the continuity of $\frac{\partial}{\partial\rho} t_2$ we obtain that the function $\dot{C}^t_\rho$ is continuous on $(-1,1) \times [0,1]^2$. In view of Proposition 2.6.1 on page 84, this function verifies the hypothesis H.5 with $\Delta = (-1,1)$.

2.6.3 Verification of hypothesis H.5 for various copula models

For simplicity of the presentation, we restrict ourselves here to the case $d = 2$. Note that some of derivations and ideas presented in this section can be found in Appendix D of Omelka et al (2009).

2.6.3.1 Archimedean copulas

Recall from (1.1.10) on page 20 the definition of the bivariate Archimedean copula C^{Arch} with generator ϕ. We have by direct computations

$$\frac{\partial}{\partial u} C^{Arch}_\theta(u,v) = \frac{\phi'_\theta(u)}{\phi'_\theta(C^{Arch}_\theta(u,v))}$$

The partial derivative of $\frac{\partial}{\partial u} C_\theta^{Arch}$ with respect to v is given by

$$c_\theta^{Arch}(u,v) = \frac{\partial^2}{\partial v \partial u} C_\theta^{Arch}(u,v) = \frac{-\phi_\theta^{'}(u) \times \phi_\theta^{''}(C_\theta^{Arch}(u,v)) \frac{\phi_\theta^{'}(v)}{\phi_\theta^{'}(C_\theta^{Arch}(u,v))}}{[\phi_\theta^{'}(C_\theta^{Arch}(u,v))]^2}$$
$$= \frac{-\phi_\theta^{'}(u)\phi_\theta^{'}(v) \times \phi_\theta^{''}(C_\theta^{Arch}(u,v))}{[\phi_\theta^{'}(C_\theta^{Arch}(u,v))]^3}$$

In the sequel, to simplify our exposition, we use the common notation C_θ for different families of Archimedean copulas (Clayton, Frank and Gumbel-Hougaard).

- **Gumbel-Hougaard** family of copulas. We recall that for a Gumbel copula

$$\phi^{'}(t) = \theta(-\ln t)^{\theta-1}\left(-\frac{1}{t}\right)$$
$$\phi^{''}(t) = \theta(\theta-1)(-\ln t)^{\theta-2}\left(\frac{1}{t^2}\right) + \theta(-\ln t)^{\theta-1}\left(\frac{1}{t^2}\right),$$

which implies that

$$c_\theta(u,v) = \psi_\theta(u,v) + \varphi_\theta(u,v),$$

where

$$\psi_\theta(u,v) = \frac{-C_\theta(u,v)}{uv} \times \frac{[-\ln u]^{\theta-1}[-\ln v]^{\theta-1}}{[-\ln(C_\theta(u,v))]^{2\theta-2}},$$
$$\varphi_\theta(u,v) = \frac{-C_\theta(u,v)}{uv} \times \frac{[-\ln u]^{\theta-1}[-\ln v]^{\theta-1}}{[-\ln(C_\theta(u,v))]^{2\theta-1}}.$$

The Fréchet-Hoeffding upper bound for a copula C_θ ($C_\theta(u,v) \leq \min\{u,v\}$) implies that: $-\ln[C_\theta(u,v)] \geq -\ln u$ and $-\ln[C_\theta(u,v)] \geq -\ln v$. Direct computations yield

$$|\psi_\theta(u,v)| = \frac{C_\theta(u,v)^{1/2} \cdot C_\theta(u,v)^{1/2}}{uv} \times \left| \frac{[-\ln u]^{\theta-1}[-\ln v]^{\theta-1}}{[-\ln(C_\theta(u,v))]^{\theta-1}[-\ln(C_\theta(u,v))]^{\theta-1}} \right|$$
$$\leq \frac{u^{1/2} \cdot v^{1/2}}{uv} \times \left| \frac{[-\ln u]^{\theta-1}[-\ln v]^{\theta-1}}{[-\ln u]^{\theta-1}[-\ln v]^{\theta-1}} \right| = \frac{1}{u^{1/2}v^{1/2}} = O\left(\frac{1}{\sqrt{uv}}\right)$$

Hence $|\psi_\theta(u,v)| = O\left(\frac{1}{\sqrt{uv}}\right)$,when $(u,v) \to (0_+, 0_+)$.

Using the fact that $\frac{C_\theta(u,v)}{uv}$ is bounded when $(u,v) \to (1_-, 1_-)$ we obtain, for the second term of density c_θ

$$\begin{aligned}
|\varphi_\theta(u,v)| &\leq \frac{[-\ln u]^{\theta-1} \times [-\ln v]^{\theta-1}}{[-\ln(C_\theta(u,v))]^{\theta-1} \times [-\ln(C_\theta(u,v))]^{\theta-1} \times [-\ln(C_\theta(u,v))]} \\
&\leq \frac{[-\ln u]^{\theta-1} \times [-\ln v]^{\theta-1}}{[-\ln u]^{\theta-1} \times [-\ln v]^{\theta-1} \times [-\ln u]^{1/2} \times [-\ln v]^{1/2}} \\
&= \frac{1}{[-\ln u]^{1/2} \times [-\ln v]^{1/2}}.
\end{aligned}$$

Since $[-\ln u]^{-1/2} \times [-\ln v]^{-1/2} = O\left((1-u)(1-v)^{-1/2}\right)$, when $(u,v) \to (1_-, 1_-)$, it implies that $|\varphi_\theta(u,v)| = O\left((1-u)(1-v)]^{-1/2}\right)$.

In view of the above facts on the functions ψ and φ, the density c_θ of a *Gumbel-Hougaard* copula is $O\left([uv(1-u)(1-v)]^{-1/2}\right)$ which means that it holds the hypothesis H.5.

- **Clayton** family of copulas. We recall that for a *Clayton* copula

$$\begin{aligned}
\phi'(t) &= -t^{-\theta-1} \\
\phi''(t) &= (\theta+1)t^{-\theta-2},
\end{aligned}$$

which implies that

$$\begin{aligned}
|c_\theta(u,v)| &= \left|\frac{(\theta+1)[C_\theta(u,v)]^{3\theta+3}}{(uv)^{\theta+1}[C_\theta(u,v)]^{\theta+2}}\right| = \left|\frac{(\theta+1)[C_\theta(u,v)]^{2\theta+1}}{u^{\theta+1}v^{\theta+1}}\right| \\
&= \left|\frac{(\theta+1)[C_\theta(u,v)]^{\theta+1/2}[C_\theta(u,v)]^{\theta+1/2}}{u^{\theta+1}v^{\theta+1}}\right| \\
&\leq \left|\frac{(\theta+1)u^{\theta+1/2}v^{\theta+1/2}}{u^{\theta+1}v^{\theta+1}}\right| = \left|\frac{(\theta+1)}{u^{1/2}v^{1/2}}\right|.
\end{aligned}$$

Hence $|c_\theta(u,v)| = O\left([uv(1-u)(1-v)]^{-1/2}\right)$ which means that density of Clayton copula satisfies the hypothesis H.5.

- **Frank** family of copulas. We recall that for a *Frank* copula

$$\phi_\theta^{'}(t) = \theta \frac{e^{\theta t}}{(e^{\theta t}-1)}$$
$$\phi_\theta^{''}(t) = -\theta^2 \frac{e^{\theta t}}{(e^{\theta t}-1)^2},$$

which implies that

$$c_\theta(u,v) = \frac{e^{\theta u} e^{\theta u}(e^\theta - 1)}{(e^\theta - 1) + (e^{\theta u}-1)(e^{\theta v}-1)}.$$

Since the function c_θ is bounded on $[0,1]^2$, for all $\theta \in \mathbb{R} \setminus \{0\}$, it is $O\left([uv(1-u)(1-v)]^{-1/2}\right)$. This implies that a density of a Frank copula satisfies the hypothesis H.5.

2.6.3.2 Gaussian (Normal) copula

Recall the definition of the bivariate Gaussian (or Normal) copula given in Example 1.1.1 on page 19. The density function c^{Ga} of the bivariate Gaussian copula C^{Ga} is given by

$$c^{Ga}(u,v) = \frac{\partial^2}{\partial u \partial v} C^{Ga}(u,v) = \frac{\varphi_2(\Phi^{-1}(u), \Phi^{-1}(v); \rho)}{\varphi(\Phi^{-1}(u))\varphi(\Phi^{-1}(v))}, \tag{2.6.7}$$

where φ_2 is defined in Example 1.1.4 on page 21 and the function φ denotes the density of the standard normal $N(0,1)$ law. From tail behavior of density-quantile function $\varphi \circ \Phi$ (see Grimshaw (1989) p.9),

$$\varphi(\Phi(1-u)) \sim u(-2\ln u)^{1/2}, \text{ as } u \to 0^+,$$

and

$$\varphi(\Phi(u)) \sim (1-u)[-2\ln(1-u)]^{1/2}, \text{ as } u \to 1_-.$$

Since the function φ_2 is bounded and

$$[\varphi(\Phi^{-1}(u))\varphi(\Phi^{-1}(v)]^{-1} = o([uv(1-u)(1-v)]^{-1}) = O([uv(1-u)(1-v)]^{-1/2}),$$

the density function c^{Ga} of Gaussian copula fulfils hypothesis H.5.

2.6.3.3 Student copula

Recall from (2.6.6) on page 88 the definition of the bivariate *Student* copula C^t (with m degrees of freedom).

The density function c^t of the bivariate *Student* copula C^t (with m degrees of freedom) is give by

$$c^t(u,v) = \frac{f_{2,m}(t_m^{-1}(u), t_m^{-1}(v); \rho)}{f_m(t_m^{-1}(u)) \times f_m(t_m^{-1}(v))},$$

where $f_{2,m}$ is a density function of the bivariate *Student* distribution and f_m is a density of *Student* univariate distribution (with m degrees of freedom). Thus, applying l'Hôpital's rule gives

$$\begin{aligned} \frac{t_m^{-1}(u)}{u} &\sim \frac{1}{f_m(t_m^{-1}(u))} \quad \text{for } u \to 0_+ \\ \frac{t_m^{-1}(u)}{1-u} &\sim -\frac{1}{f_m(t_m^{-1}(u))} \quad \text{for } u \to 1_- \end{aligned}$$

Since the function $f_{2,m}$ is bounded on $\mathbb{R}^2$ and

$$\begin{aligned} \frac{t_m^{-1}(u)}{u} &= o(u^{-1}) \quad \text{for } u \to 0_+, \\ \frac{t_m^{-1}(u)}{1-u} &= o((1-u)^{-1}) \quad \text{for } u \to 1_-, \end{aligned}$$

we obtain that

$$|c^t(u,v)| = O([uv(1-u)(1-v)]^{-1/2}),$$

which means that hypothesis H.5 holds for density of Student copula.

2.6.4 Central Limit Theorem (CLT) in L^p-space

The following fact constitutes a version of the CLT version of random variables taking their values in L^p-space (see e.g., Example 1.8.5, p.50, the discussion on p.92-93 in van der Vaart and Wellner (1996) and Ex.14, p. 205 in Araujo and Giné (1980)).

Fact 2.6.1 *Let (S, Σ, μ) be a σ-finite measure space and $Y_1, ..., Y_n$ are i.i.d., zero-mean Borel measurable maps into $L^p(S, \Sigma, \mu)$, then the sequence $n^{-1/2} \sum_{i=1}^n Y_i$ converges weakly to Y-a centered Gaussian $L^p(S, \Sigma, \mu)$-valued random variable with the covariance function of Y_1, if and only if $\mathbb{P}(\|Y_1\|_p > t) = o(t^{-2})$ as $t \to \infty$ and*

$$\int_S (\mathbb{E}Y_1^2(t))^{p/2}\, d\mu(t) < \infty.$$

(In case $p = 2$, this can be simplified to the single requirement $\mathbb{E}\|Y_1\|_2^2 < \infty$).

Proposition 2.6.3 *For every copula function C and for $j = 1, ..., d$, the partial derivative C_j exist for almost all $\mathbf{u} \in [0,1]^d$ with respect to the Lebesgue-measure. For such $\mathbf{u}$, we have*

$$0 \le C_j(\mathbf{u}) \le 1, \;\; \mathbf{u} \in [0,1]^d.$$

Proof. See Theorem 2.2.7 of Nelsen (2006). □

2.6.5 Variance function of the limiting Gaussian process $\mathbb{B}^*$

For $\mathbf{u} \in [0,1]^d$, we have

$$[\mathbb{B}^*(\mathbf{u})]^2 = \left[\mathbb{B}(\mathbf{u}) - \sum_{j=1}^d C_j(\mathbf{u})\mathbb{B}(\mathbf{1}, u_j, \mathbf{1})\right]^2 = \mathbb{B}^2(\mathbf{u}) + V_1(\mathbf{u}) + V_2(\mathbf{u}),$$

where

$$\begin{aligned}
V_1(\mathbf{u}) &= -2\sum_{j=1}^d C_j(\mathbf{u})\mathbb{B}(\mathbf{u})\mathbb{B}(\mathbf{1}, u_j, \mathbf{1}) \\
V_2(\mathbf{u}) &= \sum_{i=1}^d \sum_{j=1}^d C_i(\mathbf{u})C_j(\mathbf{u})\mathbb{B}(\mathbf{1}, u_i, \mathbf{1})\mathbb{B}(\mathbf{1}, u_j, \mathbf{1}) \\
&= \sum_{i=1}^d C_i^2(\mathbf{u})\mathbb{B}^2(\mathbf{1}, u_i, \mathbf{1}) + 2\sum_{i<j} C_i(\mathbf{u})C_j(\mathbf{u})\mathbb{B}(\mathbf{1}, u_i, \mathbf{1})\mathbb{B}(\mathbf{1}, u_j, \mathbf{1})
\end{aligned}$$

From (2.2.13) we obtain that, for $i < j$

$$\begin{aligned}\mathbb{E}[\mathbb{B}^2(\mathbf{u})] &= C(\mathbf{u})[1 - C(\mathbf{u})] \\ \mathbb{E}[\mathbb{B}(\mathbf{u})\mathbb{B}(\mathbf{1}, u_j, \mathbf{1})] &= C(\mathbf{u})[1 - u_j] \\ \mathbb{E}[\mathbb{B}^2(\mathbf{1}, u_i, \mathbf{1})] &= u_i(1 - u_i) \\ \mathbb{E}[\mathbb{B}(\mathbf{1}, u_i, \mathbf{1})\mathbb{B}(\mathbf{1}, u_j, \mathbf{1})] &= C(\mathbf{1}, u_i, u_j, \mathbf{1}) - u_i u_j\end{aligned}$$

Hence we get

$$\begin{aligned}Var(\mathbb{B}^*(\mathbf{u})) = \mathbb{E}\left\{[\mathbb{B}^*(\mathbf{u})]^2\right\} &= C(\mathbf{u})[1 - C(\mathbf{u})] - 2\sum_{j=1}^{d} C_j(\mathbf{u})C(\mathbf{u})[1 - u_j] \\ &+ \sum_{i=1}^{d} C_i^2(\mathbf{u})u_i(1 - u_i) \\ &+2 \sum_{i<j} C_i(\mathbf{u})C_j(\mathbf{u})[C(\mathbf{1}, u_i, u_j, \mathbf{1}) - u_i u_j]\end{aligned}$$

For $d = 2$ we obtain that

$$\begin{aligned}Var(\mathbb{B}^*(u, v)) = C(u, v)[1 - C(u, v)] - 2C_1(u, v)C(u, v)[1 - u] \\ - 2C_2(u, v)C(u, v)[1 - v] + C_1^2(u, v)u(1 - u) \\ + C_2^2(u, v)v(1 - v) + 2C_1(u, v)C_2(u, v)[C(u, v) - uv].\end{aligned} \tag{2.6.8}$$

For $d = 2$ and for $C(u, v) = uv$ (independence copula):

$$\begin{aligned}Var(\mathbb{B}^*(u, v)) &= uv[1 - uv] - 2vuv[1 - u] \\ &- 2uuv[1 - v] + v^2u(1 - u) \\ &= uv[1 - uv]_u v(v(1 - u)) - uv(u(1 - v)) \\ &= uv - uv^2 - u^2v + u^2v^2 \\ &= u(v - v^2) - u^2(v - v^2) \\ &= u(1 - u)v(1 - v).\end{aligned}$$

2.6.6 An example of a very non-smooth copula

For $0 < \varepsilon < 1$, a bivariate copula is constructed such that its first-order partial derivatives are not continuous on a set of Lebesgue measure $1-\varepsilon$. Let $\{r_1, r_2, ...\}$ be an enumeration of the rationals of the open interval $(0, 1)$. For every integer $n \geq 1$, let $0 < a_n < r_n < b_n < 1$ be such that $b_n - a_n < \varepsilon 2^{-n}$. Let $A = \bigcup_{n\geq 1}(a_n, b_n)$. Let λ_1 denote the Lebesgue measure on $\mathbb{R}$. Then $A \subset (0,1)$ and $\lambda_1(A) \leq \sum_{n\geq 1} \lambda_1(a_n, b_n) < \sum_{n\geq 1} \varepsilon 2^{-n} = \varepsilon$. Clearly, A is open and A is dense in $[0, 1]$.
Set $B = [0,1] \setminus A$. Then B is closed and $\lambda_1(B) > 1 - \varepsilon$. [Moreover, B is nowhere dense in $[0, 1]$, that is, for every non-empty open interval $I \subset (0, 1)$ there exists a non-empty open interval $J \subset I$ such that $J \cap B = \emptyset$. Indeed, since A is dense, $A \cap I$ is non-empty. But since A and I are open, their intersection $A \cap I$ must be open too. Hence there exists a non-empty open interval J contained in $A \cap I = I \setminus B$. The construction of the sets A and B is inspired by Example 3.1 on p. 44 in Billingsley (1995)].
Let (U, V) be a pair of random variables with the following joint distribution:

- U is uniformly distributed on $(0, 1)$;
- conditionally on $U \in A$, the variable V is uniformly distributed on A;
- conditionally on $U \in B$, the variable V is uniformly distributed on B.

The distribution function of the pair (U, V) is given by

$$\begin{aligned} C(u,v) &= P(U \leq u, V \leq v) \\ &= P(U \in A \cap [0,u], V \in A \cap [0,v]) + P(U \in B \cap [0;u], V \in B \cap [0,v]) \\ &= \frac{\lambda_1(A \cap [0,u])\lambda_1(A \cap [0,v])}{\lambda_1(A)} + \frac{\lambda_1(B \cap [0,u])\lambda_1(B \cap [0,v])}{\lambda_1(B)}, \end{aligned}$$

for $(u, v) \in [0, 1]^2$. It is readily checked that the copula C is absolutely continuous with respect to the two-dimensional Lebesgue measure λ_2. Its Radon-Nikodym derivative is given by

$$c(u,v) = \frac{1}{\lambda_1(A)} \mathbf{1}_{A\times A}(u,v) + \frac{1}{\lambda_1(B)} \mathbf{1}_{B\times B}(u,v).$$

The first-order partial derivative of C with respect to u is given by

$$C_1(u,v) = \int_0^v c(u,v')dv' = \begin{cases} \lambda_1(A \cap [0,v])/\lambda_1(A), & \text{if } u \in A, \\ \lambda_1(B \cap [0,v])/\lambda_1(B), & \text{if } u \in B, \end{cases}$$

for $(u,v) \in [0,1]^2$. Let $N = \{v \in [0,1] : \lambda_1(A \cap [0,v])/\lambda_1(A) = \lambda_1(B \cap [0,v])/\lambda_1(B))\}$. Since A and B are disjoint, the set N is a null-set, $\lambda_1(N) = 0$. Since A is dense in $[0,1]$, we see that C_1 is not continuous in $B \times ([0,1] \setminus N)$. Hence, C_1 is not continuous on a set of λ_2-measure $1 - \varepsilon$.

Chapter 3

Paper II - Generalized Framework of Mack Stochastic Chain Ladder Method

3.1 Motivation

In the study of Mack (1994), author extended his stochastic chain ladder method from Mack (1993) in two ways. Firstly, he provided different estimators of chain ladder factors. Secondly, he introduced the possibility to exclude the individual age-to-age factors (link ratios) from the estimation of the main parameters. The last extension is especially important from practical point of view. In fact, reserving actuaries often have concerns about particular events and unusual effects which make affect the estimation process. Such external knowledge is considered in setting the weights on link ratios. However, the elimination of link ratios, in general, reduces artificially the MSEP of claims reserves. To overcome this difficulty, we provide in the present paper a general reserving tool allowing the use of the expert judgment and analyze of the data. In our approach, the weights can be set differently for the estimation of the amount of reserves and for the estimation of the variance parameters. This is the common practice of actuaries even though the estimation of the parameters in that case is not optimal in the sense of Proposition 3.3.1 (ii) - (iv). The Chain Ladder model of Mack (1994) turns out to be a particular case of our general framework. In the final section we apply our results to cover the methods of claims reserving often used by practitioners. We will make an instrumental use of the notation and methods of Mack (1993), Mack (1994) and Mack (1999) .We will follow the arguments of proofs of the main results of these articles. Finally, our general framework can be applied in the case of incomplete run-off triangles (see Remark 3.6.2).

Organization of the paper. In Section 3.2 we present our notation and Section 3.3 introduces the model and derives the structure of the mean and the variance of the claims process and defines estimators of the unknown parameters. The definition of the MSEP is given in Section 3.4. In Section 3.5 we derive estimators of the conditional MSEP of the ultimate amount of claims reserves for single and for aggregated accident years. A numerical example is presented in Section 3.6. Some concluding remarks are given in Section 3.7 and Section 3.8. Finally, all proofs are provided in the Section 3.9 .

3.2 Notation

Let $C_{i,j}$ denote the cumulative payments for accident year $i \in \{1, ..., I\}$ until development year $j \in \{1, ..., J\}$, where the accident year is referred to as the year in which an event triggering insurance claims occurs. This means that the ultimate claim for accident year i is given by $C_{i,J}$. We assume that the cumulative payments $C_{i,j}$ are random variables observable for calendar years $i + j \leq I + 1$ and non-observable (to be predicted) for calendar years $i+j > I+1$. The observable cumulative payments are represented by the so-called run-off trapezoids($I > J$) or run-off triangles ($I = J$). Table 3.1 gives an example of a typical run-off triangle. In order to simplify our notation, we assume that $I = J$ (run-off triangle). However, all the results we present can be easily extended to the case when the last accident year for which data is available is greater than the last development year, i.e., $I > J$ (run-off trapezoid).

Accident Year i	Development Year j						
	1	2	3	4	j	...	J
1							
2							
3			$C_{i,j}$ (observations)				
$I-j$							
$I-2$						$C_{i,j}$ (to be predicted)	
$I-1$							
I							

Table 3.1: Run-off triangle ($I = J$)

Then the outstanding loss liabilities for accident year $i \in \{1, .., J\}$ at time $t = I$ are given by

$$R_i^I = C_{i,J} - C_{i,I-i+1}. \tag{3.2.1}$$

Let

$$D_I = \{C_{i,j} : i + j \leq I + 1; i \leq I\}, \tag{3.2.2}$$

denote the claims data available at time $t = I$.
Let define, for $1 \leq i \leq I$ and $1 \leq k \leq I$,

$$A_{i,I-i+1} = \{C_{i,j} : 1 \leq j \leq I - i + 1\},$$

and

$$B_k = \{C_{i,j} : k \leq j, i + j \leq I + 1; i \leq I\}.$$

3.3 Model Assumptions

Let define the individual development factors (or link ratios), for $1 \leq i \leq I-1$ and $1 \leq k \leq J - 1$,

$$F_{i,k} := C_{i,k+1}/C_{i,k}. \tag{3.3.1}$$

Suppose that function $g : [0, \infty) \to [0, \infty)$ is Borel measurable. Let $\delta_{i,j}$ be the *non-negative* random variables defined by

$$\delta_{i,j} := g(C_{i,j}). \tag{3.3.2}$$

Our model is formalized by the following assumptions:

(M.1) The accident years $(C_{i,1}, .., C_{i,J})_{1 \leq i \leq I}$ are independent

(M.2) There exist constants $f_k > 0$ such that

$$E(F_{i,k}|C_{i,1}, .., C_{i,k}) = f_k.$$

(M.3) There exist constants $\sigma_k^2 > 0$ such that for all $1 \leq i \leq I$ and $1 \leq k \leq J - 1$ we have

$$Var(F_{i,k}|C_{i,1}, .., C_{i,k}) = \begin{cases} \frac{\sigma_k^2}{\delta_{i,k}} & \text{if} \quad \delta_{i,k} \neq 0 \quad \text{a.s.,} \\ \infty & \text{if} \quad \delta_{i,k} = 0 \quad \text{a.s.,} \end{cases} \tag{3.3.3}$$

where a.s. means *almost surely*.

3.3.1 Model Estimators

Suppose that function $f : [0, \infty) \to [0, \infty)$ is Borel measurable. Let $\gamma_{i,j}$ be the *non-negative* random variables defined by

$$\gamma_{i,j} := f(C_{i,j}). \tag{3.3.4}$$

- Given the information D_I, the factors f_k are estimated by

$$\widehat{f}_k = \frac{\sum_{i=1}^{I-k} \gamma_{i,k} F_{i,k}}{\sum_{i=1}^{I-k} \gamma_{i,k}}, \text{ for } 1 \le k \le J-1. \tag{3.3.5}$$

  It becomes obvious from (M.3) that in order to compute correctly the variance of $\widehat{f}_k$ (see Lemma 3.9.2 in Appendix) we have to assume that

$$\text{if} \quad \delta_{i,j} = 0 \quad \text{then} \quad \gamma_{i,j} = 0. \tag{3.3.6}$$

- The variance parameters σ_k^2 are estimated by

$$\widehat{\sigma}_k^2 = \frac{1}{I-k-1} \sum_{i=}^{I-k} \delta_{i,k} (F_{i,k} - \widehat{f}_k)^2, \text{ for } 1 \le k \le J-2, \tag{3.3.7}$$

$$\widehat{\sigma}_{J-1}^2 = \min(\widehat{\sigma}_{J-2}^4 / \widehat{\sigma}_{J-3}^2, \min(\widehat{\sigma}_{J-3}^2, \widehat{\sigma}_{J-2}^2)). \tag{3.3.8}$$

- The estimator of ultimate claim amount $C_{i,I}$ is given by

$$\widehat{C}_{i,I} = C_{i,I+1-i} \cdot \prod_{k=I+1-i}^{I-1} \widehat{f}_k. \tag{3.3.9}$$

- The estimator of chain ladder reserve R_i^I defined in 3.2.1 is given by

$$\widehat{R}_i^I = \widehat{C}_{i,I} - C_{i,I-i+1}. \tag{3.3.10}$$

Proposition 3.3.1 *(i) The estimators $\widehat{f}_k$ given in (3.3.5) are the unbiased and uncorrelated.*

(ii) *If $\delta_{i,j} = \gamma_{i,j}$ for all i, j, then the estimators $\widehat{f}_k$ of f_k have the minimal variance among all unbiased estimators of f_k which are the weighted average of the observed development factors $F_{i,k}$.*

(iii) *The bias of the estimator $\widehat{\sigma}_k^2$ is given by the following formula*

$$E[\widehat{\sigma}_k^2 - \sigma_k^2] = \frac{\sigma_k^2}{I-k-1} E\left[\frac{\sum_{i=1}^{I-k} \delta_{i,k} \sum_{j=1}^{I-k} \frac{\gamma_{j,k}^2}{\delta_{j,k}}}{\left(\sum_{j=1}^{I-k} \gamma_{j,k}\right)^2} - 1 \right].$$

(iv) *If $\delta_{i,j} = \gamma_{i,j}$ for all i, j, then the estimator $\widehat{\sigma}_k^2$, given in (3.3.7) is the unbiased estimators of the parameter σ_k^2.*

(v) *Under the model assumptions (M.1) and (M.2) we have*

$$E(C_{i,I}|D_I) = C_{i,I+1-i} f_{i,I+1-i} \cdot \ldots \cdot f_{I-1}.$$

This fact and the fact that $\widehat{f}_k$ are uncorrelated implies that $\widehat{C}_{i,I}$ defined in (3.3.9)is unbiased estimator of $E(C_{i,I}|D_I)$.

(vi) *The expected values of the estimator $\widehat{C}_{i,I}$ for the ultimate claim amount defined in (3.3.9) and of the true ultimate claim $C_{i,I}$ are equal, i.e., $E(\widehat{C}_{i,k}) = E(C_{i,I})$, $2 \leq i \leq I$.*

The proof of this proposition is postponed to the appendix.

Remark 3.3.1 *If we set $\gamma_{i,j} = \delta_{i,j} = C_{i,j}$ in (3.3.3) and (3.3.5) we get the assumptions of stochastic Chain Ladder model of Mack (1993)(see also Mack (1994) and Mack (1999)). More examples of the models which are included in our general framework are presented in Section 3.6.*

3.4 (Conditional)Mean-Square Error of Prediction (MSEP)

There are many claims reserving methods available to predict the outstanding loss liabilities, the challenge is to quantify not only the claims reserves but also the uncertainty of the resulting predictors. Here, we quantify the prediction uncertainty with the aid of the most popular such measure, the so-called mean-square error of prediction (MSEP). The conditional mean-square error of prediction of the estimators $\widehat{C}_{i,I}$ and $\sum_{i=1}^{I} \widehat{C}_{i,I}$ are defined by

$$msep_{\widehat{C}_{i,I}|D_I}(C_{i,I}) := E\left[\left(\widehat{C}_{i,I} - C_{i,I}\right)^2 |D_I\right], \tag{3.4.1}$$

$$msep_{\sum_{i=1}^{I}\widehat{C}_{i,I}|D_I}\left(\sum_{i=1}^{I} C_{i,I}\right) := E\left[\left(\sum_{i=1}^{I} \widehat{C}_{i,I} - \sum_{i=1}^{I} C_{i,I}\right)^2 |D_I\right]. \tag{3.4.2}$$

Note that with regards to the conditional MSEP, it does not matter whether one considers the predictor $\widehat{C}_{i,I}$ of the ultimate claim amount or the predictor R_i^I of the claims reserves of accident year i. Both yield the same result. We adopt the convention of using the predictor of the ultimate claim amount. If the predictor $\widehat{C}_{i,I}$ is D_I-measurable, the conditional MSEP decouples as follows:

$$msep_{\widehat{C}_{iI}|D_I}(C_{iI}) = Var(C_{iI}|D_I) + (E(C_{iI}|D_I) - \widehat{C}_{iI})^2,$$

The first term on the right-hand side of the above equation is called the **conditional process variance** It represents the inherent uncertainty of the underlying model chosen for the observed data. The second term on the right-hand side is called the **conditional estimation error**, it represents the uncertainty in the estimation of the unknown model parameters.

3.5 Main Results

The statements in the present section are meant in the following sense. First, we derive the theoretical expressions of *msep* ($msep_{\widehat{C}_{i,I}|D_I}(C_{i,I})$ and $msep_{\sum_{i=1}^{I}\widehat{C}_{i,I}|D_I}\left(\sum_{i=1}^{I} C_{i,I}\right)$ respectively), which depend upon the unknown parameters. This step is legitimate as a mathematical statement. Second, we apply the so called *plug-in principle*, which is often used in statistical inference, and which consists in the replacement of the unknown parameters by their estimates. As a result we obtain plug-in estimators of the above quantities, denoted by $\widehat{msep}$. The presentation of our results follows the common practice and vocabulary of the actuarial literature (see, e.g., Mack (1993)), which may be considered as mathematically ambiguous when the quantities which are so "estimated" are random. In such a setup, one should speak, of "evaluation" rather than "estimation", but both terms are used indifferently in the literature.

Obviously, it is a serious problem to derive, by theory, the limiting properties of plug-in estimators such as $\widehat{msep}$. Aside of the Slutsky lemma, the only applicable methodology in practice makes use of simulation. We leave such investigations to future research on the problem.

The next results show how we quantify the prediction uncertainty by means of the mean-square error of prediction (MSEP) defined in (3.4.1) and (3.4.2).

3.5.1 Single Accident Year

Result 3.5.1 (Conditional MSEP estimator for a single accident year) *Under Model Assumptions (M.1)-(M.3), the conditional mean-square error of ultimate claim $C_{i,I}$, for accident year i, can be estimated by*

$$\widehat{msep}_{\widehat{C}_{i,I}|D_I}(C_{i,I}) = \left(\widehat{C}^I_{i,J}\right)^2 \cdot \left(\widehat{\Gamma}^I_{i,J} + \widehat{\Delta}^I_{i,J}\right), \tag{3.5.1}$$

where

$$\widehat{\Gamma}^I_{i,J} = \sum_{k=I-i+1}^{J-1} \frac{\widehat{\sigma}^2_k / \left(\widehat{f}^I_k\right)^2}{\widehat{\delta}_{i,k}}, \tag{3.5.2}$$

$$\widehat{\Delta}^I_{i,J} = \sum_{k=I-i+1}^{J-1} \frac{\widehat{\sigma}^2_k / \left(\widehat{f}^I_k\right)^2}{\left(\sum_{j=1}^{I-k} \gamma_{j,k}\right)^2} \cdot \sum_{j=1}^{I-k} \frac{(\gamma_{j,k})^2}{\delta_{j,k}} \cdot \mathbf{1}_{\{\delta_{j,k} \neq 0\}}, \tag{3.5.3}$$

and $\widehat{f}_j$ and $\widehat{\sigma}^2_j$ are given in (3.3.5) and (3.3.7)-(3.3.8) respectively.

3.5.2 Aggregation over Prior Accident Year

Result 3.5.2 (Conditional MSEP estimator for aggregated years) *Under Model Assumptions (M.1)-(M.3), the conditional mean-square error of total ultimate claim $\sum_{i=1}^{I} C_{i,I}$, for all accident years, can be estimated by*

$$\widehat{msep}_{\sum_{i=1}^{I}\widehat{C}_{i,I}|D_I}\left(\sum_{i=1}^{I} C_{i,I}\right) = \sum_{i=2}^{I}\widehat{msep}_{\widehat{C}_{iI}|D_I}(C_{iI})$$

$$+\sum_{i=2}^{I}\widehat{C}_{i,I}\left(\sum_{j=i+1}^{I}\widehat{C}_{j,I}\right)\sum_{k=I-i+1}^{J-1} 2\frac{\widehat{\sigma}_k^2/\left(\widehat{f}_k^I\right)^2}{\left(\sum_{l=1}^{I-k}\gamma_{l,k}\right)^2}\cdot\sum_{l=1}^{I-k}\frac{(\gamma_{l,k})^2}{\delta_{l,k}}\cdot\mathbf{1}_{\{\delta_{j,k}\neq 0\}}, \tag{3.5.4}$$

where $\widehat{f}_j$ and $\widehat{\sigma}_j^2$ are given in (3.3.5) and (3.3.7)-(3.3.8), respectively.

Remark 3.5.1 *If we set $\gamma_{i,j} = C_{i,j}$ and $\delta_{i,j} = C_{i,j}$, we obtain from our general framework the main results of Mack (1993) (see also Mack (1994) and Mack (1999)).*

3.6 Numerical example

For our example we use the run-off triangle given in Table B.1 in Appendix B (see also Table 2 in Merz and Wüthrich (2008a)). The dataset contains cumulative payments $C_{i,j}$ for accident years $i \in \{1, 2, ..., 9\}$ and development years $j \in \{1, 2, ..., 9\}$.

As in Mack (1999) we consider here the following family of models, for $1 \leq i \leq I$ and $1 \leq j \leq J$,

$$\gamma_{i,j} := w_{i,j}^{\gamma}\cdot C_{i,j}^{\alpha} \quad \text{and} \quad \delta_{i,j} := w_{i,j}^{\delta}\cdot C_{i,j}^{\beta}, \tag{3.6.1}$$

where $\alpha, \beta \geq 0$ and $w_{i,j}^{\gamma}, w_{i,j}^{\delta} \in [0, 1]$ are arbitrary weights which can be used by the actuary to downweight any outlying $F_{i,j}$ defined via (3.3.1). Hence, according to (3.3.5), the age-to-age factor $\widehat{f}_k$ is defined by

$$\widehat{f}_k = \frac{\sum_{i=1}^{I-k} w_{i,k}^{\gamma} C_{i,k}^{\alpha} F_{i,k}}{\sum_{i=1}^{I-k} w_{i,k}^{\gamma} C_{i,k}^{\alpha}}, \text{ for } 1 \leq k \leq J-1.$$

For the $\gamma_{i,j}$ and $\delta_{i,j}$ defined via (3.6.1) we define 3 methods: A, B and C, by choosing the parameters α, β and the weights $w_{i,j}^{\gamma}$ and $w_{i,j}^{\delta}$ in the following way:

A. $\alpha = \beta = 1$ and $w^{\gamma}_{i,j} = w^{\delta}_{i,j} = 1$ for all i, j (Chain Ladder of Mack (1993)).

B. $\alpha = \beta = 1$ and $w^{\gamma}_{i,j} = w^{\delta}_{i,j} = 0$ for $(i, j) \in W^0$, where W^0 is the nonempty set of indices (i, j).
For the run-off triangle from the Table B.1 the set W^0 of indices (i, j) is

$$W^0 = \{(1, 2); (2, 6); (3, 2)\}.$$

C. $\alpha = \beta = 1$ and $w^{\gamma}_{i,j} = 0$ for $(i, j) \in W^0$, where $W^0 = \{(1, 2); (2, 6); (3, 2)\}$ and $w^{\delta}_{i,j} = 1$ for all (i, j).

Remark 3.6.1 *The parameter α determine the different estimates of f_k*

1. *If $\alpha = 1$ and $w_{i,j} = 1$ for all i, j we get the classical Chain Ladder estimate of f_k*

$$\widehat{f}_k = \frac{\sum_{i=1}^{I-k} C_{i,k} F_{i,k}}{\sum_{i=1}^{I-k} C_{i,k}} = \frac{\sum_{i=1}^{I-k} C_{i,k+1}}{\sum_{i=1}^{I-k} C_{i,k}}, \text{ for } 1 \leq k \leq J-1.$$

2. *If $\alpha = 0$ and $w_{i,j} = 1$ for all i, j we get the model for which the estimators of the age-to-age factors f_k are the straightforward average of the observed individual development factors $F_{i,j}$ defined via (3.3.1), i.e.,*

$$\widehat{f}_k = \frac{1}{I-k} \sum_{i=1}^{I-k} F_{i,k}, \text{ for } 1 \leq k \leq J-1.$$

3. *If $\alpha = 2$ and $w_{i,j} = 1$ for all i, j we get the model for which the estimators of the age-to-age factors f_k are the results of an ordinary regression of $\{C_{i,k+1}\}_{i \in \{1,\ldots,I-k-1\}}$ against $\{C_{i,k}\}_{i \in \{1,\ldots,I-k\}}$ with intercept 0, i.e.,*

$$\widehat{f}_k = \frac{\sum_{i=1}^{I-k} C^2_{i,k} F_{i,k}}{\sum_{i=1}^{I-k} C^2_{i,k}} = \frac{\sum_{i=1}^{I-k} C_{i,k} C_{i,k+1}}{\sum_{i=0}^{I-k-1} C^2_{i,k}}, \text{ for } 0 \leq k \leq J-1.$$

Remark 3.6.2 (Incomplete run-off triangles) *By setting the null value of the weights $w^{\gamma}_{i,j}$ and $w^{\delta}_{i,j}$ we are able to deal with the incomplete run-off triangles. In our general framework the phenomenon of incomplete run-off triangles can be treated by setting $\gamma_{i,j} = 0$ and $\delta_{i,j} = 0$.*

Table 3.4 summarizes the estimates $\widehat{f}_j^I$ of the age-to age factors f_j from the models A, B and C. These estimates are used to compute the claims reserves $\widehat{R}_i^I$ defined in (3.3.10) for the outstanding claims liabilities R_i^I (see Table 3.2). In Table 3.3 we present the estimates (for each model from A to C) of the predictions defined in (3.4.1) and (3.4.2) and given by the formulas (3.5.1) and (3.5.4) respectively .

	$\widehat{R}_I$	
i/method	A	B, C
1	-	-
2	4 378	4 378
3	9 348	9 348
4	28 392	28 392
5	51 444	51 444
6	111 811	111 811
7	187 084	193 414
8	411 864	422 832
9	1 433 505	1 431 306
Total	**2 237 826**	**2 252 925**
	A/A	B,C/A
Total (%)	**100%**	**101%**

Table 3.2: Estimation of reserves

3.7 Conclusions (Case Study)

For the reserves and the uncertainty of reserves (see Table 3.2 and Table 3.3), we obtain the results which are very close. This is mainly due to the fact that the estimated link ratios (see Table 3.4) and their coefficients of variation are very close as well.

	method		
i	A	B	C
1	-	-	-
2	567	567	567
3	1 566	1 566	1 566
4	4 157	4 157	4 157
5	10 536	10 536	10 536
6	30 319	30 319	30 319
7	35 967	32 128	36 977
8	45 090	40 118	46 446
9	69 552	54 136	72 463
Total	**108 401**	**94 600**	**111 965**
	A/A	B/A	C/A
Total (%)	**100%**	**87%**	**103%**

Table 3.3: Estimation of $\widehat{msep}^{1/2}$

$\widehat{f}_i^I$	i							
	1	2	3	4	5	6	7	8
A	1.4759	1.0719	1.0232	1.0161	1.0063	1.0056	1.0013	1.0011
B, C	1.4705	1.0733	1.0249	1.0161	1.0062	1.0055	1.0012	1.0011

Table 3.4: Estimates of the age-to-age factors

3.8 Overall Conclusions

We presented the generalized framework for stochastic Chain Ladder method. We derived the estimators of MSEP of the ultimate amount of claims reserves in the case when we use different weights to estimate the age-to-age factors and the variance parameters. Moreover, the present paper intends to provide a general framework for Chain Ladder (CL) which can be applied in the context of measuring uncertainly of Claims Development Result(CDR) or even in the case of multivariate CL method. These are the objects of our forthcoming paper.

3.9 Mathematical Proofs

We present here the proofs of our main results.

3.9.1 Proof of Result 3.5.1

Due to the general rule $E(X-c)^2 = Var(X) + (EX-c)^2$ for any scalar c we have

$$msep_{\widehat{C}_{iI}|D_I}(C_{iI}) = E\left[\left(\widehat{C}_{iI} - C_{iI}\right)^2 |D_I\right] = Var(C_{iI}|D_I) + (E(C_{iI}|D_I) - \widehat{C}_{iI})^2. \tag{3.9.1}$$

To estimate $Var(C_{i,I}|D_I)$ we use the following

Lemma 3.9.1 *For $i = 2, ..., I$, we have,*

$$Var(C_{i,I}|D_I) = \sum_{l=I+1-i}^{I-1} E\left[\frac{C_{i,l}^2}{\delta_{i,l}}|D_I\right] \sigma_l^2 \prod_{k=l+1}^{I-1} f_k^2.$$

Proof. (Lemma 3.9.1)

For $l = I+1-i, ..., I-1$,

$$\begin{aligned}
Var(C_{i,l+1}|D_I) &= Var(C_{i,l+1}|A_{i,I+1-i}) \\
&= E\left[Var(C_{i,l+1}|A_{i,I-1})|A_{i,I+1-i}\right] + Var\left[E(C_{i,l+1}|A_{i,I-1})|A_{i,I+1-i}\right] \\
&= E\left[\sigma_l^2 \frac{C_{i,l}^2}{\delta_{i,l}}|A_{i,I+1-i}\right] + Var\left[f_l C_{i,l}|A_{i,I+1-i}\right] \\
&= \sigma_l^2 E\left[\frac{C_{i,l}^2}{\delta_{i,l}}|A_{i,I+1-i}\right] + f_l^2 Var\left[C_{i,l}|A_{i,I+1-i}\right] \\
&= \sigma_l^2 E\left[\frac{C_{i,l}^2}{\delta_{i,l}}|D_I\right] + f_l^2 Var\left[C_{i,l}|D_I\right].
\end{aligned} \tag{3.9.2}$$

We multiply the both sides by $\prod_{k=l+1}^{I-1} f_k^2$ and we take the sum over $l = I+1-i, ..., I-1$,

$$\sum_{l=I+1-i}^{I-1} Var(C_{i,l+1}|D_I) \prod_{k=l+1}^{I-1} f_k^2 = \sum_{l=I+1-i}^{I-1} \sigma_l^2 E\left[\frac{C_{i,l}^2}{\delta_{i,l}}|D_I\right] \prod_{k=l+1}^{I-1} f_k^2 + \sum_{l=I+1-i}^{I-1} Var\left[C_{i,l}|D_I\right] f_l^2 \prod_{k=l+1}^{I-1} f_k^2,$$

$$Var(C_{i,I}|D_I) + \sum_{l=I+1-i}^{I-2} Var(C_{i,l+1}|D_I) \prod_{k=l+1}^{I-1} f_k^2 = \sum_{l=I+1-i}^{I-1} \sigma_l^2 E\left[\frac{C_{i,l}^2}{\delta_{i,l}}|D_I\right] \prod_{k=l+1}^{I-1} f_k^2 + Var\left[C_{i,I+1-i}|D_I\right] \prod_{k=I+2-i}^{I-1} f_k^2 + \sum_{l=I+1-i}^{I-1} Var\left[C_{i,l}|D_I\right] \prod_{k=l}^{I-1} f_k^2. \quad (3.9.3)$$

Since $Var\left[C_{i,I+1-i}|D_I\right] = 0$ and from the fact that

$$\sum_{l=I+1-i}^{I-2} Var(C_{i,l+1}|D_I) \prod_{k=l+1}^{I-1} f_k^2 = \sum_{l=I+1-i}^{I-1} Var\left[C_{i,l}|D_I\right] \prod_{k=l}^{I-1} f_k^2,$$

we finally get the proof of Lemma 3.9.1

We estimate $Var(C_{i,l+1}|D_I)$ via Lemma 3.9.1 and by replacing the unknown parameters f_k et σ_k^2 with their estimators $\widehat{f}_k^2$ and $\widehat{\sigma}_k^2$. The quantity $E\left[\frac{C_{i,l}^2}{\delta_{i,l}}|D_I\right]$ is estimated by $\frac{\widehat{C}_{i,l}^2}{\widehat{\delta}_{i,l}}$ (since $E\left[E\left[\frac{C_{i,l}^2}{\delta_{i,l}}|D_I\right]\right] = E\left[\frac{C_{i,l}^2}{\delta_{i,l}}\right]$ and an unbiased estimate of $E\left[\frac{C_{i,l}^2}{\delta_{i,l}}\right]$ is $\frac{C_{i,l}^2}{\delta_{i,l}}$). Furthermore, since $\widehat{C}_{i,l}^2 = C_{I+1-i}^2 \prod_{k=I+1-i}^{k-1} \widehat{f}_k^2$, we obtain:

$$
\begin{aligned}
Var(C_{i,I}|D_I) &= \sum_{l=I+1-i}^{I-1} E\left[\frac{C_{i,l}^2}{\delta_{i,l}}|D_I\right]\sigma_l^2 \prod_{k=l+1}^{I-1} f_k^2 = \sum_{l=I+1-i}^{I-1} \frac{\widehat{C}_{i,l}^2}{\widehat{\delta}_{i,l}}\widehat{\sigma}_l^2 \prod_{k=l+1}^{I-1} \widehat{f}_k^2 \\
&= \sum_{l=I+1-i}^{I-1} \frac{1}{\widehat{\delta}_{i,l}} C_{i,I+1-i}^2 \prod_{k=I+1-i}^{l-1} \widehat{f}_k^2 \widehat{\sigma}_l^2 \prod_{k=l+1}^{I-1} \widehat{f}_k^2 \\
&= C_{i,I+1-i}^2 \sum_{l=I+1-i}^{I-1} \frac{\widehat{\sigma}_l^2/\widehat{f}_k^2}{\widehat{\delta}_{i,l}} \prod_{k=I+1-i}^{I-1} \widehat{f}_k^2 \\
&= C_{i,I+1-i}^2 \cdot \prod_{k=I+1-i}^{I-1} \widehat{f}_k^2 \sum_{l=I+1-i}^{I-1} \frac{\widehat{\sigma}_l^2/\widehat{f}_k^2}{\widehat{\delta}_{i,l}} = \widehat{C}_{i,I}^2 \sum_{l=I+1-i}^{I-1} \frac{\widehat{\sigma}_l^2/\widehat{f}_k^2}{\widehat{\delta}_{i,l}}.
\end{aligned}
\tag{3.9.4}
$$

We now turn to the second summand of the expression (3.9.1). Because of Proposition 3.3.1(v) and (vi) we have,

$$
(E(C_{i,I}|D_I) - \widehat{C}_{i,I})^2 = C_{i,I+1-i}^2 \left(f_{I+1-i} \cdot \ldots \cdot f_{I-1} - \widehat{f}_{I+1-i} \cdot \ldots \cdot \widehat{f}_{I+1-i}\right)^2. \tag{3.9.5}
$$

This expression cannot be estimated by replacing f_k with $\widehat{f}_k$. In order to estimate the right hand side of (3.9.5) we use the same approach as in Mack (1999). We define,

$$
\begin{aligned}
F &= f_{I+1-i} \cdot \ldots \cdot f_{I-1} - \widehat{f}_{I+1-i} \cdot \ldots \cdot \widehat{f}_{I+1-i} \\
&= S_{I+1-i} + \ldots + S_{I-1},
\end{aligned}
\tag{3.9.6}
$$

with

$$
\begin{aligned}
S_k &= \widehat{f}_{I+1-i} \cdot \ldots \cdot \widehat{f}_{k-1} f_k f_{k+1} \cdot \ldots \cdot f_{I-1} \\
&\quad - \widehat{f}_{I+1-i} \cdot \ldots \cdot \widehat{f}_{k-1} \widehat{f}_k f_{k+1} \cdot \ldots \cdot f_{I-1} \\
&= \widehat{f}_{I+1-i} \cdot \ldots \cdot \widehat{f}_{k-1} (f_k - \widehat{f}_k) f_{k+1} \cdot \ldots \cdot f_{I-1}.
\end{aligned}
\tag{3.9.7}
$$

This yields

$$F^2 = (S_{I+1-i} + \ldots + S_{I-1})^2 = \sum_{k=I+1-i}^{I-1} S_k^2 + 2 \sum_{k=I+1-i}^{I-1} \sum_{j<k}^{I-1} S_j S_k. \tag{3.9.8}$$

Following Mack (1993, 1994) we can estimate F^2 using following Proposition

Proposition 3.9.1 *We have*

$$\widehat{F^2} = \prod_{l=I+1-i}^{I-1} \widehat{f}_l^2 \sum_{k=I+1-i}^{I-1} \frac{Var(\widehat{f}_k|B_k)}{\widehat{f}_k^2}.$$

The prove this proposition we imitate the proof of the same result stated in Mack (1993, 1994).

Proof. (Proposition 3.9.1)

We replace S_k^2 with $E(S_k^2|B_k)$ and S_jS_k, with $E(S_jS_k|B_k)$. This means that we approximate S_k^2 and S_jS_k by varying and averaging as little data as possible so that as many values $C_{i,k}$ from data observed are kept fixed. Due to Proposition 3.3.1 (i) we have $E(\widehat{f}_k - f_k) = 0$ and therefore $E(S_jS_k|B_k) = 0$ for $j < k$ because all f_r , $r < k$, are scalars under B_k.

Since $E((f_k - \widehat{f}_k)^2|B_k) = Var(\widehat{f}_k|B_k)$ we obtain from (3.9.7)

$$E(S_k^2|B_k) = \widehat{f}_{I+1-i}^2 \cdot \ldots \cdot \widehat{f}_{k-1}^2 Var(\widehat{f}_k|B_k) f_{k+1}^2 \cdot \ldots \cdot f_{I-1}^2.$$

Taken together, we have replaced $F^2 = \sum_{k=I+1-i}^{I-1} S_k^2$ with $\sum_{k=I+1-i}^{I-1} E(S_k^2|B_k)$ and the unknown parameters are replaced by their estimators. Altogether, we estimate F^2 by

$$\sum_{k=I+1-i}^{I-1} \widehat{f}_{I+1-i}^2 \cdot \ldots \cdot \widehat{f}_{k-1}^2 \widehat{f}_k^2 \widehat{f}_{k+1}^2 \cdot \ldots \cdot \widehat{f}_{I-1}^2 \frac{Var(\widehat{f}_k|B_k)}{\widehat{f}_k^2}.$$

□

It remains to determine the estimate of $Var(\widehat{f}_k|B_k)$. We use the following

Lemma 3.9.2 *We assume (3.3.6). We have*

$$Var(\widehat{f}_k|B_k) = \sigma_k^2 \frac{\sum_{j=1}^{I-k} \frac{\gamma_{j,k}^2}{\delta_{j,k}} \cdot \mathbf{1}_{\{\delta_{j,k} \neq 0\}}}{\left(\sum_{j=1}^{I-k} \gamma_{j,k}\right)^2}.$$

Proof. (Lemma 3.9.2)

$$Var(\widehat{f}_k|B_k) = Var\left(\frac{\sum_{j=1}^{I-k}\gamma_{j,k}F_{j,k}}{\sum_{j=1}^{I-k}\gamma_{j,k}}|B_k\right) = \frac{\sum_{j=1}^{I-k}\gamma_{j,k}^2 Var\left(F_{i,k}|B_k\right)\cdot \mathbf{1}_{\{\delta_{j,k}\neq 0\}}}{\left(\sum_{j=1}^{I-k}\gamma_{j,k}\right)^2}, \tag{3.9.9}$$

where the second equality is due to assumption (3.3.6) and due to the convention that the product of 0 and ∞ equals to 0. □

Using (3.9.5) and Lemma 3.9.2 we estimate $E(C_{i,I}|D_I) - \widehat{C}_{i,I})^2$ by

$$\begin{aligned} &C_{i,I+1-i}^2\widehat{f}_{I+1-i}^2\cdot\ldots\cdot\widehat{f}_{I-1}^2\sum_{k=I+1-i}^{I-1}\frac{\widehat{\sigma}_k^2}{\widehat{f}_k^2}\frac{\sum_{j=1}^{I-k}\frac{\gamma_{j,k}^2}{\delta_{j,k}}\cdot\mathbf{1}_{\{\delta_{j,k}\neq 0\}}}{\left(\sum_{j=1}^{I-k}\gamma_{j,k}\right)^2} \\ &= \widehat{C}_{i,I}^2\sum_{k=I+1-i}^{I-1}\frac{\widehat{\sigma}_k^2}{\widehat{f}_k^2}\frac{\sum_{j=1}^{I-k}\frac{\gamma_{j,k}^2}{\delta_{j,k}}\cdot\mathbf{1}_{\{\delta_{j,k}\neq 0\}}}{\left(\sum_{j=1}^{I-k}\gamma_{j,k}\right)^2}. \end{aligned} \tag{3.9.10}$$

3.9.2 Proof of Result 3.5.2 (Overall standard error)

$$\widehat{msep}_{\sum_{i=1}^{I}\widehat{C}_{iI}|D_I}\left(\sum_{i=1}^{I}C_{i,I}\right) = \sum_{i=2}^{I}\widehat{msep}_{\widehat{C}_{iI}|D_I}(C_{iI}) + \sum_{2\leq i<j\leq I} 2\cdot C_{i,I+1-i}C_{j,I+1-j}F_iF_j,$$

with

$$F_i = f_{I+1-i}\cdot\ldots\cdot f_{I-1} - \widehat{f}_{I+1-i}\cdot\ldots\cdot\widehat{f}_{I+1-i},$$

$$F_i = \sum_{k=I+1-i}^{I-1} S_k^i,$$

where

$$S_k^i = \widehat{f}_{I+1-i}\cdot\ldots\cdot\widehat{f}_{k-1}(f_k - \widehat{f}_k)f_{k+1}\cdot\ldots\cdot f_{I-1}.$$

We can determine the estimator of F_iF_j in the analogous way as for F^2.

Proposition 3.9.2 *We have*

$$\widehat{F_iF_j} = \sum_{k=I+1-i}^{I-1} \frac{Var(\widehat{f}_k|B_k)}{\widehat{f}_k^2} \left(\widehat{f}_{I+1-j} \cdot \ldots \cdot \widehat{f}_{I-1}\right) \cdot \left(\widehat{f}_{I+1-i} \cdot \ldots \cdot \widehat{f}_{I-1}\right).$$

Proof. (Proposition 3.9.2)

$$E[(S_k^i)^2|B_k] = \widehat{f}_{I+1-i}^2 \cdot \ldots \cdot \widehat{f}_{k-1}^2 Var(\widehat{f}_k|B_k) f_{k+1}^2 \cdot \ldots \cdot f_{I-1}^2.$$

For $i < j$, we have

$$\begin{aligned}
\widehat{F_iF_j} &= \sum_{k=I+1-i}^{I-1} \widehat{f}_{I+1-j} \cdot \ldots \cdot \widehat{f}_{I-i} \cdot \widehat{f}_{I+1-i}^2 \cdot \ldots \cdot \widehat{f}_{k-1}^2 Var(\widehat{f}_k|B_k) \widehat{f}_{k+1}^2 \cdot \ldots \cdot \widehat{f}_{I-1}^2 \\
&= \sum_{k=I+1-i}^{I-1} \frac{Var(\widehat{f}_k|B_k)}{\widehat{f}_k^2} \widehat{f}_{I+1-j} \cdot \ldots \cdot \widehat{f}_{I-i} \cdot \widehat{f}_{I+1-i}^2 \cdot \ldots \cdot \widehat{f}_{k-1}^2 \cdot \widehat{f}_k^2 \cdot \widehat{f}_{k+1}^2 \cdot \ldots \cdot \widehat{f}_{I-1}^2 \\
&= \sum_{k=I+1-i}^{I-1} \frac{Var(\widehat{f}_k|B_k)}{\widehat{f}_k^2} \left(\widehat{f}_{I+1-j} \cdot \ldots \cdot \widehat{f}_{I-1}\right) \cdot \left(\widehat{f}_{I+1-i} \cdot \ldots \cdot \widehat{f}_{I-1}\right).
\end{aligned} \tag{3.9.11}$$

□

Finally, from Lemma 3.9.2

$$\begin{aligned}
&\sum_{2 \le i < j \le I}^{I} 2 \cdot C_{i,I+1-i} C_{j,I+1-j} \widehat{F_iF_j} \\
&= \sum_{i=2}^{I} \widehat{C}_{i,I} \sum_{j=i+1}^{I} \widehat{C}_{j,I} \sum_{k=I-i+1}^{I-1} 2 \frac{\widehat{\sigma}_k^2 / \widehat{f}_k^2}{\left(\sum_{l=1}^{I-k} \gamma_{l,k}\right)^2} \cdot \sum_{l=1}^{I-k} \frac{\gamma_{l,k}^2 \cdot \mathbf{1}_{\{\delta_{j,k} \neq 0\}}}{\delta_{l,k}}
\end{aligned} \tag{3.9.12}$$

□

3.9.3 Proof of Proposition 3.3.1

(i) See Theorem 2 p. 215 in Mack (1993).

(ii) See discussion on p.112, Corollary on p.141 and Appendix B on p.140 in Mack (1994).

(iii) We have, for $1 \leq k \leq J-2$,

$$(I-k-1)\cdot\widehat{\sigma}_k^2 = \sum_{i=1}^{I-k} \delta_{i,k}(F_{i,k}-\widehat{f}_k)^2 = \sum_{i=1}^{I-k} \delta_{i,k}F_{i,k}^2 - 2\sum_{i=1}^{I-k} \delta_{i,k}F_{i,k}\cdot\widehat{f}_k + \sum_{i=1}^{I-k} \delta_{i,k}\widehat{f}_k^{\,2}.$$

Since $\gamma_{i,j}$ and $\delta_{i,j}$ are $\sigma(C_{i,j})$ measurable ($\sigma(C_{i,j})$ denotes the σ- field generated by $C_{i,j}$), we have

$$E((I-k-1)\cdot\widehat{\sigma}_k^2|B_k) = \sum_{i=1}^{I-k} \delta_{i,k}E(F_{i,k}^2|B_k) - 2\sum_{i=1}^{I-k} \delta_{i,k}E(F_{i,k}\cdot\widehat{f}_k|B_k) + \sum_{i=1}^{I-k} \delta_{i,k}E(\widehat{f}_k^{\,2}|B_k).$$

Since $F_{i,k}$ and $F_{j,k}$ are independent for $i \neq j$, and $E(F_{i,k}^2|B_k) = \frac{\sigma_k^2}{\delta_{i,k}} + f_k^2$, we have

$$\begin{aligned}
E(F_{i,k}\cdot\widehat{f}_k|B_k) &= \frac{1}{\sum_{i=1}^{I-k}\gamma_{i,k}}\left(\sum_{j=1}^{I-k}\gamma_{j,k}\cdot E(F_{i,k}\cdot F_{j,k}|B_k)\right)\\
&= \frac{1}{\sum_{i=1}^{I-k}\gamma_{i,k}}\left(\gamma_{i,k}\cdot E(F_{i,k}^2|B_k) + \sum_{j\neq i}^{I-k}\gamma_{j,k}\cdot E(F_{i,k}|B_k)\cdot E(F_{j,k}|B_k)\right)\\
&= \frac{1}{\sum_{i=1}^{I-k}\gamma_{i,k}}\left(\gamma_{i,k}\cdot E(F_{i,k}^2|B_k) + \sum_{j\neq i}^{I-k}\gamma_{j,k}\cdot E(F_{i,k}|B_k)\cdot E(F_{j,k}|B_k)\right)\\
&= \frac{1}{\sum_{i=1}^{I-k}\gamma_{i,k}}\left(\gamma_{i,k}\cdot(\frac{\sigma_k^2}{\delta_{i,k}} + f_k^2) + \sum_{j\neq i}^{I-k}\gamma_{j,k}f_k^2\right)\\
&= \frac{1}{\sum_{i=1}^{I-k}\gamma_{i,k}}\left(\gamma_{i,k}\cdot\frac{\sigma_k^2}{\delta_{i,k}} + f_k^2\sum_{j=1}^{I-k}\gamma_{j,k}\right)\\
&= \sigma_k^2\frac{\frac{\gamma_{i,k}}{\delta_{i,k}}}{\sum_{i=1}^{I-k}\gamma_{i,k}} + f_k^2.
\end{aligned} \tag{3.9.13}$$

From Lemma 3.9.2

$$E(\widehat{f}_k^{\,2}|B_k) = Var(\widehat{f}_k|B_k) + (E(\widehat{f}_k|B_k))^2 = \sigma_k^2 \frac{\sum_{j=1}^{I-k} \frac{\gamma_{j,k}^2}{\delta_{j,k}}}{\left(\sum_{j=1}^{I-k} \gamma_{j,k}\right)^2} + f_k^2.$$

Taking together we obtain

$$\begin{aligned} E((I-k-1)\cdot\widehat{\sigma}_k^2|B_k) &= \sum_{i=1}^{I-k} \delta_{i,k} E(F_{i,k}^2|B_k) - 2\sum_{i=1}^{I-k} \delta_{i,k} E(F_{i,k}\cdot\widehat{f}_k|B_k) \\ &+ \sum_{i=1}^{I-k} \delta_{i,k} E(\widehat{f}_k^{\,2}|B_k) \\ &= \sum_{i=1}^{I-k} \delta_{i,k}(\frac{\sigma_k^2}{\delta_{i,k}} + f_k^2) - 2\sum_{i=1}^{I-k} \delta_{i,k}(\sigma_k^2 \frac{\frac{\gamma_{i,k}}{\delta_{i,k}}}{\sum_{i=1}^{I-k}\gamma_{i,k}} + f_k^2) \\ &+ \sum_{i=1}^{I-k} \delta_{i,k}(\sigma_k^2 \frac{\sum_{j=1}^{I-k} \frac{\gamma_{j,k}^2}{\delta_{j,k}}}{\left(\sum_{j=1}^{I-k} \gamma_{j,k}\right)^2} + f_k^2) \\ &= (I-k)\sigma_k^2 + f_k^2 \sum_{i=1}^{I-k} \delta_{i,k} - 2\sigma_k^2 - 2f_k^2 \sum_{i=1}^{I-k} \delta_{i,k} \\ &+ \sigma_k^2 \frac{\sum_{i=1}^{I-k} \delta_{i,k} \sum_{j=1}^{I-k} \frac{\gamma_{j,k}^2}{\delta_{j,k}}}{\left(\sum_{j=1}^{I-k} \gamma_{j,k}\right)^2} + f_k^2 \sum_{i=1}^{I-k} \delta_{i,k} \\ &= (I-k-1)\sigma_k^2 + \sigma_k^2 \left[\frac{\sum_{i=1}^{I-k} \delta_{i,k} \sum_{j=1}^{I-k} \frac{\gamma_{j,k}^2}{\delta_{j,k}}}{\left(\sum_{j=1}^{I-k} \gamma_{j,k}\right)^2} - 1 \right]. \end{aligned} \tag{3.9.14}$$

Finally

$$E(\widehat{\sigma}_k^2 - \sigma_k^2) = E[E[(\widehat{\sigma}_k^2 - \sigma_k^2)|B_k]] = \frac{\sigma_k^2}{I-k-1} E\left[\frac{\sum_{i=1}^{I-k} \delta_{i,k} \sum_{j=1}^{I-k} \frac{\gamma_{j,k}^2}{\delta_{j,k}}}{\left(\sum_{j=1}^{I-k} \gamma_{j,k}\right)^2} - 1 \right].$$

(iv) It is straightforward from (iii).

(v) See Theorem 1 p. 215 in Mack (1993).

(vi) see Appendix C p.142 in Mack (1994):

Chapter 4

Paper III - Generalized Framework for Measuring the Uncertainty of the Claims Development Result

4.1 Introduction

In Merz and Wüthrich (2008a) (MW), the authors defined the *claims development result* (CDR) at time $I+1$ for accounting year $(I, I+1]$ as the difference between two successive predictions of the total ultimate claim. The first prediction is evaluated at time I (with the available information up to time I), and the second one is made one period later at time $I+1$ (with the updated information available at time $I+1$). In their paper, MW based their study of the prediction of CDR, and of the possible fluctuations around this prediction (prediction uncertainty) on a distribution-free Chain Ladder method.

In our work, we extend the model of MW to a more general class of models based on age-to-age factors (link ratios). The Chain Ladder Model of MW turns out to be a particular case of our general model.

In the final section, we present a case study containing different applications of our general tool for reserving. We apply our theoretical results to solve many of the practical problems of actuaries. One of them is to compute the MSEP of the claims reserves in the case when the actuary wants to use different weights for individuals age-to-age factors (link ratios) for estimation of the chain ladder factors and the variance parameters. We will make an instrumental use of the notation and methods of MW and we will follow the arguments of proofs of their main results.

Organization of the paper. In Section 4.2 we present our notation and we introduce the Chain Ladder Time Series Model. In the same section we derive the structure of the mean and the variance of the claims process and defines estimators of the unknown parameters. The definition of the one-year CDR is given in Section 4.3. In Section 4.4 and 4.5 we derive estimators of the conditional MSEP of the one-year CDR for single and for aggregated accident years. Finally, a numerical example is presented in Section 4.6. The run-off triangles used in this case study are presented in Appendix B. All proofs are provided in Section 4.7.

4.2 Notation

Change in notation.

There is a slight change in notation compared to Chapter 3, namely the indices corresponding to accident year i and development year j start from 0 instead 1. The reason is that we wanted to keep the notation consistent with the study of Merz and Wüthrich (2008a) on which we based our present work.

Let $C_{i,j}$ denote the cumulative payments for accident year $i \in \{0, ..., I\}$ until development year $j \in \{0, ..., J\}$, where the accident year is referred to as the year in which an event triggering insurance claims occurs. This means that the ultimate claim for accident year i is given by $C_{i,J}$. We assume that the cumulative payments $C_{i,j}$ are random variables observable for calendar years $i + j \leq I$ and non-observable (to be predicted) for calendar years $i + j > I$. The observable cumulative payments are represented by the so-called run-off trapezoids($I > J$) or run-off triangles ($I = J$). Table 4.1 gives an example of a typical run-off triangle. In order to simplify our notation, we assume that $I = J$ (run-off triangle). However, all the results we present can be easily extended to the case when the last accident year for which data is available is greater than the last development year, i.e., $I > J$ (run-off trapezoid).

Accident Year i	Development Year j: 0	1	2	3	j	...	J
0							
1							
2				$C_{i,j}$ (observations)			
$I-j$							
$I-2$						$C_{i,j}$ (to be predicted)	
$I-1$							
I							

Table 4.1: Run-off triangle ($I = J$)

Then the outstanding loss liabilities for accident year $i \in \{0, .., J\}$ at time $t = I$ are given by

$$R_i^I = C_{i,J} - C_{i,I-i} \tag{4.2.1}$$

and at time $t = I + 1$ they are given by

$$R_i^{I+1} = C_{i,J} - C_{i,I-i+1} \tag{4.2.2}$$

Let

$$D_I = \{C_{i,j} : i + j \leq I; i \leq I\}, \tag{4.2.3}$$

denote the claims data available at time $t = I$ and

$$D_{I+1} = \{C_{i,j} : i + j \leq I + 1; i \leq I\}, \tag{4.2.4}$$

denote the claims data available at time $t = I + 1$.

4.2.1 Model Assumptions of Chain Ladder Time Series Model

The time series model defines an auto-regressive process. This model enables the evaluation of the volatility of one-year reserve risk. It is particularly useful for the derivation of the estimation error (see (4.4.3)). We introduce this approach because the model from Chapter 2 defined by assumptions (M.1)-(M.3) seems to be to general to evaluate the estimation error in the context of one-year insurance reserve risk.

Suppose that function $g : [0, \infty) \to [0, \infty)$ is Borel measurable. Let $\delta_{i,j}$ be the *non-negative* random variables defined by

$$\delta_{i,j} := g(C_{i,j}).$$

Our model is formalized by the following assumptions:

(TM.1) There exist constants $f_k > 0$, $\sigma_k > 0$, $(k = 0, ..., J - 1)$ such that for all $1 \leq j \leq J$ and $0 \leq i \leq I$ we have

$$C_{i,j} = \begin{cases} f_{j-1} \cdot C_{i,j-1} + \frac{\sigma_{j-1}}{\sqrt{\delta_{i,j-1}}} \cdot C_{i,j-1} \cdot \varepsilon_{i,j} & \text{if} \quad \delta_{i,j-1} \neq 0 \quad \text{a.s.,} \\ f_{j-1} \cdot C_{i,j-1} + \overline{\varepsilon}_{i,j} & \text{if} \quad \delta_{i,j-1} = 0 \quad \text{a.s.} \end{cases} \tag{4.2.5}$$

where a.s. states for almost surely and

- $E(\varepsilon_{i,j}|C_{i,0}) = 0$, $E(\varepsilon_{i,j}^2|C_{i,0}) = 1$,
- $E(\overline{\varepsilon}_{i,j}|C_{i,0}) = 0$, $E(\overline{\varepsilon}_{i,j}^2|C_{i,0}) = \infty$,

(TM.2) $\varepsilon_{i,j}$ and $\overline{\varepsilon}_{i,j}$, given $C_{i,0}$, are independent random variables, (TM.3) $P(C_{i,j} > 0|C_{i,0}) = 1$ for all $j \in \{1, .., J\}$ and $i \in \{0, .., I\}$,

Let define the individual development factors (individual age-to-age factors or link ratios), for $0 \le i \le I - 1$ and $0 \le k \le J - 1$,

$$F_{i,k} := C_{i,k+1}/C_{i,k}. \tag{4.2.6}$$

Remark 4.2.1 *Chain Ladder Time Series Model defined via assumptions (TM.1)-(TM.3) fulfils the following hypothesis:*

(M.1) There exist constants $f_k > 0$ and $\sigma_k^2 > 0$ such that for all $0 \le i \le I$ and $0 \le k \le J - 1$ we have

$$E(F_{i,k}|C_{i,0}, .., C_{i,k}) = E(F_{i,k}|C_{i,k}) = f_k,$$
$$Var(F_{i,k}|C_{i,0}, .., C_{i,k}) = Var(F_{i,k}|C_{i,k}) = \begin{cases} \frac{\sigma_k^2}{\delta_{i,k}} & if \quad \delta_{i,k} \neq 0 \quad a.s., \\ \infty & if \quad \delta_{i,k} = 0 \quad a.s. \end{cases} \tag{4.2.7}$$

(M.2) The accident years $(C_{i,0}, .., C_{i,J})_{0 \le i \le I}$ are independent

Remark 4.2.2 *If we set $\delta_{i,j} = C_{i,j}$ in (4.2.5) we obtain the Markov chain formulation of stochastic Chain Ladder model defined in (4.2.7) (see also Buchwalder et al (2006)). Mack assumptions are slightly weaker than the Markov chain assumptions (see Mack (1993), Mack (1994) and Mack (1999)). More examples of the models which are included in our general framework are presented in Section 4.6.*

Remark 4.2.3 *[**Limits and critics of Chain Ladder Time Series Model**]*
In Mack et al (2006), the authors presented the strongly critical opinion about the model defined via the assumptions (TM.1)-(TM.3). They mostly pointed two weak points of this model. First negative comment concerns the contradiction between the positivity of the variables $C_{i,j}$ and the independence of the variables $\varepsilon_{i,j}$ They stated that it seems to be almost impossible to define the error term $\varepsilon_{i,j}$ in such a way that the assumptions (TM.2) and (TM.3) are fulfilled (see the comment in Mack et al (2006), p. 544).
The second complaint concerns the approach applied to determine the estimation error. This approach is based on the resampling technique of the

development factors $\widehat{f}_k$. As a result, we generate the new factors $\widehat{\widehat{f}}_k$ which are independent by construction (see proof of Lemma 4.7.3). It implies that the square of the new factors $\widehat{\widehat{f}}_k$ are also independent which disregard the fact that the original development factors are conditionally negatively correlated (see Mack et al (2006) p.546-547).

Remark 4.2.4 *[**Solution for limits of Chain Ladder Time Series Model**]*
The partial remedy of the criticism of Mack et al (2006) would be to consider the Chain Ladder method in the framework of Bayesian model. In fact it is possible to generalize the assumption B2 of the Bayesian Chain Ladder in Bühlmann et al (2009) (see p.276) and to obtain (by linearization) the same estimators of (4.4.1)- (4.4.2) as obtained in the present paper (see Result 4.9, Result 4.10 and Remark 4.11 in Bühlmann et al (2009), p. 297-298). This is mentioned by the authors after the statement of Result 4.9: "One obtains equality when one linearizes the Result 4.9".

4.2.2 Model Estimators

Suppose that function $f : [0, \infty) \to [0, \infty)$ is Borel measurable. Let $\gamma_{i,j}$ be the *non-negative* random variables defined by

$$\gamma_{i,j} := f(C_{i,j}).$$

Our model is formalized by the following assumptions:

- At time $t = I$, given the information D_I, the factors f_k are estimated by
$$\widehat{f}_k^I = \frac{\sum_{i=0}^{I-k-1} \gamma_{i,k} F_{i,k}}{\sum_{i=0}^{I-k-1} \gamma_{i,k}}, \text{ for } 0 \le k \le J-1, \tag{4.2.8}$$

- At time $t = I + 1$, given the information D_{I+1}, the factors f_k are estimated by,
$$\widehat{f}_k^{I+1} = \frac{\sum_{i=0}^{I-k} \gamma_{i,k} F_{i,k}}{\sum_{i=0}^{I-k} \gamma_{i,k}}, \text{ for } 0 \le k \le J-1, \tag{4.2.9}$$

- The variance parameters σ_k^2 are estimated by

$$\widehat{\sigma}_k^2 = \frac{1}{I-k-1} \sum_{i=0}^{I-k-1} \delta_{i,k}(F_{i,k} - \widehat{f}_k)^2, \text{ for } 0 \le k \le J-2, \qquad (4.2.10)$$

$$\widehat{\sigma}_{J-1}^2 = \min(\widehat{\sigma}_{J-2}^4 / \widehat{\sigma}_{J-3}^2, \min(\widehat{\sigma}_{J-3}^2, \widehat{\sigma}_{J-2}^2)). \qquad (4.2.11)$$

It becomes obvious from (4.2.7) that in order to compute correctly the variance of $\widehat{f}_k$ (see proof of Lemma 4.7.3 (d) in Section 4.7) we have to assume that

$$\text{if } \ \delta_{i,j} = 0 \ \text{ then } \ \gamma_{i,j} = 0. \qquad (4.2.12)$$

In the following Proposition we recall the basic properties of the estimators $\widehat{f}_k$ and $\widehat{\sigma}_k^2$.

Proposition 4.2.1 *(i) The estimators $\widehat{f}_k$ given in (4.2.8) are the unbiased and uncorrelated.*

(ii) If $\delta_{i,j} = \gamma_{i,j}$ for all i, j, then the estimators $\widehat{f}_k$ of f_k have the minimal variance among all unbiased estimators of f_k which are the weighted average of the observed development factors $F_{i,k}$.

(iii) The bias of the estimator $\widehat{\sigma}_k^2$ is given by the following formula

$$E[\widehat{\sigma}_k^2 - \sigma_k^2] = \frac{\sigma_k^2}{I-k-1} E\left[\frac{\sum_{i=0}^{I-k-1} \delta_{i,k} \sum_{j=0}^{I-k-1} \frac{\gamma_{j,k}^2}{\delta_{j,k}}}{\left(\sum_{j=0}^{I-k-1} \gamma_{j,k}\right)^2} - 1 \right].$$

(iv) If $\delta_{i,j} = \gamma_{i,j}$ for all i, j, then the estimator $\widehat{\sigma}_k^2$, given in (4.2.10) is the unbiased estimators of the parameter σ_k^2.

(v) Under the model assumptions (TM.1) and (TM.2) we have

$$E(C_{i,I}|D_I) = C_{i,I+1-i} f_{i,I+1-i} \cdot \ldots \cdot f_{I-1}.$$

This fact and the fact that $\widehat{f}_k$ are uncorrelated implies that $\widehat{C}_{i,I}$ is unbiased estimator of $E(C_{i,I}|D_I)$.

(vi) The expected values of the estimator

$$\widehat{C}_{i,I} = C_{i,I+1-i} \cdot \prod_{k=I+1-i}^{I-1} \widehat{f}_k,$$

for the ultimate claims amount and of the true ultimate claims amount $C_{i,I}$ *are equal, i.e.,* $E(\widehat{C}_{i,k}) = E(C_{i,I})$, $1 \le i \le I$.

The proof of this proposition is postponed to Section 4.7.

4.3 Claims Development Result (CDR)

Definition 4.3.1 (True claims development result)

$$\begin{aligned} CDR_i(I+1) &= E[R_i^I|D_I] - (X_{i,I-i+1} + E[R_i^{I+1}|D_{I+1}]) \\ &= E[C_{i,J}|D_I] - E[C_{i,J}|D_{I+1}], \end{aligned} \tag{4.3.1}$$

where $X_{i,I-i+1} = C_{i,I-i+1} - C_{i,I-i}$ *denotes the incremental payments. Furthermore, the true aggregate CDR is given by*

$$CDR(I+1) = \sum_{i=1}^{I} CDR_i(I+1). \tag{4.3.2}$$

Remark 4.3.1 *Using the martingale property we have*

$$E[CDR_i(I+1)|D_I] = 0. \tag{4.3.3}$$

Definition 4.3.2 (Observable claims development result)

$$\widehat{CDR}_i(I+1) = \widehat{R}_i^I - (X_{i,I-i+1} + \widehat{R}_i^{I+1}) = \widehat{C}_{i,J}^I - \widehat{C}_{i,J}^{I+1}, \tag{4.3.4}$$

where, for $1 \le i \le I$,

$$\widehat{C}_{i,J}^I = C_{i,I-i} \prod_{j=I-i}^{J-1} \widehat{f}_j^I, \tag{4.3.5}$$

$$\widehat{C}_{i,J}^{I+1} = C_{i,I-i+1} \prod_{j=I-i+1}^{J-1} \widehat{f}_j^{I+1}, \tag{4.3.6}$$

$$\widehat{R}_i^I = \widehat{C}_{i,J}^I - C_{i,I-i}, \tag{4.3.7}$$

$$\widehat{R}_i^{I+1} = \widehat{C}_{i,J}^{I+1} - C_{i,I-i+1}. \tag{4.3.8}$$

Furthermore, the observable aggregate CDR is given by

$$\widehat{CDR}(I+1) = \sum_{i=1}^{I} \widehat{CDR}_i(I+1). \tag{4.3.9}$$

4.4 (Conditional)Mean-Square Error of Prediction (MSEP) of the claims development result

The conditional MSEP considered here gives the prospective solvency point of view. It quantifies the prediction uncertainty in the budget value 0 for the observable claims development result at the end of the accounting period. In the solvency margin we need to hold risk capital for possible negative deviations of $CDR_i(I+1)$ from 0. We are interested here in quantifying the following two quantities

$$msep_{\widehat{CDR}_i(I+1)|D_I}(0) = E\left[\left(\widehat{CDR}_i(I+1) - 0\right)^2 |D_I\right], \tag{4.4.1}$$

$$msep_{\sum_{i=1}^I \widehat{CDR}_i(I+1)|D_I}(0) = E\left[\left(\sum_{i=1}^{I} \widehat{CDR}_i(I+1) - 0\right)^2 |D_I\right]. \tag{4.4.2}$$

Similarly to the mean-square error of prediction of ultimate claim, the conditional MSEP of one-year claims development result decouples as follows:

$$msep_{\widehat{CDR}_i(I+1)|D_I}(0) = Var\left(\widehat{CDR}_i(I+1)|D_I\right) + \left[E\left(\widehat{CDR}_i(I+1)|D_I\right)\right]^2. \tag{4.4.3}$$

The first term on the right-hand side of the above equation is called the **conditional process variance** It represents the inherent uncertainty of the underlying model chosen for the observed data. The second term on the right-hand side is called the **conditional estimation error**, it represents the uncertainty in the estimation of the unknown model parameters.

4.5 Main Results

The statements in the present section are meant in the following sense. First, we derive the theoretical expressions of *msep* ($msep_{\widehat{CDR}_i(I+1)|D_I}(0)$ and $msep_{\sum_{i=1}^{I}\widehat{CDR}_i(I+1)|D_I}(0)$ respectively), which depend upon the unknown parameters. This step is legitimate as a mathematical statement. Second, we apply the so called *plug-in principle*, which is often used in statistical inference, and which consists in the replacement of the unknown parameters by their estimates. As a result we obtain plug-in estimators of the above quantities, denoted by $\widehat{msep}$. The presentation of our results follows the common practice and vocabulary of the actuarial literature (see, e.g., Merz and Wüthrich (2008a)), which may be considered as mathematically ambiguous when the quantities which are so "estimated" are random. In such a setup, one should speak, of "evaluation" rather than "estimation", but both terms are used indifferently in the literature.

Obviously, it is a serious problem to derive, by theory, the limiting properties of plug-in estimators such as $\widehat{msep}$. Aside of the Slutsky lemma, the only applicable methodology in practice makes use of simulation. We leave such investigations to future research on the problem.

The next results show how one can quantify the prediction uncertainty by means of the mean-square error of prediction (MSEP) defined in (4.4.1) and (4.4.2).

4.5.1 Single Accident Year - Exact Formula

Result 4.5.1 (Conditional MSEP estimator for a single accident year) *Under Model Assumptions (TM.1)-(TM.3), the conditional mean-square error of one year claims development result for a single accident year can be estimated by*

$$\widehat{msep}_{\widehat{CDR}_i(I+1)|D_I}(0) = \left(\widehat{C}_{i,J}^{I}\right)^2 \cdot \left(\widehat{\Gamma}_{i,J}^{I} + \widehat{\Delta}_{i,J}^{I}\right), \tag{4.5.1}$$

where

$$\widehat{\Gamma}^I_{i,J} = \prod_{l=I-i}^{J-1}\left(\frac{\widehat{\sigma}_l^2/\widehat{f}_l^2}{(\beta_l^I)^2}\cdot\omega_l+1\right)+\prod_{l=I-i+1}^{J-1}\left(\alpha_l^2\cdot\frac{\widehat{\sigma}_l^2/\widehat{f}_l^2}{(\beta_l^I)^2}\cdot\omega_l+1\right)$$
$$-2\prod_{l=I-i+1}^{J-1}\left(\alpha_l\cdot\frac{\widehat{\sigma}_l^2/\widehat{f}_l^2}{(\beta_l^I)^2}\cdot\omega_l+1\right), \tag{4.5.2}$$

$$\widehat{\Delta}^I_{i,J} = \left(\widehat{C}_{i,J}\right)^2\cdot\left\{\left[1+\frac{\widehat{\sigma}_{I-i}^2/\left(\widehat{f}_{I-i}^I\right)^2}{\delta_{i,I-i}}\right]\cdot\prod_{l=I-i+1}^{J-1}\left[1+\frac{\widehat{\sigma}_l^2/\left(\widehat{f}_l^I\right)^2\frac{\gamma_{I-l,l}^2}{\delta_{I-l,l}}}{\left(\beta_l^{I+1}\right)^2}\right]-1\right\}, \tag{4.5.3}$$

$$\omega_l = \sum_{i=0}^{I-l-1}\frac{\gamma_{i,l}^2}{\delta_{i,l}}, \tag{4.5.4}$$

$$\alpha_l = \frac{\beta_l^I}{\beta_l^{I+1}}, \tag{4.5.5}$$

$$\beta_l^I = \sum_{i=0}^{I-l-1}\gamma_{i,l}, \tag{4.5.6}$$

$$\beta_l^{I+1} = \sum_{i=0}^{I-l}\gamma_{i,l}, \tag{4.5.7}$$

and $\widehat{f}_l$ *and* $\widehat{\sigma}_l^2$ *are given in (4.2.8) and (4.2.10)-(4.2.11) respectively.*

4.5.2 Single Accident Year-Approximating Formula

Result 4.5.2 (Conditional MSEP estimator for a single accident year) *Under assumptions of Result 4.5.1, the estimate of the mean-square error of claims development result for a single accident year can be approximated by*

$$\widehat{msep}_{\widehat{CDR}_i(I+1)|D_I}(0) = \left(\widehat{C}_{i,J}^I\right)^2\cdot\left(\widehat{\Gamma}^I_{i,J}+\widehat{\Delta}^I_{i,J}\right), \tag{4.5.8}$$

where

$$\widehat{\Gamma}^I_{i,J} \cong \frac{\widehat{\sigma}^2_{I-i}/f^2_{I-i}}{(\beta^I_{I-i})^2} \cdot \omega_{I-i} + \sum_{l=I-i+1}^{J-1} \left(\frac{\gamma_{I-l,l}}{\beta^{I+1}_l}\right)^2 \frac{\widehat{\sigma}_l{}^2/f^2_l}{(\beta^I_l)^2} \cdot \omega_l, \tag{4.5.9}$$

$$\widehat{\Delta}^I_{i,J} \cong \frac{\widehat{\sigma}^2_{I-i}/\left(\widehat{f}^I_{I-i}\right)^2}{\delta_{i,I-i}} + \sum_{l=I-i+1}^{J-1} \frac{\widehat{\sigma}^2_l/\left(\widehat{f}^I_l\right)^2 \frac{\gamma^2_{I-l,l}}{\delta_{I-l,l}}}{\left(\beta^{I+1}_l\right)^2}. \tag{4.5.10}$$

and β^I_l, β^{I+1}_l, ω_l, $\widehat{f}_l$ *and* $\widehat{\sigma}^2_l$ *are given in (4.5.6), (4.5.7), (4.5.4), (4.2.8) and (4.2.10)-(4.2.11) respectively.*

4.5.3 Aggregation over Prior Accident Year - Exact Formula

Result 4.5.3 (Conditional MSEP estimator for aggregated years)

Under Model Assumptions (TM.1)-(TM.3), the conditional mean-square error of one-year claims development result for aggregated accident years, can be estimated by

$$\widehat{msep}_{\sum_{i=1}^I \widehat{CDR}_i(I+1)|D_I}(0) = \sum_{i=1}^I \widehat{msep}_{\widehat{CDR}_i(I+1)|D_I}(0) + 2\sum_{k>i>0} \widehat{C}^I_{i,J} \cdot \widehat{C}^I_{k,J} \left(\widehat{\Upsilon}^I_{i,J} + \widehat{\Phi}^I_{i,J}\right), \tag{4.5.11}$$

where

$$\widehat{\Upsilon}^I_{i,J} = \left(1 + \frac{\widehat{\sigma}^2_{I-i}/\left(\widehat{f}^I_{I-i}\right)^2}{\beta^{I+1}_{I-i}} \cdot \frac{\gamma_{i,I-i}}{\delta_{i,I-i}}\right) \cdot \prod_{l=I-i+1}^{J-1} \left(1 + \frac{\widehat{\sigma}^2_l/\left(\widehat{f}^I_l\right)^2}{\left(\beta^{I+1}_l\right)^2} \cdot \frac{\gamma^2_{I-l,l}}{\delta_{I-l,l}}\right) - 1, \tag{4.5.12}$$

$$\widehat{\Phi}^I_{i,J} = \prod_{l=I-i}^{J-1}\left(\frac{\widehat{\sigma}_l^2/\left(\widehat{f}_l^I\right)^2}{(\beta_l^I)^2}\cdot\omega_l + 1\right) - \prod_{l=I-i}^{J-1}\left(\alpha_l\cdot\frac{\widehat{\sigma}_l^2/\left(\widehat{f}_l^I\right)^2}{(\beta_l^I)^2}\cdot\omega_l + 1\right)$$
$$- \prod_{l=I-i+1}^{J-1}\left(\alpha_l\cdot\frac{\widehat{\sigma}_l^2/\left(\widehat{f}_l^I\right)^2}{(\beta_l^I)^2}\cdot\omega_l + 1\right) + \prod_{l=I-i+1}^{J-1}\left(\alpha_l^2\cdot\frac{\widehat{\sigma}_l^2/\left(\widehat{f}_l^I\right)^2}{(\beta_l^I)^2}\cdot\omega_l + 1\right), \tag{4.5.13}$$

and β_l^I, β_l^{I+1}, ω_l, α_l, $\widehat{f}_l$ and $\widehat{\sigma}_l^2$ are given in (4.5.6), (4.5.7), (4.5.4),(4.5.5), (4.2.8) and (4.2.10)-(4.2.11) respectively.

4.5.4 Aggregation over Prior Accident Year - Approximating Formula

Result 4.5.4 (Conditional MSEP estimator for aggregated years) *Under assumptions of Result 4.5.3, the estimate of the mean-square error of claims development result for aggregated accident years, can be approximated by*

$$\begin{aligned}\widehat{msep}_{\sum_{i=1}^I \widehat{CDR}_i(I+1)|D_I}(0) = &\sum_{i=1}^{I}\widehat{msep}_{\widehat{CDR}_i(I+1)|D_I}(0)\\ &+ 2\sum_{k>i>0}\widehat{C}^I_{i,J}\cdot\widehat{C}^I_{k,J}\left(\widehat{\Upsilon}^I_{i,J}+\widehat{\Phi}^I_{i,J}\right),\end{aligned} \tag{4.5.14}$$

where

$$\widehat{\Upsilon}^I_{i,J} \cong \frac{\widehat{\sigma}_{I-i}^2/\left(\widehat{f}_{I-i}^I\right)^2}{\beta_{I-i}^{I+1}}\cdot\frac{\gamma_{i,I-i}}{\delta_{i,I-i}} + \sum_{l=I-i+1}^{J-1}\frac{\widehat{\sigma}_l^2/\left(\widehat{f}_l^I\right)^2}{\left(\beta_l^{I+1}\right)^2}\cdot\frac{\gamma_{I-l,l}^2}{\delta_{I-l,l}}, \tag{4.5.15}$$

$$\widehat{\Phi}^I_{i,J} \cong \frac{\gamma_{i,I-i}}{\beta_l^{I+1}}\cdot\frac{\widehat{\sigma}_{I-i}^2/\left(\widehat{f}_{I-i}\right)^2}{(\beta_{I-i}^I)^2}\cdot\omega_{I-i} + \sum_{l=I-i+1}^{J-1}\frac{\widehat{\sigma_l}^2/\left(\widehat{f}_l^I\right)^2}{(\beta_l^I)^2}\cdot\omega_l\left(\frac{\gamma_{I-l,l}}{\beta_l^{I+1}}\right)^2, \tag{4.5.16}$$

and β_l^I, β_l^{I+1}, ω_l, $\widehat{f}_l$ and $\widehat{\sigma}_l^2$ are given in (4.5.6), (4.5.7), (4.5.4), (4.2.8) and (4.2.10)-(4.2.11) respectively.

Remark 4.5.1 *If we set $\gamma_{i,j} = C_{i,j}$ and $\delta_{i,j} = C_{i,j}$, we obtain from our general framework the main results of Merz and Wüthrich (2008a).*

4.6 Numerical example

In this section, we present a case study illustrating the use of our theoretical results. For our example we use two run-off triangles: Triange 1 (**T1**) given in Table B.1 in Appendix B (see also Table 2 in Merz and Wüthrich (2008a)) and Triange 2 (**T2**) given in Table B.2 in Appendix B (see also Table 1, p.22 in England and Verrall (2002)).
These datasets contain cumulative payments $C_{i,j}$, where $(i,j) \in \{0,1,...,8\} \times \{0,1,...,8\}$ for **T1** and $(i,j) \in \{0,1,...,9\} \times \{0,1,...,9\}$ for **T2**.

As in Mack (1999) we consider here the following family of models, for $0 \le i \le I$ and $0 \le j \le J$,

$$\gamma_{i,j} = w^{\gamma}_{i,j} \cdot C^{\alpha}_{i,j} \quad \text{and} \quad \delta_{i,j} = w^{\delta}_{i,j} \cdot C^{\beta}_{i,j}, \tag{4.6.1}$$

where $\alpha, \beta \ge 0$ and $w^{\gamma}_{i,j}, w^{\delta}_{i,j} \in [0,1]$ are arbitrary weights which can be used by the actuary to downweight any outlying link ratios $F_{i,j}$ defined via (4.2.6). Hence, according to (4.2.8) and (4.2.10) , the estimators $\widehat{f}_k$ and $\widehat{\sigma}^2_k$ are defined by

$$\widehat{f}_k = \frac{\sum_{i=0}^{I-k-1} w^{\gamma}_{i,k} C^{\alpha}_{i,k} F_{i,k}}{\sum_{i=0}^{I-k-1} w^{\gamma}_{i,k} C^{\alpha}_{i,k}}, \text{ for } 0 \le k \le J-1,$$

$$\widehat{\sigma}^2_k = \frac{1}{I-k-1} \sum_{i=0}^{I-k-1} w^{\delta}_{i,j} \cdot C^{\beta}_{i,j} (F_{i,k} - \widehat{f}_k)^2, \text{ for } 0 \le k \le J-2.$$

4.6.1 Cas study 1 - no adjustments of link ratios

In this study we use all data information in estimation of the model parameters f_k and σ^2_k, so we put all the weights $w^{\gamma}_{i,j}$ and $w^{\delta}_{i,j}$ equal to 1. We define below 3 different setups $A1 - A3$ of our model by specifying only the parameters α and β.

A-1 $\alpha = \beta = 0$ and $w^{\gamma}_{i,j} = w^{\delta}_{i,j} = 1$ for all i, j.

In this case, we get the model for which the estimators of the age-to-age factors f_k are the "unweighted average" of the observed link ratios $F_{i,j}$ defined via (4.2.6), i.e.,

$$\widehat{f}_k = \frac{1}{I-k-1} \sum_{i=0}^{I-k-1} F_{i,k}, \text{ for } 0 \leq k \leq J-1.$$

A-2 $\alpha = \beta = 1$ and $w_{i,j}^{\gamma} = w_{i,j}^{\delta} = 1$ for all i, j.
In this case, we get the classical Chain Ladder estimate of f_k

$$\widehat{f}_k = \frac{\sum_{i=0}^{I-k-1} C_{i,k} F_{i,k}}{\sum_{i=0}^{I-k-1} C_{i,k}} = \frac{\sum_{i=0}^{I-k-1} C_{i,k+1}}{\sum_{i=1}^{I-k} C_{i,k}}, \text{ for } 0 \leq k \leq J-1.$$

A-3 $\alpha = \beta = 2$ and $w_{i,j}^{\gamma} = w_{i,j}^{\delta} = 1$ for all i, j.
In this case, we get the model for which the estimators of the development factors f_k are the results of an ordinary regression of $\{C_{i,k+1}\}_{i \in \{1,\ldots,I-k-1\}}$ against $\{C_{i,k}\}_{i \in \{1,\ldots,I-k\}}$ with intercept 0, i.e.,

$$\widehat{f}_k = \frac{\sum_{i=0}^{I-k-1} C_{i,k}^2 F_{i,k}}{\sum_{i=0}^{I-k-1} C_{i,k}^2} = \frac{\sum_{i=0}^{I-k-1} C_{i,k} C_{i,k+1}}{\sum_{i=0}^{I-k-1} C_{i,k}^2}, \text{ for } 0 \leq k \leq J-1.$$

Concerning the run-off triangle T1, Table 4.4 and Table 4.5 summarize the estimates $\widehat{f}_j^I$ of the factors f_j together with their coefficients of variation $CV(\widehat{f}_j)$. These estimates are used to compute the claims reserves $\widehat{R}_i^I$ for the outstanding claims liabilities R_i^I (see Table 4.3). In Table 4.2 we present the estimates of the predictions defined in (4.4.1) and (4.4.2) and given by the approximating formulas (4.5.8)-(4.5.10) and (4.5.14)-(4.5.16) respectively .

Conclusions (Methods A1-A3 applied to the run-off triangle T1).
For the reserves and the uncertainty of the CDR (see Table 4.2 and Table 4.3), the obtained results are very close. We explain that by the close results of the estimated link ratios and their coefficients of variation (see Tables 4.4 and 4.5).

Conclusions (Methods A1-A3 applied to the run-off triangle T2).
Regarding to the uncertainty of the CDR (see Table 4.6), we observe the large deviations of different setups A1-A3. This is mainly due to two effects: the divergence of the estimated reserves (see Table 4.7) and high differences in the first estimated link ratios f_k (see Table 4.8) and their volatility illustrated by the coefficients of variation $CV(f_k)$ (see Table4.9).

	$\widehat{msep}_{\widehat{CDR}_i(I+1)\|D_I}(0)^{1/2}$		
i/method	A1(T1)	A2(T1)	A3(T1)
0	-	-	-
1	563	567	572
2	1 501	1 488	1 475
3	3 863	3 923	3 982
4	9 634	9 723	9 812
5	28 320	28 443	28 563
6	20 460	20 954	21 475
7	27 485	28 119	28 783
8	52 017	53 320	54 690
Total	**79 749**	**81 080**	**82 468**
	A1/A2	A2/A2	A3/A2
Total (%)	**98%**	**100%**	**102%**

Table 4.2: Estimates of prediction uncertainty of claims development result (CDR)

	R_i^I		
i/method	A1(T1)	A2(T1)	A3(T1)
0	-	-	-
1	4 378	4 378	4 378
2	9 373	9 348	9 322
3	28 495	28 392	28 292
4	51 797	51 444	51 097
5	112 875	111 811	110 736
6	188 445	187 084	185 699
7	413 491	411 864	410 206
8	1 434 720	1 433 505	1 432 291
Total	**2 243 574**	**2 237 826**	**2 232 020**
	A1/A2	A2/A2	A3/A2
Total (%)	**100,3%**	**100,0%**	**99,7%**

Table 4.3: Estimates of the claims reserves

	$\widehat{f}_j^I$		
j/method	A1(T1)	A2(T1)	A3(T1)
0	1,48	1,48	1,48
1	1,07	1,07	1,07
2	1,02	1,02	1,02
3	1,02	1,02	1,02
4	1,01	1,01	1,01
5	1,01	1,01	1,01
6	1,00	1,00	1,00
7	1,00	1,00	1,00

Table 4.4: Estimates of the age-to-age factors

	$CV(\widehat{f}_j^I)$		
j/method	A1(T1)	A2(T1)	A3(T1)
0	0,48%	0,49%	0,49%
1	0,27%	0,27%	0,27%
2	0,21%	0,21%	0,21%
3	0,31%	0,31%	0,31%
4	0,12%	0,12%	0,12%
5	0,05%	0,05%	0,05%
6	0,02%	0,02%	0,02%
7	0,01%	0,01%	0,01%

Table 4.5: Estimates of the coefficient of variation of link ratios

	$\widehat{msep}_{\widehat{CDR}_i(I+1)\mid D_I}(0)^{1/2}$		
i/method	A1 (T2)	A2 (T2)	A3 (T2)
1	203	206	209
2	636	579	530
3	408	396	389
4	1 591	1 303	1 076
5	1 467	1 668	1 887
6	1 161	1 185	1 223
7	6 593	4 693	3 571
8	5 316	4 708	4 455
9	89 999	24 142	11 636
Total	**90 838**	**25 681**	**13 853**
	A1/A2	A2/A2	A3/A2
Total (%)	**354%**	**100%**	**54%**

Table 4.6: Estimates of prediction uncertainty of claims development result (CDR)

	R_i^I		
i/method	A1 (T2)	A2 (T2)	A3 (T2)
0	-	-	-
1	154	154	154
2	642	617	592
3	1 695	1 635	1 575
4	2 844	2 745	2 647
5	3 952	3 647	3 343
6	5 890	5 438	5 015
7	12 368	10 911	10 154
8	12 385	10 653	9 625
9	53 729	16 343	10 671
Total	**93 659**	**52 143**	**43 776**
	A1/A2	A2/A2	A3/A2
Total (%)	**180%**	**100%**	**84%**

Table 4.7: Estimates of the claims reserves

	$\widehat{f}_j^I$		
j/method	A1 (T2)	A2 (T2)	A3 (T2)
0	8,21	3,00	2,22
1	1,70	1,62	1,57
2	1,31	1,27	1,26
3	1,18	1,17	1,16
4	1,13	1,11	1,10
5	1,04	1,04	1,04
6	1,03	1,03	1,03
7	1,02	1,02	1,02
8	1,01	1,01	1,01

Table 4.8: Estimates of the age-to-age factors

	$CV(\widehat{f}_j^I)$		
j/method	A1 (T2)	A2 (T2)	A3 (T2)
0	50,1%	37,7%	18,5%
1	9,9%	8,4%	6,9%
2	9,1%	7,1%	5,6%
3	2,3%	2,2%	2,0%
4	3,0%	3,2%	3,3%
5	2,4%	2,2%	1,9%
6	0,5%	0,5%	0,5%
7	1,5%	1,5%	1,5%
8	0,9%	0,8%	0,8%

Table 4.9: Estimates of the coefficient of variation of link ratios

4.6.2 Cas study 2 - with adjustments of link ratios $F_{i,k}$

We define below 3 other setups $B1-B3$ in which we exclude some individual age to age factors $F_{i,k}$ from the estimation of parameters f_k et σ_k^2. The factors to be omitted in estimation are defined by two sets of indices (i, j), W_0^γ (for f_k) and W_0^δ (for σ_k^2) . In the first place we consider that $W_0^\gamma = W_0^\delta := W_0$. Note that this assumption implies that $\gamma_{i,j} = \delta_{i,j}$, for all i, j, which means that we obtain the optimal properties of the estimators for f_k et σ_k^2 (see Proposition 4.2.1). For the run-off triangle **T2** from the Table B.2 the set W_0 is defined by

$$W_0 = \{(1, 0), (4, 0), (6, 1), (7, 1), (1, 5))\}.$$

Note that we use the same weights in estimation of f_k et σ_k^2 parameters. The definition of W_0 means that the following link ratios are excluded from the estimation of $\widehat{f}_k$ et $\widehat{\sigma}_k^2$: $\{F_{1,0}, F_{4,0}, F_{6,1}, F_{7,1}, F_{1,5}\}$.

B-1 $\alpha = \beta = 0$, $w^{\gamma}_{i,j} = w^{\delta}_{i,j} = 0$ for $(i,j) \in W_0$ and $w^{\gamma}_{i,j} = w^{\delta}_{i,j} = 1$ for $(i,j) \notin W_0$.

B-2 $\alpha = \beta = 1$, $w^{\gamma}_{i,j} = w^{\delta}_{i,j} = 0$ for $(i,j) \in W_0$ and $w^{\gamma}_{i,j} = w^{\delta}_{i,j} = 1$ for $(i,j) \notin W_0$.

B-3 $\alpha = \beta = 2$, $w^{\gamma}_{i,j} = w^{\delta}_{i,j} = 0$ for $(i,j) \in W_0$ and $w^{\gamma}_{i,j} = w^{\delta}_{i,j} = 1$ for $(i,j) \notin W_0$.

	$\widehat{msep}_{\widehat{CDR}_i(I+1)\mid D_I}(0)^{1/2}$		
i/method	B1 (T2)	B2 (T2)	B3 (T2)
1	203	206	209
2	636	579	530
3	408	396	389
4	1 247	1 080	929
5	1 488	1 686	1 900
6	1 177	1 197	1 230
7	6 698	4 745	3 598
8	2 193	2 238	2 475
9	12 202	8 792	6 966
Total	**15 371**	**11 442**	**9 443**
	B1/B2	B2/B2	B3/B2
Total (%)	**134%**	**100%**	**83%**

Table 4.10: Estimates of prediction uncertainty of claims development result (CDR)

Conclusions (Methods B1-B3 applied to run-off triangle T2).
For the uncertainty of the CDR (see Table 4.10), we see that the fact of excluding the link ratios considered atypical, reduced significantly the differences between results of the setups B1-B3 applied to T2. It can be explained by the fact that the differences of reserves and coefficients of variation of estimated link ratios also decreased in the same manner (see Tables 4.11 ,4.12 and 4.13).

	R_i^I		
i/method	B1 (T2)	B2 (T2)	B3 (T2)
0	-	-	-
1	154	154	154
2	642	617	592
3	1 695	1 635	1 575
4	3 306	3 071	2 868
5	4 267	3 867	3 490
6	6 180	5 638	5 148
7	12 774	11 182	10 333
8	10 505	9 561	8 984
9	19 366	12 247	9 583
Total	**58 890**	**47 971**	**42 727**
	B1/B2	B2/B2	B3/B2
Total (%)	**123%**	**100%**	**89%**

Table 4.11: Estimates of the claims reserves

Finally we define 3 last setups $C1 - C3$ for which the sets W_0^γ and W_0^δ are different. Often in actuarial practice, we consider W_0^δ as an empty set which means that we want to take into account all information in order to estimate the parameters σ_k^2. However, in the present case of run-off triangle **T2** we observe (without any inside knowledge) that the factor $F_{1,0}$ $(= 40.4)$ is clearly an outlier so we decided to excluded it from estimation of the variance parameters σ_k^2. Therefore, we set

$$W_0^\delta = \{(1,0)\}.$$

Note that the parameters f_k are estimated with the same weights as in the setups B1-B3, i.e.,

$$W_0^\gamma = W_0 = \{(1,0); (4,0); (6,1); (7,1), (1,5))\}.$$

C-1 $\alpha = \beta = 0$, $w_{i,j}^\gamma = 0$ for $(i,j) \in W_0$ and $w_{i,j}^\gamma = 1$ for $(i,j) \notin W_0$, $w_{i,j}^\delta = 0$ for $(i,j) \in W_0^\delta$ and $w_{i,j}^\delta = 1$ for $(i,j) \notin W_0^\delta$.

C-2 $\alpha = \beta = 1$, $w_{i,j}^\gamma = 0$ for $(i,j) \in W_0$ and $w_{i,j}^\gamma = 1$ for $(i,j) \notin W_0$, $w_{i,j}^\delta = 0$ for $(i,j) \in W_0^\delta$ and $w_{i,j}^\delta = 1$ for $(i,j) \notin W_0^\delta$.

C-3 $\alpha = \beta = 2$, $w_{i,j}^\gamma = 0$ for $(i,j) \in W_0$ and $w_{i,j}^\gamma = 1$ for $(i,j) \notin W_0$, $w_{i,j}^\delta = 0$ for $(i,j) \in W_0^\delta$ and $w_{i,j}^\delta = 1$ for $(i,j) \notin W_0^\delta$.

	$\widehat{f}_j^I$		
j/method	B1 (T2)	B2 (T2)	B3 (T2)
0	3,52	2,50	2,12
1	1,49	1,50	1,49
2	1,31	1,27	1,26
3	1,18	1,17	1,16
4	1,13	1,11	1,10
5	1,06	1,05	1,05
6	1,03	1,03	1,03
7	1,02	1,02	1,02
8	1,01	1,01	1,01

Table 4.12: Estimates of the age-to-age factors

	$CV(\widehat{f}_j^I)$		
j/method	B1 (T2)	B2 (T2)	B3 (T2)
0	19,5%	17,9%	12,4%
1	5,0%	4,6%	4,3%
2	9,1%	7,1%	5,6%
3	2,3%	2,2%	2,0%
4	3,0%	3,2%	3,3%
5	2,0%	1,9%	1,7%
6	0,5%	0,5%	0,5%
7	1,5%	1,5%	1,5%
8	0,9%	0,8%	0,8%

Table 4.13: Estimates of the coefficient of variation of link ratios

	$\widehat{msep}_{\widehat{CDR}_i(I+1)\mid D_I}(0)^{1/2}$		
i/method	C1 (T2)	C2 (T2)	C3 (T2)
1	203	206	209
2	636	579	530
3	408	396	389
4	1 754	1 395	1 127
5	1 503	1 694	1 905
6	1 193	1 207	1 237
7	6 703	4 750	3 602
8	6 032	5 090	4 669
9	18 734	13 799	10 628
Total	**22 074**	**16 482**	**13 130**
	C1/C2	C2/C2	C3/C2
Total (%)	**134%**	**100%**	**80%**

Table 4.14: Estimates of prediction uncertainty of claims development result (CDR)

	$\widehat{f}_j^I$		
j/method	C1 (T2)	C2 (T2)	C3 (T2)
0	3,52	2,50	2,12
1	1,49	1,50	1,49
2	1,31	1,27	1,26
3	1,18	1,17	1,16
4	1,13	1,11	1,10
5	1,06	1,05	1,05
6	1,03	1,03	1,03
7	1,02	1,02	1,02
8	1,01	1,01	1,01

Table 4.15: Estimates of the age-to-age factors

	$CV(\widehat{f}_j^I)$		
j/method	C1 (T2)	C2 (T2)	C3 (T2)
0	27,8%	27,1%	18,2%
1	14,3%	10,7%	8,1%
2	9,1%	7,1%	5,6%
3	2,3%	2,2%	2,0%
4	3,0%	3,2%	3,3%
5	2,9%	2,5%	2,1%
6	0,5%	0,5%	0,5%
7	1,5%	1,5%	1,5%
8	0,9%	0,8%	0,8%

Table 4.16: Estimates of the coefficient of variation of link ratios

Conclusions (Methods C1-C3 applied to the run-off triangle T2).
For the reserves (see Table 4.11) , we observe that comparing to the setups B1-B3 we obtain the same results because we used the same set of the zero weights W_0^γ.
Concerning the uncertainty of CDR given in Table 4.14, we obtain the higher estimates of these quantities. This is mainly due to the fact that we took almost all (except one) link ratios in estimation of parameters σ_k. It means that we reduced the set of the zero weights W_0^δ and in consequence, we increased the "volatility" of the reserves.

Overall Conclusions

In our case study we considered several setups of our universal tool of reserving and measuring the one year volatility of the reserves. In the daily practice of actuaries we are often confronted to choose the proper model. This is not an easy issue in the case of run-off triangle T2 unlike to the run-off triangle T1 where all the methods give almost the same results. In our case study concerning the triangle T2, we proceeded without any inside knowledge about the reserving procedure and historical data. Therefore we obtained rather large range of possible values of estimators. This shows that the choice of model for reserving processes is still an open challenging problem and underlines the importance of statistical inference methods to properly assess the model structure in each case.

Finally, in the present chapter, we presented the approach which can be used within the Solvency II framework (standard formula calibration (USPs methods) or (full or partial) internal models for reserve risk).

4.7 Mathematical Proofs

For the notation convenience and to simplify our exposition, we will consider only the case where $\delta_{i,k} \neq 0$ (a.s.) in the assumptions of Chain Ladder Time Series Model presented in (4.2.5). It will become obvious from our arguments that our results hold for general assumptions. Therefore, from the assumption (4.2.12), to cover the general case one should add the term $\mathbf{1}_{\{\delta_{j,k}\neq 0\}}$ (if necessary), where $\mathbf{1}_A$ denotes the indicator function for event A.
As already mentioned, the proofs of our main results are obtained by mimicking the arguments from Merz and Wüthrich (2008a). They are derived from the following lemmas:

4.7.1 Useful Lemmas

Lemma 4.7.1 *Under the assumptions of the Chain Lader Time Series Model (see Section 4.2.1), we have*

(a) $C_{i,I-i+1}, \widehat{f}^{I+1}_{I-i+1}, ..., \widehat{f}^{I+1}_{J-1}$ *are conditionally independent with respect to* D_I

(b) $E\left(\widehat{f}^{I+1}_l | D_I\right) = \frac{\beta^I_l}{\beta^{I+1}_l} \cdot \widehat{f}^I_l + f_l \cdot \frac{\gamma_{I-l,l}}{\beta^{I+1}_l}$, *where* β^I_l *and* β^{I+1}_l *are defined in (4.5.6) and (4.5.7) respectively*

(c) $E\left[\widehat{C}^{I+1}_{i,j} | D_I\right] = C_{i,I-i} \cdot f_{I-i} \cdot \prod_{l=I-i+1}^{j-1} E\left[\widehat{f}^{I+1}_l | D_I\right]$

(d) $E\left[C^2_{i,I-i+1} | D_I\right] = f^2_{I-i} \cdot C^2_{i,I-i} + \frac{\sigma^2_{I-i}}{\delta_{i,I-i}} \cdot C^2_{i,I-i}$

(e) $E\left[(\widehat{f}^{I+1}_l)^2 | D_I\right] = \left(\frac{\sum_{i=0}^{I-l-1} \frac{\gamma_{i,l}}{C_{i,l}} \cdot C_{i,l+1}}{\beta^{I+1}_l} + \frac{\gamma_{l,I-l} \cdot f_l}{\beta^{I+1}_l}\right)^2 + \frac{\sigma^2_l \cdot \frac{\gamma^2_{I-l,l}}{\delta_{I-l,l}}}{\left(\beta^{I+1}_l\right)^2}.$

(f) We have,

$$E\left[C_{i,I-i+1} \cdot \widehat{f}^{I+1}_{I-i} | D_I\right] = \frac{1}{\beta^{I+1}_{I-i}} \left[\left(\sum_{k=0}^{i-1} \frac{\gamma_{k,I-i}}{C_{k,I-i}} \cdot C_{k,I-i+1}\right) \cdot f_{I-i} \cdot C_{i,I-i}\right] + \frac{1}{\beta^{I+1}_{I-i}} \left[\sigma^2_{I-i} \cdot C_{i,I-i} \cdot \frac{\gamma_{i,I-i}}{\delta_{i,I-i}} + \gamma_{i,I-i} \cdot f^2_{I-i} \cdot C_{i,I-i}\right]$$

The following Lemma correspond to Lemma 3.3, p.123, of Merz and Wüthrich (2007)

Lemma 4.7.2

$$E[\widehat{CDR}_i(I+1)|D_I] = C_{i,I-i}\left(\prod_{l=I-i}^{J-1}\widehat{f}_l^I - f_{I-i}^I \cdot \prod_{l=I-i+1}^{J-1}\left(\frac{\beta_l^I}{\beta_l^{I+1}}\widehat{f}_l^I + f_l \cdot \frac{\gamma_{I-l,l}}{\beta_l^{I+1}}\right)\right).$$

Lemma 4.7.3 *Set* $\alpha_l = \frac{\beta_l^I}{\beta_l^{I+1}}$. *We have*

(a) $E\left[\left(\widehat{f}_l^I\right)^2|D_I\right] = Var\left(\widehat{f}_l^I|D_I\right) + f_l^2$

(b) $E\left[\left(\alpha_l \cdot \widehat{f}_l^I + f_l \cdot \frac{\gamma_{I-l,l}}{\beta_l^{I+1}}\right)^2|D_I\right] = \alpha_l^2 \cdot Var\left(\widehat{f}_l^I|D_I\right) + f_l^2$

(c) *For* $i = 1, ..., I$,

$$E\left[\prod_{l=I-i}^{J-1}\widehat{f}_l^I \cdot f_{I-i} \cdot \prod_{l=I-i+1}^{J-1}\left(\alpha_l \cdot \widehat{f}_l^I + f_l \cdot \frac{\gamma_{I-l,l}}{\beta_l^{I+1}}\right)|D_I\right]$$
$$= f_{I-i}^2 \cdot \prod_{l=I-i+1}^{J-1}\left[\alpha_l \cdot Var\left(\widehat{f}_l^I|D_I\right) + f_l^2\right]$$

(d) $Var\left(\widehat{f}_l^I|D_I\right) = \frac{\sigma_l^2}{(\beta_l^I)^2}\sum_{i=0}^{I-l-1}\frac{\gamma_{i,l}^2}{\delta_{i,l}}$

Lemma 4.7.4 *We have*

(a) $\widehat{E}\left(\widehat{f}_j^{I+1}|D_I\right) = \widehat{f}_j^I$

(b) $\widehat{E}\left[\left(\widehat{f}_j^{I+1}\right)^2|D_I\right] = \left(\widehat{f}_j^I\right)^2 \cdot \left(1 + \frac{\left(\widehat{\sigma}_j^I\right)^2/\left(\widehat{f}_j^I\right)^2}{\left(\beta_j^{I+1}\right)^2} \cdot \frac{\gamma_{I-j,j}^2}{\delta_{I-j,j}}\right)$

(c) $\widehat{E}\left[C_{i,I-i+1} \cdot \widehat{f}_{I-i}^{I+1}|D_I\right] = C_{i,I-i} \cdot \widehat{f}_{I-i}^2\left(1 + \frac{\widehat{\sigma}_{I-i}^2/\widehat{f}_{I-i}^2}{\beta_{I-i}^{I+1}}\frac{\gamma_{i,I-i}}{\delta_{i,I-i}}\right)$

Lemma 4.7.5 *For*

$$\delta_l := \frac{\beta_l^I}{\beta_l^{I+1}} \widehat{f}_l^I + f_l \cdot \frac{\gamma_{I-l,l}}{\beta_l^{I+1}}, \tag{4.7.1}$$

we have (see p. 128-129 in Merz and Wüthrich (2007))

(a) $E\left[\left(\widehat{f}_l^I\right)^2 |D_I\right] = \frac{\sigma_j^2}{(\beta_j^I)^2} \sum_{i=0}^{I-j-1} \frac{\gamma_{i,j}^2}{\delta_{i,j}} + f_l^2$

(b) $E\left[\delta_l^2|D_I\right] = \left(\frac{\beta_j^I}{\beta_j^{I+1}}\right)^2 \cdot \frac{\sigma_j^2}{(\beta_j^I)^2} \sum_{i=0}^{I-j-1} \frac{\gamma_{i,j}^2}{\delta_{i,j}} + f_l^2$

(c) $E\left[\widehat{f}_l^I \cdot \delta_l|D_I\right] = \frac{\beta_j^I}{\beta_j^{I+1}} \cdot \frac{\sigma_j^2}{(\beta_j^I)^2} \sum_{i=0}^{I-j-1} \frac{\gamma_{i,j}^2}{\delta_{i,j}} + f_l^2$

The proofs of Lemmas 4.7.1-4.7.5 are postponed to section 4.7.6.

4.7.2 Proof of Result 4.5.1

From (4.4.1) we have

$$msep_{\widehat{CDR}_i(I+1)|D_I}(0) := E\left[\left(\widehat{CDR}_i(I+1) - 0\right)^2 |D_I\right] = E\left[\left(\widehat{CDR}_i(I+1)\right)^2 |D_I\right].$$

Since

$$Var\left(\widehat{CDR}_i(I+1)|D_I\right) := E\left[\left(\widehat{CDR}_i(I+1)\right)^2 |D_I\right] - \left[E\left(\widehat{CDR}_i(I+1)|D_I\right)\right]^2,$$

we obtain

$$msep_{\widehat{CDR}_i(I+1)|D_I}(0) = Var\left(\widehat{CDR}_i(I+1)|D_I\right) + \left[E\left(\widehat{CDR}_i(I+1)|D_I\right)\right]^2. \tag{4.7.2}$$

In view of (4.7.2) to prove Result 4.5.1 it is enough to show

$$\widehat{Var}\left(\widehat{CDR}_i(I+1)|D_I\right) = \left(\widehat{C}_{i,J}^I\right)^2 \cdot \widehat{\Gamma}_{i,J}^I, \tag{4.7.3}$$

$$\left[\widehat{E}\left(\widehat{CDR}_i(I+1)|D_I\right)\right]^2 = \left(\widehat{C}_{i,J}^I\right)^2 \cdot \widehat{\Delta}_{i,J}^I, \tag{4.7.4}$$

where $\widehat{\Gamma}^I_{i,J}$ and $\widehat{\Delta}^I_{i,J}$ are given via (4.5.9) and (4.5.10) respectively.

Proof of (4.7.3). From (4.3.4) we have

$$\begin{aligned} Var\left(\widehat{CDR}_i(I+1)|D_I\right) &= Var\left(\widehat{C}^I_{i,J} - \widehat{C}^{I+1}_{i,J}|D_I\right) \\ &= Var\left(\widehat{C}^I_{i,J}|D_I\right) + Var\left(\widehat{C}^{I+1}_{i,J}|D_I\right) \\ &- 2 \cdot Cov\left(\widehat{C}^I_{i,J}, \widehat{C}^{I+1}_{i,J}|D_I\right) = Var\left(\widehat{C}^{I+1}_{i,J}|D_I\right), \end{aligned}$$

where the last equality is the consequence of: $E(f(Z)|Z) = f(Z)$ and $Var(f(Z)|Z) = 0$, for any real function f.
From (4.3.6),

$$
\begin{aligned}
&Var\left(\widehat{C}_{i,J}^{I+1}|D_I\right)\\
&= E\left[\left(C_{i,I-i+1}\prod_{l=I-i+1}^{J-1}\widehat{f}_l^{I+1}\right)^2|D_I\right] - \left[E\left(C_{i,I-i+1}\prod_{l=I-i+1}^{J-1}\widehat{f}_l^{I+1}|D_I\right)\right]^2\\
&= E\left[C_{i,I-i+1}^2|D_I\right]\cdot\prod_{l=I-i+1}^{J-1}E\left[(\widehat{f}_l^{I+1})^2|D_I\right] - \left(E\left[C_{i,I-i+1}|D_I\right]\right)^2\cdot\prod_{l=I-i+1}^{J-1}\left(E\left[(\widehat{f}_l^{I+1})|D_I\right]\right)^2\\
&= \left[f_{I-i}^2\cdot C_{i,I-i}^2 + \frac{\sigma_{I-i}^2}{\delta_{i,I-i}}\cdot C_{i,I-i}^2\right]\cdot\prod_{l=I-i+1}^{J-1}E\left[(\widehat{f}_l^{I+1})^2|D_I\right]\\
&- f_{I-i}^2\cdot C_{i,I-i}^2\cdot\prod_{l=I-i+1}^{J-1}\left(E\left[(\widehat{f}_l^{I+1})|D_I\right]\right)^2\\
&= \left[f_{I-i}^2\cdot C_{i,I-i}^2 + \frac{\sigma_{I-i}^2}{\delta_{i,I-i}}\cdot C_{i,I-i}^2\right]\cdot\prod_{l=I-i+1}^{J-1}\left[\left(\widehat{f}_l^I\right)^2 + \frac{\widehat{\sigma}_l^2}{\left(\beta_l^{I+1}\right)^2}\cdot\frac{\gamma_{I-l,l}^2}{\delta_{I-l,l}}\right]\\
&- f_{I-i}^2\cdot C_{i,I-i}^2\cdot\prod_{l=I-i+1}^{J-1}\left(\widehat{f}_l^I\right)^2\\
&= \left(\widehat{f}_{I-i}^I\right)^2\cdot C_{i,I-i}^2\cdot\left[1+\frac{\widehat{\sigma}_{I-i}^2/\left(\widehat{f}_{I-i}^I\right)^2}{\delta_{i,I-i}}\right]\cdot\prod_{l=I-i+1}^{J-1}\left(\widehat{f}_l^I\right)^2\left[1+\frac{\widehat{\sigma}_l^2/\left(\widehat{f}_l^I\right)^2}{\left(\beta_l^{I+1}\right)^2}\cdot\frac{\gamma_{I-l,l}^2}{\delta_{I-l,l}}\right]\\
&- \left(\widehat{f}_{I-i}^I\right)^2\cdot C_{i,I-i}^2\cdot\prod_{l=I-i+1}^{J-1}\left(\widehat{f}_l^I\right)^2\\
&= C_{i,I-i}^2\cdot\left(\widehat{f}_{I-i}^I\right)^2\cdot\prod_{l=I-i+1}^{J-1}\left(\widehat{f}_l^I\right)^2\cdot\left[1+\frac{\widehat{\sigma}_{I-i}^2/\left(\widehat{f}_{I-i}^I\right)^2}{\delta_{i,I-i}}\right]\cdot\prod_{l=I-i+1}^{J-1}\left[1+\frac{\widehat{\sigma}_l^2/\left(\widehat{f}_l^I\right)^2\frac{\gamma_{I-l,l}^2}{\delta_{I-l,l}}}{\left(\beta_l^{I+1}\right)^2}\right]\\
&- C_{i,I-i}^2\cdot\left(\widehat{f}_{I-i}^I\right)^2\cdot\prod_{l=I-i+1}^{J-1}\left(\widehat{f}_l^I\right)^2\\
&= C_{i,I-i}^2\cdot\prod_{l=I-i}^{J-1}\left(\widehat{f}_l^I\right)^2\cdot\left[1+\frac{\widehat{\sigma}_{I-i}^2/\left(\widehat{f}_{I-i}^I\right)^2}{\delta_{i,I-i}}\right]\cdot\prod_{l=I-i+1}^{J-1}\left[1+\frac{\widehat{\sigma}_l^2/\left(\widehat{f}_l^I\right)^2\frac{\gamma_{I-l,l}^2}{\delta_{I-l,l}}}{\left(\beta_l^{I+1}\right)^2}\right]\\
&- C_{i,I-i}^2\cdot\prod_{l=I-i}^{J-1}\left(\widehat{f}_l^I\right)^2
\end{aligned}
\tag{4.7.5}
$$

Finally we obtain,

$$
\begin{aligned}
&Var\left(\widehat{C}_{i,J}^{I+1}|D_I\right)\\
&= C_{i,I-i}^2 \cdot \prod_{l=I-i}^{J-1}\left(\widehat{f}_l^I\right)^2 \cdot \left\{\left[1+\frac{\widehat{\sigma}_{I-i}^2/\left(\widehat{f}_{I-i}^I\right)^2}{\delta_{i,I-i}}\right] \cdot \prod_{l=I-i+1}^{J-1}\left[1+\frac{\widehat{\sigma}_l^2/\left(\widehat{f}_l^I\right)^2 \frac{\gamma_{I-l,l}^2}{\delta_{I-l,l}}}{\left(\beta_l^{I+1}\right)^2}\right]-1\right\}\\
&= \left(\widehat{C}_{i,J}\right)^2 \cdot \left\{\left[1+\frac{\widehat{\sigma}_{I-i}^2/\left(\widehat{f}_{I-i}^I\right)^2}{\delta_{i,I-i}}\right] \cdot \prod_{l=I-i+1}^{J-1}\left[1+\frac{\widehat{\sigma}_l^2/\left(\widehat{f}_l^I\right)^2 \frac{\gamma_{I-l,l}^2}{\delta_{I-l,l}}}{\left(\beta_l^{I+1}\right)^2}\right]-1\right\}
\end{aligned}
$$

The second equality in (4.7.5) is a consequence of conditional independence assumption (see Lemma 4.7.1 (a)).
The third and fourth equality is due to Lemma 4.7.1 (d) and Lemma 4.7.4 (a),(b). The remaining equalities are provided by replacing the unknown parameters (σ_k^2, f_k) by their estimates $(\widehat{\sigma}_k^2, \widehat{f}_k)$ and by the simple computations.

Proof of (4.7.4).
The left hand side of (4.7.4) can be estimated by the following term (see p.128 of Merz and Wüthrich (2007) and (5.15), p.23, of Merz et al (2008))

$$\widehat{E}\left[\left(E[\widehat{CDR}_i(I+1)|D_I]\right)^2|D_I\right].$$

Set $\xi_i = E[\widehat{CDR}_i(I+1)|D_I]$. We have

$$
\begin{aligned}
\left[\widehat{E}\left(\widehat{CDR}_i(I+1)|D_I\right)\right]^2 &= \widehat{E}\left[\left(E[\widehat{CDR}_i(I+1)|D_I]\right)^2|D_I\right]\\
&= \widehat{E}\left[\xi_i^2|D_I\right].
\end{aligned}
\tag{4.7.6}
$$

From Lemma 4.7.2 we may write

$$
\begin{aligned}
\xi_i^2 &= \left(E[\widehat{CDR}_i(I+1)|D_I]\right)^2 \\
&= C_{i,I-i}^2 \left(\prod_{l=I-i}^{J-1} \widehat{f}_l^I - f_{I-i}^I \cdot \prod_{l=I-i+1}^{J-1} \left(\frac{\beta_l^I}{\beta_l^{I+1}} \widehat{f}_l^I + f_l \cdot \frac{\gamma_{I-l,l}}{\beta_l^{I+1}} \right) \right)^2 \\
&= C_{i,I-i}^2 \left(\prod_{l=I-i}^{J-1} \left(\widehat{f}_l^I\right)^2 + f_{I-i}^2 \cdot \prod_{l=I-i+1}^{J-1} \left(\frac{\beta_l^I}{\beta_l^{I+1}} \widehat{f}_l^I + f_l \cdot \frac{\gamma_{I-l,l}}{\beta_l^{I+1}} \right)^2 \right) \\
&- C_{i,I-i}^2 \left(2 \cdot \prod_{l=I-i}^{J-1} \widehat{f}_l^I \cdot f_{I-i} \cdot \prod_{l=I-i+1}^{J-1} \left(\frac{\beta_l^I}{\beta_l^{I+1}} \widehat{f}_l^I + f_l \cdot \frac{\gamma_{I-l,l}}{\beta_l^{I+1}} \right) \right)
\end{aligned}
\tag{4.7.7}
$$

Set $\omega_j := \sum_{i=0}^{I-j-1} \frac{\gamma_{i,j}^2}{\delta_{i,j}}$. Furthermore, from (4.7.7) and Lemma 4.7.3 (a)-(d) we obtain

$$
\begin{aligned}
&\widehat{E}\left[\xi_i^2|D_I\right] \\
&= C_{i,I-i}^2\left(\prod_{l=I-i}^{J-1}\left(Var\left(\widehat{f}_l^I|D_I\right)+f_l^2\right)+f_{I-i}^2\cdot\prod_{l=I-i+1}^{J-1}\left(\alpha_l^2\cdot Var\left(\widehat{f}_l^I|D_I\right)+f_l^2\right)\right) \\
&- 2\cdot C_{i,I-i}^2\left(\prod_{l=I-i+1}^{J-1}\left(\alpha_l\cdot Var\left(\widehat{f}_l^I|D_I\right)+f_l^2\right)\right) \\
&= C_{i,I-i}^2\left(\prod_{l=I-i}^{J-1}\left(\frac{\sigma_l^2}{(\beta_l^I)^2}\cdot\omega_l+f_l^2\right)+f_{I-i}^2\cdot\prod_{l=I-i+1}^{J-1}\left(\alpha_l^2\cdot\frac{\sigma_l^2}{(\beta_l^I)^2}\cdot\omega_l+f_l^2\right)\right) \\
&- 2\cdot C_{i,I-i}^2\left(f_{I-i}^2\cdot\prod_{l=I-i+1}^{J-1}\left(\alpha_l\cdot\frac{\sigma_l^2}{(\beta_l^I)^2}\cdot\omega_l+f_l^2\right)\right) \\
&= C_{i,I-i}^2\left(\prod_{l=I-i}^{J-1}f_l^2\cdot\prod_{l=I-i}^{J-1}\left(\frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2}\cdot\omega_l+1\right)\right) \\
&+ C_{i,I-i}^2\left(f_{I-i}^2\cdot\prod_{l=I-i+1}^{J-1}f_l^2\cdot\prod_{l=I-i+1}^{J-1}\left(\alpha_l^2\cdot\frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2}\cdot\omega_l+1\right)\right) \\
&- 2\cdot C_{i,I-i}^2\left(f_{I-i}^2\cdot\prod_{l=I-i+1}^{J-1}f_l^2\cdot\prod_{l=I-i+1}^{J-1}\left(\alpha_l\cdot\frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2}\cdot\omega_l+1\right)\right) \\
&= C_{i,I-i}^2\cdot\prod_{l=I-i}^{J-1}f_l^2\left(\prod_{l=I-i}^{J-1}\left(\frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2}\cdot\omega_l+1\right)+\prod_{l=I-i+1}^{J-1}\left(\alpha_l^2\cdot\frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2}\cdot\omega_l+1\right)\right) \\
&- 2C_{i,I-i}^2\cdot\prod_{l=I-i}^{J-1}f_l^2\left(\prod_{l=I-i+1}^{J-1}\left(\alpha_l\cdot\frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2}\cdot\omega_l+1\right)\right)
\end{aligned}
\tag{4.7.8}
$$

Finally, we plug in the estimators of $(\sigma_k^2,\ f_k)$ in the above equation and we obtain the claim (4.7.4).

4.7.3 Proof of Result 4.5.2

The approximating formulas for $\widehat{\Gamma}_{i,J}^{I}$ and $\widehat{\Delta}_{i,J}^{I}$ are obtained by using the following approximation (see Merz and Wüthrich (2008a), p.19)

$$\prod_{k=1}^{m}(1+x_k) \cong 1+\sum_{k=1}^{m} x_k, \text{ for } x_k \text{ small.} \tag{4.7.9}$$

Concerning the term for $\widehat{\Gamma}_{i,J}^{I}$ (see (4.5.9)) it is a straightforward application of approximation (4.7.9) to (4.7.5).
Concerning the term for $\widehat{\Delta}_{i,J}^{I}$, from the proof of Result 4.5.1 we have

$$\begin{aligned}
&\widehat{E}\left[\xi_i^2|D_I\right] \\
&= C_{i,I-i}^2 \cdot \prod_{l=I-i}^{J-1} f_l^2 \left(1+\sum_{l=I-i}^{J-1} \frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2}\cdot\omega_l + 1 + \sum_{l=I-i+1}^{J-1} \alpha_l^2 \cdot \frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2}\cdot\omega_l - 2\right) \\
&+ C_{i,I-i}^2 \cdot \prod_{l=I-i}^{J-1} f_l^2 \left(2\sum_{l=I-i+1}^{J-1} \alpha_l \cdot \frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2}\cdot\omega_l\right) \\
&= C_{i,I-i}^2 \cdot \prod_{l=I-i}^{J-1} f_l^2 \left(\frac{\sigma_{I-i}^2/f_{I-i}^2}{(\beta_{I-i}^I)^2}\cdot\omega_{I-i} + \sum_{l=I-i+1}^{J-1} \frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2}\cdot\omega_l + \sum_{l=I-i+1}^{J-1} \alpha_l^2 \cdot \frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2}\cdot\omega_l\right) \\
&- C_{i,I-i}^2 \cdot \prod_{l=I-i}^{J-1} f_l^2 \left(2\sum_{l=I-i+1}^{J-1} \alpha_l \cdot \frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2}\cdot\omega_l\right) \\
&= C_{i,I-i}^2 \cdot \prod_{l=I-i}^{J-1} f_l^2 \left(\frac{\sigma_{I-i}^2/f_{I-i}^2}{(\beta_{I-i}^I)^2}\cdot\omega_{I-i} + \sum_{l=I-i+1}^{J-1} (1-\alpha_l)^2 \frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2}\cdot\omega_l\right) \\
&= C_{i,I-i}^2 \cdot \prod_{l=I-i}^{J-1} f_l^2 \left(\frac{\sigma_{I-i}^2/f_{I-i}^2}{(\beta_{I-i}^I)^2}\cdot\omega_{I-i} + \sum_{l=I-i+1}^{J-1} \left(\frac{\gamma_{I-l,l}}{\beta_l^{I+1}}\right)^2 \frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2}\cdot\omega_l\right)
\end{aligned} \tag{4.7.10}$$

where the last equality is due to the fact,

$$(1-\alpha_l)^2 = \left(\frac{\gamma_{I-l,l}}{\beta_l^{I+1}}\right)^2. \tag{4.7.11}$$

By replacing the unknown parameters (σ_k^2, f_k) in (4.7.10) by their estimators we get (4.5.9).

4.7.4 Proof of Result 4.5.3

Let denote $\widehat{X}_l := \widehat{CDR}_l(I+1)$. Then we have

$$\begin{aligned} msep_{(\widehat{X}_i+\widehat{X}_k|D_I)}(0) &:= E\left[\left(\widehat{X}_i+\widehat{X}_k-0\right)^2|D_I\right] \\ &= E\left[\left(\widehat{X}_i-0\right)^2|D_I\right]+E\left[\left(\widehat{X}_k-0\right)^2|D_I\right]+2\cdot E\left[\widehat{X}_i\cdot\widehat{X}_k|D_I\right] \\ &= msep_{\widehat{X}_i|D_I}(0)+msep_{\widehat{X}_k|D_I}(0)+2\cdot Cov\left[\widehat{X}_i,\widehat{X}_k|D_I\right] \\ &+2\cdot E\left[\widehat{X}_i|D_I\right]\cdot E\left[\widehat{X}_k|D_I\right] \end{aligned} \tag{4.7.12}$$

Let set $V_{i,k} := Cov\left[\widehat{X}_i,\widehat{X}_k|D_I\right]$ and $W_{i,k} := E\left[\widehat{X}_i|D_I\right]\cdot E\left[\widehat{X}_k|D_I\right]$.
From (4.7.12) to prove (4.5.11) it is enough to show

$$\widehat{V}_{i,k} = \widehat{C}^I_{i,J}\cdot\widehat{C}^I_{k,J}\cdot\widehat{\Upsilon}^I_{i,J} \tag{4.7.13}$$

and

$$\widehat{W}_{i,k} = \widehat{C}^I_{i,J}\cdot\widehat{C}^I_{k,J}\cdot\widehat{\Lambda}^I_{i,J} \tag{4.7.14}$$

Proof of (4.7.13).
Recall from (4.3.4) that $\widehat{X}_i = \widehat{C}^I_{i,J}-\widehat{C}^{I+1}_{i,J}$. Hence
$V_{i,k} := Cov\left[\widehat{X}_i,\widehat{X}_k|D_I\right] = Cov\left[\widehat{C}^I_{i,J}-\widehat{C}^{I+1}_{i,J},\widehat{C}^I_{k,J}-\widehat{C}^{I+1}_{k,J}|D_I\right] = Cov\left[\widehat{C}^{I+1}_{i,J},\widehat{C}^{I+1}_{k,J}|D_I\right]$.
Furthermore

$$Cov\left[\widehat{C}^{I+1}_{i,J},\widehat{C}^{I+1}_{k,J}|D_I\right] := E\left[\widehat{C}^{I+1}_{i,J}\cdot\widehat{C}^{I+1}_{k,J}|D_I\right]-E\left[\widehat{C}^{I+1}_{i,J}|D_I\right]\cdot E\left[\widehat{C}^{I+1}_{k,J}|D_I\right].$$

- Term $E\left[\widehat{C}^{I+1}_{i,J}\cdot\widehat{C}^{I+1}_{k,J}|D_I\right]$:
  Recall from (4.3.6) that we can write for $k>i$

$$\widehat{C}^{I+1}_{i,J}\cdot\widehat{C}^{I+1}_{k,J} = \left(C_{k,I-k+1}\cdot\prod_{j=I-k+1}^{I-i-1}\widehat{f}^{I+1}_j\cdot\widehat{f}^{I+1}_{I-i}\cdot\prod_{j=I-i+1}^{J-1}\widehat{f}^{I+1}_j\right)\cdot C_{i,I-i+1}\cdot\prod_{j=I-i+1}^{J-1}\widehat{f}^{I+1}_j. \tag{4.7.15}$$

Hence from (4.7.15) and Lemma 4.7.1 (a) we obtain

$$E\left[\widehat{C}_{i,J}^{I+1}\cdot\widehat{C}_{k,J}^{I+1}|D_I\right]=E\left[C_{k,I-k+1}|D_I\right]\cdot\prod_{j=I-k+1}^{I-i-1}E\left[\widehat{f}_j^{I+1}|D_I\right]\cdot E\left[C_{i,I-i+1}\cdot\widehat{f}_{I-i}^{I+1}|D_I\right]\cdot\prod_{j=I-i+1}^{J-1}E\left[\left(\widehat{f}_j^{I+1}\right)^2|D_I\right]. \tag{4.7.16}$$

- Term $E\left[\widehat{C}_{i,J}^{I+1}|D_I\right]\cdot E\left[\widehat{C}_{k,J}^{I+1}|D_I\right]$.
  By using (4.3.6) we obtain

$$\begin{aligned}
&E\left[\widehat{C}_{i,J}^{I+1}|D_I\right]\cdot E\left[\widehat{C}_{k,J}^{I+1}|D_I\right]\\
&=E\left[C_{k,I-k+1}|D_I\right]\cdot\prod_{j=I-k+1}^{J-1}E\left[\widehat{f}_j^{I+1}|D_I\right]\\
&\times E\left[C_{i,I-i+1}|D_I\right]\cdot\prod_{j=I-i+1}^{J-1}E\left[\widehat{f}_j^{I+1}|D_I\right]\\
&=E\left[C_{k,I-k+1}|D_I\right]\cdot\prod_{j=I-k+1}^{I-i-1}E\left[\widehat{f}_j^{I+1}|D_I\right]\cdot E\left[\widehat{f}_{I-i}^{I+1}|D_I\right]\cdot\prod_{j=I-i+1}^{J-1}E\left[\widehat{f}_j^{I+1}|D_I\right]\\
&\times E\left[C_{i,I-i+1}|D_I\right]\cdot\prod_{j=I-i+1}^{J-1}E\left[\widehat{f}_j^{I+1}|D_I\right]\\
&=E\left[C_{k,I-k+1}|D_I\right]\cdot\prod_{j=I-k+1}^{I-i-1}E\left[\widehat{f}_j^{I+1}|D_I\right]\cdot E\left[\widehat{f}_{I-i}^{I+1}|D_I\right]\\
&\times E\left[C_{i,I-i+1}|D_I\right]\cdot\prod_{j=I-i+1}^{J-1}\left(E\left[\widehat{f}_j^{I+1}|D_I\right]\right)^2
\end{aligned} \tag{4.7.17}$$

Hence from (4.7.15) and (4.7.17)

$$\begin{aligned} V_{i,k} &:= Cov\left[\widehat{C}_{i,J}^{I+1}, \widehat{C}_{k,J}^{I+1}|D_I\right] \\ &= E\left[C_{k,I-k+1}|D_I\right] \cdot \prod_{j=I-k+1}^{I-i-1} E\left[\widehat{f}_j^{I+1}|D_I\right] \cdot \Bigg[E\left[C_{i,I-i+1} \cdot \widehat{f}_{I-i}^{I+1}|D_I\right] \\ &\times \prod_{j=I-i+1}^{J-1} E\left[\left(\widehat{f}_j^{I+1}\right)^2|D_I\right] \\ &- E\left[\widehat{f}_{I-i}^{I+1}|D_I\right] \cdot E\left[C_{i,I-i+1}|D_I\right] \cdot \prod_{j=I-i+1}^{J-1}\left(E\left[\widehat{f}_j^{I+1}|D_I\right]\right)^2\Bigg] \end{aligned} \quad (4.7.18)$$

We apply Lemma 4.7.4 in (4.7.18),

$$
\begin{aligned}
\widehat{V}_{i,k} &:= \widehat{Cov}\left[\widehat{C}_{i,J}^{I+1}, \widehat{C}_{k,J}^{I+1}|D_I\right] \\
&= E\left[C_{k,I-k+1}|D_I\right] \cdot \prod_{j=I-k+1}^{I-i-1} E\left[\widehat{f}_j^{I+1}|D_I\right] \\
&\times \left[E\left[C_{i,I-i+1}\cdot \widehat{f}_{I-i}^{I+1}|D_I\right] \cdot \prod_{j=I-i+1}^{J-1} E\left[\left(\widehat{f}_j^{I+1}\right)^2|D_I\right]\right. \\
&\left. - E\left[\widehat{f}_{I-i}^{I+1}|D_I\right]\cdot E\left[C_{i,I-i+1}|D_I\right]\cdot \prod_{j=I-i+1}^{J-1}\left(E\left[\widehat{f}_j^{I+1}|D_I\right]\right)^2\right] \\
&= C_{k,I-k}\cdot \widehat{f}_{I-k}^I \cdot \prod_{j=I-k+1}^{I-i-1} \widehat{f}_j^I \left[\left(\widehat{f}_{I-i}^I\right)^2 \cdot C_{i,I-i}\cdot \left(1+\frac{\left(\widehat{\sigma}_{I-i}^I\right)^2/\left(\widehat{f}_{I-i}^I\right)^2}{\beta_{I-i}^{I+1}}\cdot\frac{\gamma_{i,I-i}}{\delta_{i,I-i}}\right)\right. \\
&\times \prod_{j=I-i+1}^{J-1}\left(\widehat{f}_j^I\right)^2 \cdot \left(1+\frac{\left(\widehat{\sigma}_j^I\right)^2/\left(\widehat{f}_j^I\right)^2}{\left(\beta_j^{I+1}\right)^2}\cdot\frac{\gamma_{I-j,j}^2}{\delta_{I-j,j}}\right) \\
&\left. - \widehat{f}_{I-i}^I \cdot C_{i,I-i}\cdot \widehat{f}_{I-i}^I \cdot \prod_{j=I-i+1}^{J-1}\left(\widehat{f}_j^I\right)^2\right] \\
&= \widehat{C}_{k,I-i}\left[C_{i,I-i}\cdot\left(\widehat{f}_{I-i}^I\right)^2\cdot\left(1+\frac{\left(\widehat{\sigma}_{I-i}^I\right)^2/\left(\widehat{f}_{I-i}^I\right)^2}{\beta_{I-i}^{I+1}}\cdot\frac{\gamma_{i,I-i}}{\delta_{i,I-i}}\right)\cdot\prod_{j=I-i+1}^{J-1}\left(\widehat{f}_j^I\right)^2\right. \\
&\left.\times\left(1+\frac{\left(\widehat{\sigma}_j^I\right)^2/\left(\widehat{f}_j^I\right)^2}{\left(\beta_j^{I+1}\right)^2}\cdot\frac{\gamma_{I-j,j}^2}{\delta_{I-j,j}}\right) - C_{i,I-i}\cdot\left(\widehat{f}_{I-i}^I\right)^2\cdot\prod_{j=I-i+1}^{J-1}\left(\widehat{f}_j^I\right)^2\right] \\
&= \widehat{C}_{k,I-i}\cdot C_{i,I-i}\cdot\left(\widehat{f}_{I-i}^I\right)^2\cdot\prod_{j=I-i+1}^{J-1}\left(\widehat{f}_j^I\right)^2\left[\left(1+\frac{\left(\widehat{\sigma}_{I-i}^I\right)^2/\left(\widehat{f}_{I-i}^I\right)^2}{\beta_{I-i}^{I+1}}\cdot\frac{\gamma_{i,I-i}}{\delta_{i,I-i}}\right)\right. \\
&\left.\times\prod_{j=I-i+1}^{J-1}\left(1+\frac{\left(\widehat{\sigma}_j^I\right)^2/\left(\widehat{f}_j^I\right)^2}{\left(\beta_j^{I+1}\right)^2}\cdot\frac{\gamma_{I-j,j}^2}{\delta_{I-j,j}}\right)-1\right] \\
&= \widehat{C}_{k,J}\cdot\widehat{C}_{i,J}\left(1+\frac{\left(\widehat{\sigma}_{I-i}^I\right)^2/\left(\widehat{f}_{I-i}^I\right)^2}{\beta_{I-i}^{I+1}}\cdot\frac{\gamma_{i,I-i}}{\delta_{i,I-i}}\right) \\
&\times\prod_{j=I-i+1}^{J-1}\left(1+\frac{\left(\widehat{\sigma}_j^I\right)^2/\left(\widehat{f}_j^I\right)^2}{\left(\beta_j^{I+1}\right)^2}\cdot\frac{\gamma_{I-j,j}^2}{\delta_{I-j,j}}\right)-\widehat{C}_{k,J}\cdot\widehat{C}_{i,J}.
\end{aligned}
\tag{4.7.19}
$$

Proof of (4.7.14). Recall that

$$\delta_l = \frac{\beta_l^I}{\beta_l^{I+1}} \widehat{f}_l^I + f_l \cdot \frac{\gamma_{I-l,l}}{\beta_l^{I+1}}.$$

Let define

$$A_i := \prod_{l=I-i}^{J-1} \widehat{f}_l^I - f_{I-i}^I \cdot \prod_{l=I-i+1}^{J-1} \delta_l. \tag{4.7.20}$$

Hence

$$A_i \cdot A_k = \left(\prod_{l=I-i}^{J-1} \widehat{f}_l^I - f_{I-i}^I \cdot \prod_{l=I-i+1}^{J-1} \delta_l \right) \cdot \left(\prod_{l=I-k}^{J-1} \widehat{f}_l^I - f_{I-k}^I \cdot \prod_{l=I-k+1}^{J-1} \delta_l \right) \tag{4.7.21}$$

$$\begin{aligned} A_i \cdot A_k = & \prod_{l=I-i}^{J-1} \widehat{f}_l^I \cdot \prod_{l=I-k}^{J-1} \widehat{f}_l^I - \prod_{l=I-i}^{J-1} \widehat{f}_l^I \cdot f_{I-k} \cdot \prod_{l=I-k+1}^{J-1} \delta_l - f_{I-i} \cdot \prod_{l=I-i+1}^{J-1} \delta_l \\ & \times \prod_{l=I-k}^{J-1} \widehat{f}_l^I + f_{I-i} \cdot f_{I-k} \cdot \prod_{l=I-i+1}^{J-1} \delta_l \cdot \prod_{l=I-k+1}^{J-1} \delta_l \\ = & \prod_{l=I-i}^{J-1} \widehat{f}_l^I \cdot \prod_{l=I-k}^{I-i-1} \widehat{f}_l^I \cdot \prod_{l=I-i}^{J-1} \widehat{f}_l^I - \prod_{l=I-i}^{J-1} \widehat{f}_l^I \cdot f_{I-k} \cdot \prod_{l=I-k+1}^{I-i-1} \delta_l \cdot \prod_{l=I-i}^{J-1} \delta_l \\ & - f_{I-i} \cdot \prod_{l=I-i+1}^{J-1} \delta_l \cdot \prod_{l=I-k}^{I-i-1} \widehat{f}_l^I \cdot \prod_{l=I-i}^{J-1} \widehat{f}_l^I \\ & + f_{I-i} \cdot f_{I-k} \cdot \prod_{l=I-i+1}^{J-1} \delta_l \cdot \prod_{l=I-k+1}^{I-i} \delta_l \cdot \prod_{l=I-i+1}^{J-1} \delta_l \\ = & \prod_{l=I-i}^{J-1} \left(\widehat{f}_l^I \right)^2 \cdot \prod_{l=I-k}^{I-i-1} \widehat{f}_l^I - f_{I-k}^I \cdot \prod_{l=I-k+1}^{I-i-1} \delta_l \cdot \prod_{l=I-i}^{J-1} \widehat{f}_l^I \cdot \delta_l \\ & - f_{I-i}^I \cdot \prod_{l=I-i+1}^{J-1} \widehat{f}_l^I \cdot \delta_l \cdot \widehat{f}_{I-i}^I \cdot \prod_{l=I-k}^{I-i-1} \widehat{f}_l^I + f_{I-i}^I \cdot f_{I-k}^I \cdot \prod_{l=I-i+1}^{J-1} \delta_l^2 \cdot \prod_{l=I-k+1}^{I-i} \delta_l \end{aligned} \tag{4.7.22}$$

Furthermore, from (4.7.22) and since $E\left[\delta_l|D_I\right] = f_l$ and $E\left[\widehat{f}_l^I|D_I\right] = f_l$,

$$\begin{aligned}
E\left[A_i \cdot A_k|D_I\right] &= \prod_{l=I-i}^{J-1} E\left[\left(\widehat{f}_l^I\right)^2|D_I\right] \cdot \prod_{l=I-k}^{I-i-1} E\left[\widehat{f}_l^I|D_I\right] \\
&- f_{I-k}^I \cdot \prod_{l=I-k+1}^{I-i-1} E\left[\delta_l|D_I\right] \cdot \prod_{l=I-i}^{J-1} E\left[\widehat{f}_l^I \cdot \delta_l|D_I\right] \\
&- f_{I-i}^I \cdot \prod_{l=I-i+1}^{J-1} E\left[\widehat{f}_l^I \cdot \delta_l|D_I\right] \cdot E\left[\widehat{f}_{I-i}^I|D_I\right] \cdot \prod_{l=I-k}^{I-i-1} E\left[\widehat{f}_l^I|D_I\right] \\
&+ f_{I-i}^I \cdot f_{I-k}^I \cdot \prod_{l=I-i+1}^{J-1} E\left[\delta_l^2|D_I\right] \cdot \prod_{l=I-k+1}^{I-i} E\left[\delta_l|D_I\right] \\
&= \prod_{l=I-i}^{J-1} E\left[\left(\widehat{f}_l^I\right)^2|D_I\right] \cdot \prod_{l=I-k}^{I-i-1} f_l - f_{I-k}^I \cdot \prod_{l=I-k+1}^{I-i-1} f_l \cdot \prod_{l=I-i}^{J-1} E\left[\widehat{f}_l^I \cdot \delta_l|D_I\right] \\
&- f_{I-i}^I \cdot \prod_{l=I-i+1}^{J-1} E\left[\widehat{f}_l^I \cdot \delta_l|D_I\right] \cdot f_{I-i} \cdot \prod_{l=I-k}^{I-i-1} f_l + f_{I-i}^I \cdot f_{I-k}^I \\
&\times \prod_{l=I-i+1}^{J-1} E\left[\delta_l^2|D_I\right] \cdot \prod_{l=I-k+1}^{I-i} f_l \\
&= \prod_{l=I-k}^{I-i-1} f_l \cdot \left(\prod_{l=I-i}^{J-1} E\left[\left(\widehat{f}_l^I\right)^2|D_I\right] - \prod_{l=I-i}^{J-1} E\left[\widehat{f}_l^I \cdot \delta_l|D_I\right]\right) \\
&- \prod_{l=I-k}^{I-i-1} f_l \cdot \left(\left(f_{I-i}^I\right)^2 \cdot \prod_{l=I-i+1}^{J-1} E\left[\widehat{f}_l^I \cdot \delta_l|D_I\right] + \left(f_{I-i}^I\right)^2 \cdot \prod_{l=I-i+1}^{J-1} E\left[\delta_l^2|D_I\right]\right)
\end{aligned} \tag{4.7.23}$$

Recall $\omega_j = \sum_{i=0}^{I-j-1} \frac{\gamma_{i,j}^2}{\delta_{i,j}}$. From Lemma 4.7.5 (a)-(c) we obtain

$$
\begin{aligned}
&E\left[A_i \cdot A_k | D_I\right] \\
&= \prod_{l=I-k}^{I-i-1} f_l \cdot \left[\prod_{l=I-i}^{J-1} \left(\frac{\sigma_l^2}{(\beta_l^I)^2} \cdot \omega_l + f_l^2 \right) - \prod_{l=I-i}^{J-1} \left(\frac{\beta_l^I}{\beta_l^{I+1}} \cdot \frac{\sigma_l^2}{(\beta_l^I)^2} \cdot \omega_l + f_l^2 \right) \right. \\
&\left. - f_{I-i}^2 \cdot \prod_{l=I-i+1}^{J-1} \left(\frac{\beta_l^I}{\beta_l^{I+1}} \cdot \frac{\sigma_l^2}{(\beta_l^I)^2} \cdot \omega_l + f_l^2 \right) + f_{I-i}^2 \cdot \prod_{l=I-i+1}^{J-1} \left(\left(\frac{\beta_l^I}{\beta_l^{I+1}} \right)^2 \cdot \frac{\sigma_l^2}{(\beta_l^I)^2} \cdot \omega_l + f_l^2 \right) \right] \\
&= \prod_{l=I-k}^{I-i-1} f_l \cdot \left[\prod_{l=I-i}^{J-1} f_l^2 \cdot \prod_{l=I-i}^{J-1} \left(\frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2} \cdot \omega_l + 1 \right) - \prod_{l=I-i}^{J-1} f_l^2 \cdot \prod_{l=I-i}^{J-1} \left(\frac{\beta_l^I}{\beta_l^{I+1}} \cdot \frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2} \cdot \omega_l + 1 \right) \right. \\
&- f_{I-i}^2 \cdot \prod_{l=I-i+1}^{J-1} f_l^2 \cdot \prod_{l=I-i+1}^{J-1} \left(\frac{\beta_l^I}{\beta_l^{I+1}} \cdot \frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2} \cdot \omega_l + 1 \right) \\
&\left. + f_{I-i}^2 \cdot \prod_{l=I-i+1}^{J-1} f_l^2 \prod_{l=I-i+1}^{J-1} \left(\left(\frac{\beta_l^I}{\beta_l^{I+1}} \right)^2 \cdot \frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2} \cdot \omega_l + 1 \right) \right] \\
&= \prod_{l=I-k}^{I-i-1} f_l \cdot \prod_{l=I-i}^{J-1} f_l^2 \cdot \left[\prod_{l=I-i}^{J-1} \left(\frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2} \cdot \omega_l + 1 \right) - \prod_{l=I-i}^{J-1} \left(\frac{\beta_l^I}{\beta_l^{I+1}} \cdot \frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2} \cdot \omega_l + 1 \right) \right. \\
&\left. - \prod_{l=I-i+1}^{J-1} \left(\frac{\beta_l^I}{\beta_l^{I+1}} \cdot \frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2} \cdot \omega_l + 1 \right) + \prod_{l=I-i+1}^{J-1} \left(\left(\frac{\beta_l^I}{\beta_l^{I+1}} \right)^2 \cdot \frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2} \cdot \omega_l + 1 \right) \right]
\end{aligned}
\tag{4.7.24}
$$

Replacing the unknown parameters $(\sigma_k^2,\ f_k)$ by their estimators $(\widehat{\sigma}_k^2,\ \widehat{f}_k)$ completes the proof of this claim.

4.7.5 Proof of Result 4.5.4

Regarding to the term $\widehat{\Upsilon}_{i,J}^I$, we apply the approximation from (4.7.9) to equation (4.7.19). This finishes the proof of (4.5.15).

For the term $\widehat{\Phi}_{i,J}^I$, we use the approximation (4.7.9) applied to the last term of (4.7.24),

$$\begin{aligned}
&E\left[A_i \cdot A_k | D_I\right] \\
&= \prod_{l=I-k}^{I-i-1} f_l \cdot \prod_{l=I-i}^{J-1} f_l^2 \cdot \left[\sum_{l=I-i}^{J-1} \frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2} \cdot \omega_l + 1 - \sum_{l=I-i}^{J-1} \frac{\beta_l^I}{\beta_l^{I+1}} \cdot \frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2} \cdot \omega_l - 1 \right. \\
&\left. - \sum_{l=I-i+1}^{J-1} \frac{\beta_l^I}{\beta_l^{I+1}} \cdot \frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2} \cdot \omega_l - 1 + \sum_{l=I-i+1}^{J-1} \left(\frac{\beta_l^I}{\beta_l^{I+1}}\right)^2 \cdot \frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2} \cdot \omega_l + 1 \right] \\
&= \prod_{l=I-k}^{I-i-1} f_l \cdot \prod_{l=I-i}^{J-1} f_l^2 \cdot \left[\frac{\sigma_{I-i}^2/f_{I-i}^2}{(\beta_{I-i}^I)^2} \cdot \omega_{I-i} - \frac{\beta_{I-i}^I}{\beta_{I-i}^{I+1}} \cdot \frac{\sigma_{I-i}^2/f_{I-i}^2}{(\beta_{I-i}^I)^2} \cdot \omega_{I-i} + \sum_{l=I-i+1}^{J-1} \frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2} \cdot \omega_l \right. \\
&\left. - 2 \cdot \sum_{l=I-i+1}^{J-1} \frac{\beta_l^I}{\beta_l^{I+1}} \cdot \frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2} \cdot \omega_l + \sum_{l=I-i+1}^{J-1} \left(\frac{\beta_l^I}{\beta_l^{I+1}}\right)^2 \cdot \frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2} \cdot \omega_l \right] \\
&= \prod_{l=I-k}^{I-i-1} f_l \cdot \prod_{l=I-i}^{J-1} f_l^2 \cdot \left[\frac{\sigma_{I-i}^2/f_{I-i}^2}{(\beta_{I-i}^I)^2} \cdot \omega_{I-i} \left(1 - \frac{\beta_{I-i}^I}{\beta_{I-i}^{I+1}}\right) + \sum_{l=I-i+1}^{J-1} \frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2} \cdot \omega_l \left(1 - \frac{\beta_l^I}{\beta_l^{I+1}}\right)^2 \right]
\end{aligned} \tag{4.7.25}$$

Since $\left(1 - \frac{\beta_l^I}{\beta_l^{I+1}}\right) = \frac{\gamma_{I-l,l}}{\beta_l^{I+1}}$ from (4.7.25) we conclude

$$E\left[A_i \cdot A_k | D_I\right] = \prod_{l=I-k}^{I-i-1} f_l \cdot \prod_{l=I-i}^{J-1} f_l^2 \cdot \left[\frac{\gamma_{i,I-i}}{\beta_l^{I+1}} \cdot \frac{\sigma_{I-i}^2/f_{I-i}^2}{(\beta_{I-i}^I)^2} \cdot \omega_{I-i} + \sum_{l=I-i+1}^{J-1} \frac{\sigma_l^2/f_l^2}{(\beta_l^I)^2} \cdot \omega_l \left(\frac{\gamma_{I-l,l}}{\beta_l^{I+1}}\right)^2 \right] \tag{4.7.26}$$

We plug in the estimators $(\widehat{f}_k, \widehat{\sigma}_k^2)$ in (4.7.26),

$$\begin{aligned}
\widehat{E}\left[A_i \cdot A_k | D_I\right] = \prod_{l=I-k}^{I-i-1} \widehat{f}_l^I \cdot \prod_{l=I-i}^{J-1} \left(\widehat{f}_l^I\right)^2 \cdot &\left[\frac{\gamma_{i,I-i}}{\beta_l^{I+1}} \cdot \frac{\widehat{\sigma}_{I-i}^2/\left(\widehat{f}_{I-i}\right)^2}{(\beta_{I-i}^I)^2} \cdot \omega_{I-i} \right. \\
&\left. + \sum_{l=I-i+1}^{J-1} \frac{\widehat{\sigma}_l^2/\left(\widehat{f}_l^I\right)^2}{(\beta_l^I)^2} \cdot \omega_l \left(\frac{\gamma_{I-l,l}}{\beta_l^{I+1}}\right)^2 \right].
\end{aligned} \tag{4.7.27}$$

Furthermore,

$$\widehat{W}_{i,k} = C_{i,I-i} \cdot C_{k,I-k} \cdot \widehat{E}\left[A_i \cdot A_k | D_I\right]$$

$$= C_{i,I-i} \cdot C_{k,I-k} \cdot \prod_{l=I-k}^{I-i-1} \widehat{f}_l^I \cdot \prod_{l=I-i}^{J-1} \left(\widehat{f}_l^I\right)^2 \cdot \left[\frac{\gamma_{i,I-i}}{\beta_l^{I+1}} \cdot \frac{\widehat{\sigma}_{I-i}^2 / \left(\widehat{f}_{I-i}\right)^2}{(\beta_{I-i}^I)^2} \cdot \omega_{I-i} \right.$$

$$\left. + \sum_{l=I-i+1}^{J-1} \frac{\widehat{\sigma}_l{}^2 / \left(\widehat{f}_l^I\right)^2}{(\beta_l^I)^2} \cdot \omega_l \left(\frac{\gamma_{I-l,l}}{\beta_l^{I+1}}\right)^2 \right]$$

$$= \widehat{C}_{i,J}^I \cdot \widehat{C}_{k,J}^I \cdot \left[\frac{\gamma_{i,I-i}}{\beta_l^{I+1}} \cdot \frac{\widehat{\sigma}_{I-i}^2 / \left(\widehat{f}_{I-i}\right)^2}{(\beta_{I-i}^I)^2} \cdot \omega_{I-i} + \sum_{l=I-i+1}^{J-1} \frac{\widehat{\sigma}_l{}^2 / \left(\widehat{f}_l^I\right)^2}{(\beta_l^I)^2} \cdot \omega_l \left(\frac{\gamma_{I-l,l}}{\beta_l^{I+1}}\right)^2 \right]. \tag{4.7.28}$$

□

4.7.6 Proofs of useful Lemmas

Proof. (Lemma 4.7.1)
(a) For $l = 0, ..., J-1$, we have

$$\widehat{f}_l^{I+1} = \frac{\beta_l^I}{\beta_l^{I+1}} \widehat{f}_l^I + \frac{\frac{\gamma_{I-l,l}}{C_{I-l,l}} C_{I-l,l+1}}{\beta_l^{I+1}}, \tag{4.7.29}$$

and from (4.2.5)

$$C_{i,I-i+1} = f_{I-i} \cdot C_{i,I-i} + \frac{\sigma_{I-i}}{\sqrt{\delta_{i,I-i}}} \cdot C_{i,I-i} \cdot \varepsilon_{i,I-i+1}.$$

Hence $C_{i,I-i+1}, \widehat{f}_{I-i+1}^{I+1}, ..., \widehat{f}_{J-1}^{I+1}$, given D_I, are the functions of $\varepsilon_{i,I-i+1}, ..., \varepsilon_{I-J+1,J}$ which are independent by assumption (TM.2).
(b) Since $\widehat{f}_l$ is D_I measurable, from (4.7.29) we have,

$$E\left[\widehat{f}_l^{I+1} | D_I\right] = \frac{\beta_l^I}{\beta_l^{I+1}} \widehat{f}_l^I + \frac{\gamma_{I-l,l}}{C_{I-l,l}\beta_l^{I+1}} \cdot E[C_{I-l,l+1} | D_I].$$

Since $C_{I-l,l+1} = f_l \cdot C_{I-l,l} + \frac{\sigma_l}{\sqrt{\delta_{I-l,l}}} \cdot C_{I-l,l} \cdot \varepsilon_{I-l,l+1}$, we obtain

$$E\left[\widehat{f}_l^{I+1} | D_I\right] = \frac{\beta_l^I}{\beta_l^{I+1}} \widehat{f}_l^I + \frac{\gamma_{I-l,l}}{C_{I-l,l}\beta_l^{I+1}} (f_l \cdot C_{I-l,l} + \frac{\sigma_l}{\sqrt{\gamma_{I-l,l}}} \cdot C_{I-l,l} \cdot E\left[\varepsilon_{I-l,l+1} | D_I\right]).$$

There remains to use the assumption $E\left[\varepsilon_{I-l,l+1}|D_I\right]) = 0$.
(c) From (4.3.6) we have

$$\widehat{C}_{i,j}^{I+1} = C_{i,I-i+1} \prod_{l=I-i+1}^{j-1} \widehat{f}_l^{I+1}.$$

Since $E\left[C_{i,I-i+1}|D_I\right] = C_{i,I-i} \cdot f_{I-i}$ there remains to apply Lemma 4.7.1(a).
(d) From (4.2.5) we have

$$C_{i,I-i+1} = f_{I-i} \cdot C_{i,I-i} + \frac{\sigma_{I-i}}{\sqrt{\gamma_{i,I-i}}} \cdot C_{i,I-i} \cdot \varepsilon_{i,I-i+1}.$$

We compute $C_{i,I-i+1}^2$ and to prove the claim (d) there remains to apply the model assumptions, $E\left[\varepsilon_{i,I-i+1}|D_I\right] = 0$ and $E\left[\varepsilon_{i,I-i+1}^2|D_I\right] = 1$.
(e)
For $l = 0, ..., J-1$, we have from definition of $\widehat{f}_l^{I+1}$ (see (4.2.9)),

$$\widehat{f}_l^{I+1} = \frac{\sum_{i=0}^{I-l} \frac{\gamma_{i,l}}{C_{i,l}} C_{i,l+1}}{\beta_l^{I+1}} + \frac{\frac{\gamma_{I-l,l}}{C_{I-l,l}} C_{I-l,l+1}}{\beta_l^{I+1}}.$$

Since β_l^{I+1} is D_I measurable we obtain

$$\begin{aligned} E[(\widehat{f}_l^{I+1})^2|D_I] = & \left(\frac{\sum_{i=0}^{I-l} \frac{\gamma_{i,l}}{C_{i,l}} C_{i,l+1}}{\beta_l^{I+1}}\right)^2 + 2\frac{\sum_{i=0}^{I-l} \frac{\gamma_{i,l}}{C_{i,l}} C_{i,l+1}}{\beta_l^{I+1}} \cdot \frac{\frac{\gamma_{I-l,l}}{C_{I-l,l}}}{\beta_l^{I+1}} \cdot E[C_{I-l,l+1}|D_I] \\ & + \left(\frac{\frac{\gamma_{I-l,l}}{C_{I-l,l}}}{\beta_l^{I+1}}\right)^2 \cdot E[C_{I-l,l+1}^2|D_I]. \end{aligned}$$

Using $E\left[C_{i,I-i+1}|D_I\right] = f_{I-i} C_{i,I-i}$ and Lemma 4.7.1 (d) (with $i = I-l$) we obtain,

$$\begin{aligned} E[(\widehat{f}_l^{I+1})^2|D_I] = & \left(\frac{\sum_{i=0}^{I-l} \frac{\gamma_{i,l}}{C_{i,l}} C_{i,l+1}}{\beta_l^{I+1}}\right)^2 + 2\frac{\sum_{i=0}^{I-l} \frac{\gamma_{i,l}}{C_{i,l}} C_{i,l+1}}{\beta_l^{I+1}} \cdot \frac{\frac{\gamma_{I-l,l}}{C_{I-l,l}}}{\beta_l^{I+1}} \cdot f_l C_{I-l,l} \\ & + \left(\frac{\frac{\gamma_{I-l,l}}{C_{I-l,l}}}{\beta_l^{I+1}}\right)^2 \cdot (f_l^2 \cdot C_{I-l,l}^2 + \frac{\sigma_l^2}{\delta_{I-l,l}} \cdot C_{I-l,l}^2) \\ = & \left(\frac{\sum_{i=0}^{I-l-1} \frac{\gamma_{i,l}}{C_{i,l}} \cdot C_{i,l+1}}{\beta_l^{I+1}} + \frac{\gamma_{l,I-l} \cdot f_l}{\beta_l^{I+1}}\right)^2 + \frac{\sigma_l^2 \cdot \frac{\gamma_{I-l,l}^2}{\delta_{I-l,l}}}{\left(\beta_l^{I+1}\right)^2}. \end{aligned}$$

Therefore, we conclude claim (e).
(f) For $l = 0, ..., J-1$, we have (see (4.2.9))

$$\widehat{f}_l^{I+1} = \frac{\sum_{i=0}^{I-l} \frac{\gamma_{i,l}}{C_{i,l}} C_{i,l+1}}{\beta_l^{I+1}}.$$

Hence, for $l = I-i$,

$$\widehat{f}_{I-i}^{I+1} = \frac{\sum_{k=0}^{i-1} \frac{\gamma_{k,I-i}}{C_{k,I-i}} C_{k,I-i+1}}{\beta_{I-i}^{I+1}} + \frac{\frac{\gamma_{i,I-i}}{C_{i,I-i}} C_{i,I-i+1}}{\beta_{I-i}^{I+1}}.$$

We derive from Lemma 4.7.1 (d),

$$\begin{aligned}
&E\left[C_{i,I-i+1} \widehat{f}_{I-i}^{I+1} | D_I\right] \\
&= (\beta_{I-i}^{I+1})^{-1} \left[\sum_{k=0}^{i-1} \frac{\gamma_{k,I-i}}{C_{k,I-i}} C_{k,I-i+1} \cdot E\left[C_{i,I-i+1} | D_I\right] + \frac{\gamma_{i,I-i}}{C_{i,I-i}} E\left[C_{i,I-i+1}^2 | D_I\right]\right]. \\
&= (\beta_{I-i}^{I+1})^{-1} \left[\sum_{k=0}^{i-1} \frac{\gamma_{k,I-i}}{C_{k,I-i}} C_{k,I-i+1} \cdot f_{I-i} C_{i,I-i} + \frac{\gamma_{i,I-i}}{C_{i,I-i}} \left(f_{I-i}^2 \cdot C_{i,I-i}^2 + \frac{\sigma_{I-i}^2}{\delta_{i,I-i}} \cdot C_{i,I-i}^2\right)\right] \\
&= (\beta_{I-i}^{I+1})^{-1} \left[\left(\sum_{k=0}^{i-1} \frac{\gamma_{k,I-i}}{C_{k,I-i}} C_{k,I-i+1}\right) \cdot f_{I-i} C_{i,I-i} + \frac{\gamma_{i,I-i}}{C_{i,I-i}} \left(f_{I-i}^2 \cdot C_{i,I-i}^2 + \frac{\sigma_{I-i}^2}{\delta_{i,I-i}} \cdot C_{i,I-i}^2\right)\right] \\
&= (\beta_{I-i}^{I+1})^{-1} \left[\left(\sum_{k=0}^{i-1} \frac{\gamma_{k,I-i}}{C_{k,I-i}} C_{k,I-i+1}\right) \cdot f_{I-i} C_{i,I-i} + \gamma_{i,I-i} f_{I-i}^2 \cdot C_{i,I-i} + \frac{\sigma_{I-i}^2}{\delta_{i,I-i}} \cdot C_{i,I-i} \gamma_{i,I-i}\right].
\end{aligned}$$

□

Proof. (Lemma 4.7.2)
From Definition 4.3.2 of $\widehat{CDR}_i(I+1)$ we have

$$E[\widehat{CDR}_i(I+1)|D_I] = E[\widehat{C}_{i,J}^I - \widehat{C}_{i,J}^{I+1}|D_I] = \widehat{C}_{i,J}^I - E[\widehat{C}_{i,J}^{I+1}|D_I],$$

where the last equality is due to the fact that $\widehat{C}_{i,J}^I$ is D_I measurable. From $E\left[C_{i,I-i+1}|D_I\right] = f_{I-i} C_{i,I-i}$ and Lemma 4.7.1 (a) we obtain

$$E[\widehat{CDR}_i(I+1)|D_I] = C_{i,I-i}\prod_{j=I-i}^{J-1}\widehat{f}_j^I - E[C_{i,I-i+1}\prod_{j=I-i+1}^{J-1}\widehat{f}_j^{I+1}|D_I]$$

$$= C_{i,I-i}\prod_{j=I-i}^{J-1}\widehat{f}_j^I - E[C_{i,I-i+1}|D_I]\cdot\prod_{j=I-i+1}^{J-1}E[\widehat{f}_j^{I+1}|D_I]$$

$$= C_{i,I-i}\left(\prod_{j=I-i}^{J-1}\widehat{f}_j^I - f_{I-i}\prod_{j=I-i+1}^{J-1}E[\widehat{f}_j^{I+1}|D_I]\right).$$

The application of Lemma 4.7.1 (b) completes the proof of Lemma 4.7.2.
□

Proof. (Lemma 4.7.3)
Following the "resampling technique" from Merz and Wüthrich (2007) p.127, we may write,

$$\widehat{f}_j^I = f_j + \frac{\sigma_j}{\beta_j^I}\sum_{i=0}^{I-j-1}\frac{\gamma_{i,j}}{\sqrt{\delta_{i,j}}}\cdot\widetilde{\varepsilon}_{i,j+1}, \tag{4.7.30}$$

where $\varepsilon_{i,j}$, $\widetilde{\varepsilon}_{i,j}$ are independent and identically distributed.
(a) From (4.7.30) we obtain $E(\widehat{f}_l^I|D_I) = f_l$. Thus

$$E\left[\left(\widehat{f}_l^I\right)^2|D_I\right] = Var\left(\widehat{f}_l^I|D_I\right) + \left[E(\widehat{f}_l^I|D_I)\right]^2 = Var\left(\widehat{f}_l^I|D_I\right) + f_l^2.$$

(b) Recall $\alpha_l = \frac{\beta_l^I}{\beta_l^{I+1}}$. We have, $f_l\cdot\frac{\gamma_{I-l,l}}{\beta_l^{I+1}} = f_l - \alpha_l f_l$, It implies that $\alpha_l\widehat{f}_l^I + f_l\cdot\frac{\gamma_{I-l,l}}{\beta_l^{I+1}} = \alpha_l(\widehat{f}_l^I - f_l) + f_l$. Furthermore

$$E\left[\left(\alpha_l\widehat{f}_l^I + f_l\cdot\frac{\gamma_{I-l,l}}{\beta_l^{I+1}}\right)^2|D_I\right] = E\left[\left(\alpha_l(\widehat{f}_l^I - f_l) + f_l\right)^2|D_I\right] = \alpha_l^2E\left[\left(\widehat{f}_l^I - f_l\right)^2|D_I\right]$$

$$+ 2\alpha_lE\left[\left(\widehat{f}_l^I - f_l\right)|D_I\right] + f_l^2 = \alpha_l^2Var(\widehat{f}_l^I|D_I) + f_l^2.$$

(c) Let set

$$\eta_i = \prod_{l=I-i}^{J-1}\widehat{f}_l^I\cdot f_{I-i}\cdot\prod_{l=I-i+1}^{J-1}\left(\alpha_l\cdot\widehat{f}_l^I + f_l\cdot\frac{\gamma_{I-l,l}}{\beta_l^{I+1}}\right) = \widehat{f}_{I-i}^I\cdot f_{I-i}\prod_{l=I-i+1}^{J-1}\widehat{f}_l^I\left[\alpha_l(\widehat{f}_l^I - f_l) + f_l\right].$$

We have

$$\begin{aligned} E[\eta_i|D_I] &= f_{I-i}E[\widehat{f}^I_{I-i}|D_I]\prod_{l=I-i+1}^{J-1}E[\widehat{f}^I_l[\alpha_l(\widehat{f}^I_l - f_l)+f_l|D_I] \\ &= f_{I-i}\cdot f_{I-i}\prod_{l=I-i+1}^{J-1}\left\{\alpha_l E[(\widehat{f}^I_l)^2|D_I] - \alpha_l f_l E[\widehat{f}^I_l|D_I] + f_l E[\widehat{f}^I_l|D_I]\right\} \\ &= f^2_{I-i}\prod_{l=I-i+1}^{J-1}\left\{\alpha_l E[(\widehat{f}^I_l)^2|D_I] - \alpha_l (E[\widehat{f}^I_l|D_I])^2 + f_l^2\right\} \\ &= f^2_{I-i}\prod_{l=I-i+1}^{J-1}\left\{\alpha_l (Var[\widehat{f}^I_l|D_I])^2 + f_l^2\right\}. \end{aligned}$$

(d)
From (4.7.30),

$$Var\left(\widehat{f}^I_l|D_I\right) = \frac{\sigma_l^2}{\beta_l^I}\sum_{i=0}^{I-l-1}\frac{\gamma_{i,l}^2}{\delta_{i,l}}Var\left(\widetilde{\varepsilon}_{i,l+1}|D_I\right).$$

The fact $Var\left(\widetilde{\varepsilon}_{i,l+1}|D_I\right) = 1$ completes the proof of this claim. □

Proof. (Lemma 4.7.4)
To prove this lemma it is enough to plug in the estimators $\widehat{f}^I_l$ and $\widehat{\sigma}^I_l$ instead of f_l and σ^I_l in Lemma 4.7.1 (b), (d) and (f) respectively. Therefore we obtain:

(a) We replace f_l by $\widehat{f}^I_l$ in Lemma 4.7.1 (b). We have

$$E\left(\widehat{f}^{I+1}_l|D_I\right) = \frac{\beta_l^I}{\beta_l^{I+1}}\cdot\widehat{f}^I_l + \widehat{f}^I_l\cdot\frac{\gamma_{I-l,l}}{\beta_l^{I+1}} = \widehat{f}^I_l\left(\frac{\gamma_{I-l,l}+\beta_l^I}{\beta_l^{I+1}}\right) = \widehat{f}^I_l.$$

(b) From Lemma 4.7.1 (d) we obtain

$$\widehat{E}\left[\left(\widehat{f}_l^{I+1}\right)^2 |D_I\right] = \left(\frac{\sum_{i=0}^{I-l-1} \frac{\gamma_{i,l}}{C_{i,l}} \cdot C_{i,l+1}}{\beta_l^{I+1}} + \frac{\gamma_{l,I-l} \cdot f_l}{\beta_l^{I+1}}\right)^2 + \frac{\sigma_l^2 \cdot \frac{\gamma_{I-l,l}^2}{\delta_{I-l,l}}}{\left(\beta_l^{I+1}\right)^2}$$

$$= \left(\frac{\beta_l^I}{\beta_l^{I+1}} \widehat{f}_l^I + \frac{\gamma_{l,I-l}}{\beta_l^{I+1}} \widehat{f}_l^I\right)^2 + \frac{\widehat{\sigma}_l^2 \cdot \frac{\gamma_{I-l,l}^2}{\delta_{I-l,l}}}{\left(\beta_l^{I+1}\right)^2} = \left(\widehat{f}_l^I\right)^2 \left(\frac{\beta_l^I}{\beta_l^{I+1}} + \frac{\gamma_{l,I-l}}{\beta_l^{I+1}}\right)^2 + \frac{\widehat{\sigma}_l^2 \cdot \frac{\gamma_{I-l,l}^2}{\delta_{I-l,l}}}{\left(\beta_l^{I+1}\right)^2}$$

$$= \left(\widehat{f}_l^I\right)^2 + \frac{\widehat{\sigma}_l^2 \cdot \frac{\gamma_{I-l,l}^2}{\delta_{I-l,l}}}{\left(\beta_l^{I+1}\right)^2}.$$

(c) From Lemma 4.7.1 (f) we have

$$\widehat{E}\left[C_{i,I-i+1} \cdot \widehat{f}_{I-i}^{I+1} |D_I\right] = \frac{1}{\beta_{I-i}^{I+1}} \left[\left(\sum_{k=0}^{i-1} \frac{\gamma_{k,I-i}}{C_{k,I-i}} \cdot C_{k,I-i+1}\right) \cdot f_{I-i} \cdot C_{i,I-i}\right]$$

$$+ \frac{1}{\beta_{I-i}^{I+1}} \left[\sigma_{I-i}^2 \cdot C_{i,I-i} \cdot \frac{\gamma_{i,I-i}}{\delta_{i,I-i}} + \gamma_{i,I-i} \cdot f_{I-i}^2 \cdot C_{i,I-i}\right]$$

$$= \frac{1}{\beta_{I-i}^{I+1}} \left[\beta_{I-i}^I \cdot \widehat{f}_{I-i} \cdot \widehat{f}_{I-i} \cdot C_{i,I-i} + \widehat{\sigma}_{I-i}^2 \cdot C_{i,I-i} \cdot \frac{\gamma_{i,I-i}}{\delta_{i,I-i}} + \gamma_{i,I-i} \cdot \widehat{f}_{I-i}^2 \cdot C_{i,I-i}\right]$$

$$= \frac{1}{\beta_{I-i}^{I+1}} \left[C_{i,I-i} \cdot \widehat{f}_{I-i}^2 (\beta_{I-i}^I + \gamma_{i,I-i}) + \widehat{\sigma}_{I-i}^2 \cdot C_{i,I-i} \cdot \frac{\gamma_{i,I-i}}{\delta_{i,I-i}}\right]$$

$$= C_{i,I-i} \cdot \widehat{f}_{I-i}^2 \left(1 + \frac{\widehat{\sigma}_{I-i}^2 / \widehat{f}_{I-i}^2}{\beta_{I-i}^{I+1}} \frac{\gamma_{i,I-i}}{\delta_{i,I-i}}\right).$$

□

Proof. (Lemma 4.7.5)
(a) From Lemma 4.7.3 (a) and (d) we have

$$E\left[\left(\widehat{f}_l^I\right)^2 |D_I\right] = Var\left(\widehat{f}_l^I |D_I\right) + f_l^2 = \frac{\sigma_l^2}{(\beta_l^I)^2} \sum_{i=0}^{I-l-1} \frac{\gamma_{i,l}^2}{\delta_{i,l}} + f_l^2.$$

(b) Recall that $\delta_l = \frac{\beta_l^I}{\beta_l^{I+1}}(\widehat{f}_l^I - f_l) + f_l$. From Lemma 4.7.3 (b) and (d) we have

$$E\left[\delta_l^2|D_I\right] = E\left[\left(\frac{\beta_l^I}{\beta_l^{I+1}}\cdot\widehat{f}_l^I + f_l\cdot\frac{\gamma_{I-l,l}}{\beta_l^{I+1}}\right)^2|D_I\right] = \left(\frac{\beta_l^I}{\beta_l^{I+1}}\right)^2\cdot Var\left(\widehat{f}_l^I|D_I\right) + f_l^2$$
$$= \left(\frac{\beta_l^I}{\beta_l^{I+1}}\right)^2\cdot\frac{\sigma_l^2}{(\beta_l^I)^2}\sum_{i=0}^{I-l-1}\frac{\gamma_{i,l}^2}{\delta_{i,l}} + f_l^2. \tag{4.7.31}$$

(c) We have from Lemma 4.7.3 (d)

$$\begin{aligned}E\left[\widehat{f}_l^I\cdot\delta_l|D_I\right] &= E\left[\widehat{f}_l^I\cdot\frac{\beta_l^I}{\beta_l^{I+1}}(\widehat{f}_l^I - f_l) + \widehat{f}_l^I\cdot f_l|D_I\right]\\ &= \frac{\beta_l^I}{\beta_l^{I+1}}E\left[(\widehat{f}_l^I)^2|D_I\right] - \frac{\beta_l^I}{\beta_l^{I+1}}f_l\cdot E\left[\widehat{f}_l^I|D_I\right] + f_l E\left[\widehat{f}_l^I|D_I\right]\\ &= \frac{\beta_l^I}{\beta_l^{I+1}}\left[E\left[(\widehat{f}_l^I)^2|D_I\right] - \left\{E\left[\widehat{f}_l^I|D_I\right]\right\}^2\right] + f_l^2\\ &= \frac{\beta_l^I}{\beta_l^{I+1}}Var\left(\widehat{f}_l^I|D_I\right) + f_l^2 = \frac{\beta_l^I}{\beta_l^{I+1}}\frac{\sigma_l^2}{(\beta_l^I)^2}\sum_{i=0}^{I-l-1}\frac{\gamma_{i,l}^2}{\delta_{i,l}} + f_l^2.\end{aligned}$$

□

4.7.7 Proof of Proposition 4.2.1

(i) See Theorem 2 p. 215 in Mack (1993).

(ii) See discussion on p.112, Corollary on p.141 and Appendix B on p.140 in Mack (1994).

(iii) We have, for $0 \le k \le J-2$,

$$\begin{aligned}(I-k-1)\cdot\widehat{\sigma}_k^2 &= \sum_{i=0}^{I-k-1}\delta_{i,k}(F_{i,k}-\widehat{f}_k)^2 = \sum_{i=0}^{I-k-1}\delta_{i,k}F_{i,k}^2\\ &\quad - 2\sum_{i=0}^{I-k-1}\delta_{i,k}F_{i,k}\cdot\widehat{f}_k + \sum_{i=0}^{I-k-1}\delta_{i,k}\widehat{f}_k^{\,2}.\end{aligned} \tag{4.7.32}$$

Since $\delta_{i,j}$ are $\sigma(C_{i,j})$ measurable, we have

$$E((I-k-1)\cdot\widehat{\sigma}_k^2|B_k) = \sum_{i=0}^{I-k-1}\delta_{i,k}E(F_{i,k}^2|B_k) - 2\sum_{i=0}^{I-k-1}\delta_{i,k}E(F_{i,k}\cdot\widehat{f}_k|B_k) + \sum_{i=0}^{I-k-1}\delta_{i,k}E(\widehat{f}_k^{\,2}|B_k). \tag{4.7.33}$$

Since $F_{i,k}$ and $F_{j,k}$ are independent for $i \neq j$, $E(F_{i,k}^2|B_k) = \frac{\sigma_k^2}{\delta_{i,k}} + f_k^2$ and $\gamma_{i,j}$ are $\sigma(C_{i,j})$ measurable, we have

$$\begin{aligned}
&E(F_{i,k}\cdot\widehat{f}_k|B_k)\\
&= \frac{1}{\sum_{i=0}^{I-k-1}\gamma_{i,k}}\left(\sum_{j=0}^{I-k-1}\gamma_{j,k}\cdot E(F_{i,k}\cdot F_{j,k}|B_k)\right)\\
&= \frac{1}{\sum_{i=0}^{I-k-1}\gamma_{i,k}}\left(\gamma_{i,k}\cdot E(F_{i,k}^2|B_k) + \sum_{j\neq i}^{I-k-1}\gamma_{j,k}\cdot E(F_{i,k}|B_k)\cdot E(F_{j,k}|B_k)\right)\\
&= \frac{1}{\sum_{i=0}^{I-k-1}\gamma_{i,k}}\left(\gamma_{i,k}\cdot E(F_{i,k}^2|B_k) + \sum_{j\neq i}^{I-k-1}\gamma_{j,k}\cdot E(F_{i,k}|B_k)\cdot E(F_{j,k}|B_k)\right)\\
&= \frac{1}{\sum_{i=0}^{I-k-1}\gamma_{i,k}}\left(\gamma_{i,k}\cdot(\frac{\sigma_k^2}{\delta_{i,k}} + f_k^2) + \sum_{j\neq i}^{I-k-1}\gamma_{j,k}f_k^2\right)\\
&= \frac{1}{\sum_{i=0}^{I-k-1}\gamma_{i,k}}\left(\gamma_{i,k}\cdot\frac{\sigma_k^2}{\delta_{i,k}} + f_k^2\sum_{j=0}^{I-k-1}\gamma_{j,k}\right)\\
&= \sigma_k^2\frac{\frac{\gamma_{i,k}}{\delta_{i,k}}}{\sum_{i=0}^{I-k-1}\gamma_{i,k}} + f_k^2.
\end{aligned} \tag{4.7.34}$$

From Lemma 4.7.3 (d)

$$E(\widehat{f}_k^{\,2}|B_k) = Var(\widehat{f}_k|B_k) + (E(\widehat{f}_k|B_k))^2 = \sigma_k^2\frac{\sum_{j=0}^{I-k-1}\frac{\gamma_{j,k}^2}{\delta_{j,k}}}{\left(\sum_{j=0}^{I-k-1}\gamma_{j,k}\right)^2} + f_k^2.$$

Taking together we obtain

$$
\begin{aligned}
&E((I-k-1)\cdot\widehat{\sigma}_k^2|B_k)\\
&=\sum_{i=0}^{I-k-1}\delta_{i,k}E(F_{i,k}^2|B_k)-2\sum_{i=0}^{I-k-1}\delta_{i,k}E(F_{i,k}\cdot\widehat{f}_k|B_k)+\sum_{i=0}^{I-k-1}\delta_{i,k}E(\widehat{f}_k^{\,2}|B_k)\\
&=\sum_{i=0}^{I-k-1}\delta_{i,k}(\frac{\sigma_k^2}{\delta_{i,k}}+f_k^2)-2\sum_{i=0}^{I-k-1}\delta_{i,k}(\sigma_k^2\frac{\frac{\gamma_{i,k}}{\delta_{i,k}}}{\sum_{i=0}^{I-k-1}\gamma_{i,k}}+f_k^2)\\
&+\sum_{i=0}^{I-k-1}\delta_{i,k}(\sigma_k^2\frac{\sum_{j=0}^{I-k-1}\frac{\gamma_{j,k}^2}{\delta_{j,k}}}{\left(\sum_{j=0}^{I-k-1}\gamma_{j,k}\right)^2}+f_k^2)\\
&=(I-k)\sigma_k^2+f_k^2\sum_{i=0}^{I-k-1}\delta_{i,k}-2\sigma_k^2-2f_k^2\sum_{i=0}^{I-k-1}\delta_{i,k}\\
&+\sigma_k^2\frac{\sum_{i=0}^{I-k-1}\delta_{i,k}\sum_{j=0}^{I-k-1}\frac{\gamma_{j,k}^2}{\delta_{j,k}}}{\left(\sum_{j=0}^{I-k-1}\gamma_{j,k}\right)^2}+f_k^2\sum_{i=0}^{I-k-1}\delta_{i,k}\\
&=(I-k-1)\sigma_k^2+\sigma_k^2\left[\frac{\sum_{i=0}^{I-k-1}\delta_{i,k}\sum_{j=0}^{I-k-1}\frac{\gamma_{j,k}^2}{\delta_{j,k}}}{\left(\sum_{j=0}^{I-k-1}\gamma_{j,k}\right)^2}-1\right].
\end{aligned}
\tag{4.7.35}
$$

Finally

$$
E(\widehat{\sigma}_k^2-\sigma_k^2)=E[E[(\widehat{\sigma}_k^2-\sigma_k^2)|B_k]]=\frac{\sigma_k^2}{I-k-1}E\left[\frac{\sum_{i=0}^{I-k-1}\delta_{i,k}\sum_{j=0}^{I-k-1}\frac{\gamma_{j,k}^2}{\delta_{j,k}}}{\left(\sum_{j=0}^{I-k-1}\gamma_{j,k}\right)^2}-1\right].
$$

(iv) It is straightforward from (iii).

(v) See Theorem 1 p. 215 in Mack (1993).

(vi) see Appendix C p.142 in Mack (1994).

Appendices

Appendix A

Limitations and Future Research for Paper I

The main aim of this section is to provide tentative guidelines to establish a stronger version of Theorem 2.3.1, under sharp conditions on the weight function w.

Let us summarize the well-known results on the weak convergence of empirical copula process. Under suitable regularity conditions (see e.g., **Condition 1** and **Condition 2** defined by (1.1.11) and (1.1.12) , on page 21 and 24 respectively) on the first partial derivatives C_j of copula C, we have, as $n \to \infty$,

- $\mathbb{C}_n \rightsquigarrow \mathbb{B}^*$ in the space $\ell^\infty([0,1]^d)$ (see Segers (2012)),
- For any function $w \in \ell^\infty([0,1]^d)$: $w\mathbb{C}_n \rightsquigarrow w\mathbb{B}^*$ in the space $\ell^\infty([0,1]^d)$ (easily checked from results of Segers (2012)),
- For any function $w \in L^2([0,1]^d)$: $w\mathbb{C}_n \rightsquigarrow w\mathbb{B}^*$ in the space $L^2([0,1]^d)$ (see Theorem 2.3.1),

where $\mathbb{C}_n$ and $\mathbb{B}^*$ are defined via (2.2.11) and (2.2.14) respectively.

We wish to solve the following
General Problem:
For an non-square integrable function $w : [0,1]^d \mapsto \mathbb{R}$ ($w \notin L^2$), find a general condition implying the convergence, as $n \to \infty$,

$$w\mathbb{C}_n \rightsquigarrow w\mathbb{B}^* \text{ in the space } L^2([0,1]^d). \tag{A.0.1}$$

It can be noted that in our informal exposition which follows, we do not provide a complete solution of (A.0.1), but rather, we present some ideas which are likely to provide a solution to this problem. The details are left for future and on-going research.

A.1 Weak convergence of $w\mathbb{C}_n$ - case of independence copula

In the particular case of the independence copula, the solution of the ***General Problem*** stated in A.0.1 could be useful in constructing tests of independence based on the empirical copula process. A typical example being given by Anderson-Darling type test statistics. We present in this section two methods about how one could deal with the ***General Problem***.

Method 1 - convergence in weighted sup-norm.
Let D^* be the set of all those functions $q \geq 0$ which are continuous on $[0,1]$, strictly positive on $(0,1]$ and non-decreasing. Assume that q and $\widetilde{q}$ are from the class D^*. Denote $|\mathbf{u}| = \prod_{i=1}^n u_i$. The main purpose of this approach is to provide the sufficient condition on the functions q and $\widetilde{q}$ for following convergence

$$\sup_{\mathbf{u}\in[0,1]^d} |\mathbb{C}_n(\mathbf{u}) - \mathbb{B}^*(\mathbf{u})|/q(|\mathbf{u}|)\widetilde{q}(1-|\mathbf{u}|) \to 0,$$

in probability, as $n \to \infty$. By the elementary arguments, if w is chosen in such a way that

$$\int_{[0,1]^d} w(\mathbf{u})^2 q(|\mathbf{u}|)\widetilde{q}(1-|\mathbf{u}|)\, d\mathbf{u} < \infty,$$

then $w\mathbb{C}_n \rightsquigarrow w\mathbb{B}^*$ in the space $L^2([0,1]^d)$ as $n \to \infty$. For more details about the techniques based on weighted sup-norms refer to Shao and Yu (1996) and Einmahl et al (1988) (see Theorem 2.1 p.196, Theorem 2.2 p.198 and Corollary 2.1 p.199).

Method 2 - direct computations.

The second possible way to address our ***General Problem*** is based on direct computations.

To simplify exposition, we limit ourselves in the following to $d = 2$. Let C be the bivariate independence copula function, i.e., $C(u,v) = uv$. We recall the definition of the empirical copula process $\mathbb{C}_n$ for this same copula. Namely,

$$\mathbb{C}_n(u,v) = n^{1/2}(\widetilde{C}_n(u,v) - uv), \ for \ \ (u,v) \in [0,1]^2. \tag{A.1.1}$$

Here $\widetilde{C}_n$ is the empirical copula function defined in (2.2.9). Recall the definition of the following empirical processes. We write

$$\begin{aligned}
\alpha_n(u,v) &:= n^{1/2}(H_n(u,v) - uv),\\
\alpha_{n;U}(u) &:= n^{1/2}(F_n(u) - u),\\
\alpha_{n;V}(v) &:= n^{1/2}(G_n(v) - v),\\
\beta_{n;U}(u) &:= n^{1/2}(F_n^-(u) - u),\\
\beta_{n;V}(v) &:= n^{1/2}(G_n^-(v) - v).
\end{aligned}$$

The function H_n is a bivariate empirical function based on the sample $(U_1,V_1),...,(U_n,V_n)$, with c.d.f F, i.e.,

$$H_n(u,v) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{U_i \le u, V_i \le v\}}.$$

We have the following decomposition of $\mathbb{C}_n$ (see, e.g., the proof of Theorem 2.3.1)

$$\begin{aligned}
\mathbb{C}_n(u,v) &= \alpha_n(F_n^-(u), G_n^-) + u\beta_{n;V}(v) + v\beta_{n;U}(u) + n^{-1/2}\beta_{n;U}(u)\beta_{n;V}(v)\\
&= \mathbb{C}_{n;0}(u,v) + R_{1n}(u,v) + R_{2n}(u,v) + R_{3n}(u,v),
\end{aligned}$$

where

$$\begin{aligned}
\mathbb{C}_{n;0}(u,v) &= \alpha_n(u,v) - v\alpha_{n;U}(u) - u\alpha_{n;V}(v),\\
R_{1n}(u,v) &= \alpha_n(F_n^-(u), G_n^-(v)) - \alpha_n(u,v),\\
R_{2n}(u,v) &= v(\alpha_{n;U}(u) + \beta_{n;U}(u)) + u(\alpha_{n;V}(v) + \beta_{n;V}(v)),\\
R_{3n}(u,v) &= n^{-1/2}\beta_{n;U}(u)\beta_{n;V}(v).
\end{aligned}$$

Basic arguments show that, in order to prove the convergence $w\mathbb{C}_n \rightsquigarrow w\mathbb{B}^*$, it is sufficient to show that, as $n \to \infty$,

$$w\mathbb{C}_{n;0} \rightsquigarrow w\mathbb{B}^*, \tag{A.1.2}$$

and

$$wR_{in} \to 0 \text{ in probability, for } i = 1, 2, 3. \tag{A.1.3}$$

Convergence of $w\mathbb{C}_{n;0}$.
To find the condition on w which implies this convergence we suggest the use of the techniques developed in Einmahl et al (1988).
Convergence of wR_{3n}.
We show this convergence for $w = w_0$, where

$$w_0^2(u,v) = 1/Var(\mathbb{B}^*(u,v)) = \frac{1}{u(1-u)\cdot v(1-v)}.$$

Since $w_0R_{3n} = n^{-1/2}[\frac{1}{u(1-u)} \cdot \beta_{n;U}(u)][\frac{1}{v(1-v)} \cdot \beta_{n;V}(v)]$, from Corollary 3 in Mason (1984) (p.248, applied with: $\nu_1 = \nu_2 = 0, M = 1, g(u) = u(1-u)$ and $g(v) = v(1-v)$ respectively) we get: $\frac{1}{I(1-I)} \cdot \beta_{n;U} \rightsquigarrow B$ and $\frac{1}{I(1-I)} \cdot \beta_{n;V} \rightsquigarrow B$, where I denotes the identity function on $[0,1]$ and B denotes a Brownian bridge on $[0,1]$. Hence, as $n \to \infty$, $w_0R_{3n} \rightsquigarrow 0$, which implies that $w_0R_{3n} \to 0$ in probability in $L^2([0,1]^2)$.
Convergence of wR_{in} for $i = 1, 2$. The convergence to 0 in probability of wR_{in}, for $i = 1, 2$, is implied (through the Markov and Jensen inequalities, and the Fubini theorem) by the following convergence

$$\int_{[0,1]^2} \mathbb{E}[wR_{in}(u,v)]^2 dudv \to 0 \text{ as } n \to \infty \text{ for}, i = 1, 2. \tag{A.1.4}$$

The condition (A.1.4) is fulfilled, for $i = 1, 2$, if the function $(u,v) \mapsto \mathbb{E}[wR_{in}(u,v)]^2$ is uniformly integrable and $\mathbb{E}[R_{in}(u,v)]^2 \to 0$ almost everywhere on $[0,1]^2$ as $n \to \infty$.
Uniform integrability of $\mathbb{E}[wR_{1n}]^2$.
We have

$$\begin{aligned}
\alpha_n(F_n^-(u), G_n^-(v)) - \alpha_n(u,v) &= n^{1/2}\left\{H_n(F_n^-(u), G_n^-) - H_n(u,v)\right\} \\
&+ n^{1/2}\left\{C(u,v) - C(F_n^-(u), G_n^-)\right\} \\
&\leq n^{1/2}\left\{|F_n(F_n^-(u)) - F_n(u)| + |G_n(G_n^-(v)) - G_n(v)|\right\} \\
&+ n^{1/2}\left\{|F_n^-(u) - u| + |G_n^-(v) - v|\right\}.
\end{aligned}$$

where the second inequality is due to Fact A.4.1 (below), when applied to H_n and C .
The claim about uniform integrability of $\mathbb{E}[wR_{1n}]^2$ is correct, given that the functions

$$w^2 \cdot \mathbb{E}\left[n^{1/2}\left\{|F_n \circ F_n^- - F_n| + |G_n \circ G_n^- - G_n|\right\}\right]^2,$$

and

$$w^2 \cdot \mathbb{E}\left[n^{1/2}\left\{|F_n^- - I| + |G_n^- - I|\right\}\right]^2,$$

are uniformly integrable, where I denotes the identity function on [0,1].
Convergence of $\mathbb{E}R_{1n}^2$.
We have

$$R_{1n}^2 = \left\{\alpha_n(F_n^-(u), G_n^-(v)) - \alpha_n(u,v)\right\}^2 \leq \left\{\sup_{|u-s|\times|v-t|\leq a_n} |\alpha_n(u,v) - \alpha_n(s,t)|\right\}^2,$$

where $a_n = n^{-1/2}(\log\log n)^{1/2}$. Hence

$$\mathbb{E}R_{1n}^2 \leq \mathbb{E}\left[\sup_{|u-s|\times|v-t|\leq a_n} |\alpha_n(u,v) - \alpha_n(s,t)|\right]^2,$$

and from Corollary A.4.1(see Auxiliary facts below), we get $\mathbb{E}R_{1n}^2 \to 0$.

Uniform integrability of $\mathbb{E}[wR_{2n}]^2$.
We have

$$R_{2n}^2 \leq [n^{1/2}\{[F_n^-(u) - u + F_n(u) - u] + [G_n^-(v) - v + G_n(v) - v]\}]^2.$$

Hence, the claim about uniform integrability of $\mathbb{E}[wR_{2n}]^2$ is implied by uniformly integrability of the function

$$w^2 \cdot \mathbb{E}[n^{1/2}\{[F_n^- - I + F_n - I] + [G_n^- - I + G_n - I]\}]^2.$$

Convergence of $\mathbb{E}R_{2n}^2$.
We have

$$\begin{aligned} n^{-1/2}[\alpha_{n;U}(u) + \beta_{n;U}(u)] &= [F_n^-(u) - u] + [F_n(u) - u] + F_n(F_n^-(u)) - F_n(F_n^-(u)) \\ &= [F_n(F_n^-(u)) - u] - [F_n(F_n^-(u)) - F_n^-(u) - (F_n(u) - u)]. \end{aligned}$$

From (1.6) (p. 370) in Shorack (1982): $\sqrt{n}\|F_n(F_n^-) - I\|_\infty \leq \frac{1}{\sqrt{n}}$. Hence

$$\begin{aligned}
[\alpha_{n;U}(u) + \beta_{n;U}(u)]^2 &\leq 2\{\sqrt{n}[F_n(F_n^-(u)) - u]\}^2 \\
&+ 2\{[\alpha_{n;U}(F_n^-(u)) - \alpha_{n;U}(u)]\}^2 \\
&\leq 2\{\sqrt{n}\|F_n(F_n^-) - I\|_\infty\}^2 \\
&+ 2\{\sup_{0<u<1} |\alpha_{n;U}(F_n^-(u)) - \alpha_{n;U}(u)|\}^2 \\
&\leq 2\frac{1}{n} + 2\left\{\sup_{t\leq s;|s-t|\leq a_n} |\alpha_{n;U}(s) - \alpha_{n;U}(t)|\right\}^2.
\end{aligned}$$

From Corollary A.4.1 we get that $\mathbb{E}[\alpha_{n;U} + \beta_{n;U}]^2 \to 0$ a.e on $[0,1]$ as $n \to \infty$. Let define

$$\begin{aligned}
\varphi_1(u,v) :=& n^{1/2}\left\{|F_n(F_n^-(u)) - F_n(u)| + |G_n(G_n^-(v)) - G_n(v)|\right\}, \\
\varphi_2(u,v) :=& n^{1/2}\left\{|F_n^-(u) - u| + |G_n^-(v) - v|\right\}, \\
\varphi_3(u,v) :=& n^{1/2}\{[F_n^-(u) - u + F_n(u) - u] + [G_n^-(v) - v + G_n(v) - v]\}.
\end{aligned}$$

We present below the list of potential sufficient conditions providing the solution for *general problem* in the case of independence copula:

C.1 The function w is such that $w\mathbb{C}_{n;0} \rightsquigarrow w\mathbb{B}^*$ in space L^2,

C.2 The function w is such

$$w = O(w_0) = O\left(\frac{1}{\sqrt{uv(1-u)(1-v)}}\right)$$

(required for convergence of wR_{3n}),

C.3 The function $\mathbb{E}[w \times \varphi_1]^2$ is uniformly integrable (required for uniform integrability of $\mathbb{E}[wR_{1n}]^2$),

C.4 The function $\mathbb{E}[w \times \varphi_2]^2$ is uniformly integrable (required for uniform integrability of $\mathbb{E}[wR_{1n}]^2$),

C.5 The function $\mathbb{E}[w \times \varphi_3]^2$ is uniformly integrable (required for uniform integrability of $\mathbb{E}[wR_{2n}]^2$),

Remark A.1.1 *From the statistical point of view, it would be interesting to verify whether the conditions C.1 and C.3-C.5 are fulfilled by the function* w_0. *The function* w_0 *is not square integrable and it is equal to* $1/Var(\mathbb{B}^*)$. *Therefore, the positive answer to that query would give the possibility to construct the statistical test of independence based on empirical copula process and on Anderson-Darling type goodness-of-fit statistic.*

A.2 Weak convergence of $w\mathbb{C}_n$ - case of general copula

One of possible way to approach the *general problem* for general copula function, would be to use the general decomposition of copula process (see, e.g., the proof of Theorem 2.3.1) i.e.,

$$\mathbb{C}_n = \mathbb{C}_{n;0} + R_{1n} + R_{2n} + R_{3n},$$

where

$$\begin{aligned}
\mathbb{C}_{n;0} &= \alpha_n + \sum_{j=1}^{d} C_j \alpha_{nj}, \\
R_{1,n} &= \alpha_n(G_{n1}^{-1}, ..., G_{nd}^{-1}) - \alpha_n, \\
R_{2,n} &= \sum_{j=1}^{d} C_j(\alpha_{nj} + \beta_{nj}), \\
R_{3,n} &= \sqrt{n}[C(G_{n1}^{-1}, ..., G_{nd}^{-1}) - C] - \sum_{j=1}^{d} C_j \beta_{nj},
\end{aligned}$$

and the function C_j is the *j-th* partial derivative of copula function C. Thereafter, we can mimick the techniques presented above in **Method 2**.

A.3 Goodness-of-fit test based on the random weighted function

In Section 2.4.3 we introduced the statical test based on the empirical copula process and on the random weighted function (see 2.4.5). In the present section we consider some other possible choices of the random weight functions.

- $w(u, v, \theta_n) = \dfrac{1}{[var(\mathbb{B}^*(u, v, \theta_n))]^\alpha}$, where $0 < \alpha < 1$ and $var(\mathbb{B}^*(u, v, \theta_n))$ is the estimator of the variance function of the process $\mathbb{B}^*$ defined in (2.2.14). The function $var(\mathbb{B}^*(u, v, \theta_n))$ is computed as follows: in expression of $var(\mathbb{B}^*(u, v)) = \mathbb{E}\left[\mathbb{B}^*(u, v)\right]^2$ the copula function C_θ is replaced by its parametric estimator C_{θ_n} and the partial derivatives $\partial C_\theta/\partial u$ and $\partial C_\theta/\partial v$ are replaced by their consistent estimators (see Rémillard and Scaillet (2009), Prop. A.2) $\partial C_\theta^{(n)}/\partial u$ and $\partial C_\theta^{(n)}/\partial v$ respectively, given by

$$\frac{\partial C_\theta^{(n)}}{\partial u}(u, v) := \frac{1}{2n^{-1/2}} \left\{ \widetilde{C}_n(u + n^{-1/2}, v) - \widetilde{C}_n(u - n^{-1/2}, v) \right\},$$

  and

$$\frac{\partial C_\theta^{(n)}}{\partial v}(u, v) := \frac{1}{2n^{-1/2}} \left\{ \widetilde{C}_n(u, v + n^{-1/2}) - \widetilde{C}_n(u, v - n^{-1/2}) \right\}.$$

- $w(u, v, \theta_n) = \dfrac{1}{[var(\mathbb{B}^{**}(u, v))_n]^\gamma}$, where $0 < \gamma < 1$ and $var(\mathbb{B}^{**}(u, v))_n$ is the estimator of the variance function $var(\mathbb{B}^{**}(u, v))$ of the limit process $\mathbb{B}^{**}$ form Proposition 2.4.1. In order to compute $var(\mathbb{B}^{**}(u, v))_n$ we may use the multiplier technique (see, e.g, Kojadinovic et al (2010),Rémillard and Scaillet (2009)).

A.4 Auxiliary facts

A.4.1 Moments of the modulus of continuity of the empirical process

Definition A.4.1 *Let consider the half-open rectangles $R = R(s, t)$ on $\mathbb{R}^d$ given by $(s_1, t_1] \times ... \times (s_d, t_d]$ and by*

$$\mathcal{R} = \{R(s, t) : R(s, t) \subset [0, 1]^d\},$$

we denote the class of all such rectangles contained in the unit cube. The modulus of continuity of empirical process that will be considered here is *based on the partition $\mathcal{P}_m \subset \mathcal{R}$ of the unit cube into the squares of equal size*

$$R = R_{k_1...k_d} = \left(\frac{k_1-1}{m}, \frac{k_1}{m}\right] \times ... \times \left(\frac{k_d-1}{m}, \frac{k_d}{m}\right], \; m \in \mathbb{N}, \; k_1, .., k_d \in \{1, ..., m\}.$$

Then this modulus of continuity of empirical process α_n is defined as

$$M(\alpha_n; m) = \max_{R \in \mathcal{P}_m} \sup_{s,t \in R} |\alpha_n(s) - \alpha_n(t)|.$$

We present now the main result of Einmahl and Ruymgaart (1987) and the corollary which can be derived from this result.

Theorem A.4.1 *Let $\delta \in (0, \frac{1}{2})$ be arbitrary. For integers $m \geq 2$ and $k \in \mathbb{N}$ there exist numbers $0 < C_1 = C_1(d, k, \delta) \leq C_2 = C_2(d, k, \delta) < \infty$ such that for any copula function C we have*

$$C_1 \left(\frac{\log m}{m}\right)^{k/2} \leq \mathbb{E}[M(\alpha_n; m)]^k \leq C_2 \left(\frac{\log m}{m}\right)^{k/2}.$$

Corollary A.4.1 *For any sequence a_n of real numbers such that $a_n \to 0$ and any integer $k \in \mathbb{N}$ we have*

$$\mathbb{E}\left[\sup_{t \leq s; |s-t| \leq a_n} |\alpha_n(s) - \alpha_n(t)|\right]^k \to 0 \text{ as } n \to \infty, \text{ and } |s-t| = \prod_{i=1}^{d} |s_i - t_i|.$$

Fact A.4.1 *For every bivariate cumulative distribution function H and its margins F and G*

$$|H(x,y) - H(x',y')| \leq |F(x) - F(x')| + |F(y) - F(y')|.$$

A.5 Asymptotic normality of multivariate rank order statistics

We consider multivariate rank statistics of the form

$$R_n = \frac{1}{n} \sum_{i=1}^{n} J(F_{n1}(X_i(1)), ..., F_{nd}(X_i(d))),$$

where J is a measurable real function on $[0,1]^d$. The asymptotic normality of R_n has been investigated, subject to suitable regularity conditions imposed on J by many authors (see, e.g., Fermanian et al (2004) and the references therein). We establish a result of the kind without the assumption of continuity on $[0,1]^d$ of the first partial derivatives of C. We will impose conditions on J similar to that of Theorem 6 of (Fermanian et al, 2004, p.854).

Proposition A.5.1 *Let H have continuous marginals. Assume that J is of bounded variation, continuous from above, with discontinuities of the first kind and such that the measure dJ is absolutely continuous with respect to the Lebegue measure λ_d, i.e., fulfilling*

$$dJ \ll \lambda_d. \tag{A.5.1}$$

Then, as $n \to \infty$,

$$n^{1/2}(R_n - \mathbb{E}R_n) \xrightarrow{d} \int_{[0,1]^d} \mathbb{B}^*(\mathbf{u}) \frac{\partial^d J}{\partial u_1 ... \partial u_d}(\mathbf{u})\, d\mathbf{u}.$$

Proof. We have

$$\begin{aligned} n^{1/2}(R_n - \mathbb{E}R_n) &= \int_{[0,1]^d} \sqrt{n}\{\overline{C}_n(\mathbf{u}) - C(\mathbf{u})\}\, dJ(\mathbf{u}) + O(n^{-1/2}), \\ &= \int_{[0,1]^d} \sqrt{n}\{\overline{C}_n(\mathbf{u}) - C(\mathbf{u})\} \frac{\partial^d J}{\partial u_1 ... \partial u_d}(\mathbf{u})\, d\mathbf{u} + O(n^{-1/2}). \end{aligned} \tag{A.5.2}$$

The first equality in (A.5.2) is an easy adaptation of the proof of Theorem 6 of Fermanian et al (2004), and the second equality is a consequence of the assumption (A.5.1). Since $n^{1/2}|C_n - \overline{C}_n|_\infty \to 0$ a.s, as $n \to \infty$ one can replace in Theorem 2.3.1 C_n by $\overline{C}_n$ and obtain a version of this theorem for $w\overline{\mathbb{C}}_n = wn^{1/2}(\overline{C}_n - C)$. Finally, the asymptotic normality of R_n follows from (A.5.2), in combination with Theorem 2.3.1 when applied to $\overline{\mathbb{C}}_n$ with the weight function $w = \frac{\partial^d J}{\partial u_1 ... \partial u_d} \in L^2\left([0,1]^d\right)$ (refer to (A.5.1)). We make use of the continuous mapping theorem, and applied to the transformation:

$$L^2\left([0,1]^d\right) \ni f \mapsto \int_{[0,1]^d} f(\mathbf{u})\, d\mathbf{u}.$$

The continuity of this last unctional follows from the Cauchy-Schwarz inequality.

A.6 GOF tests for copulas-weak convergence of T_n statistic

In the sequel, $\xrightarrow{\mathbb{P}}$ refers to convergence in (outer) probability and $\rightsquigarrow$ denotes the weak convergence.

The goal is to show that, for any function $w \in L^2([0,1]^d)$, the weak convergence

$$w\mathbb{C}_n \rightsquigarrow w\mathbb{B}^*,$$

in the space $L^2([0,1]^d)$, implies that the statistic

$$\int_{[0,1]^d} \mathbb{C}_n^2(\mathbf{u})\ d\overline{C}_n(\mathbf{u}),$$

converges weakly to the corresponding quadratic functional of the Gaussian process.
To prove this result it is sufficient to show the following

Proposition A.6.1 (Conjecture) *If a sequence $\mathbb{G}_n$ of processes is tight with respect to the L^2* ***norm on the space*** *$L^2[0,1]^d$ of square integrable functions with respect to the Lebesgue measure, then, as $n \to \infty$,*

$$\triangle_n := \int_{[0,1]^d} \mathbb{G}_n(\mathbf{u})\ d\overline{C}_n(\mathbf{u}) - \int_{[0,1]^d} \mathbb{G}_n(\mathbf{u})\ d\overline{C}(\mathbf{u}) \xrightarrow{\mathbb{P}} 0.$$

The similar fact to the above proposition concerning the space l^∞ has been proved in Genest et al (2013) (see Proposition 4, see also,Tsukahara (2000), p.10).

Proposition A.6.2 *If a sequence $\mathbb{G}_n$ of processes is tight with respect to the* ***uniform norm on the space*** *$\mathcal{C}[0,1]^d$ of continuous functions on $[0,1]^d$, then, as $n \to \infty$,*

$$\triangle_n := \int_{[0,1]^d} \mathbb{G}_n(\mathbf{u})\ dC_n(\mathbf{u}) - \int_{[0,1]^d} \mathbb{G}_n(\mathbf{u})\ dC(\mathbf{u}) \xrightarrow{\mathbb{P}} 0.$$

In particular, we are interested in the previous result in the case where the empirical process $\mathbb{G}_n = R_{3,n}^2$ where

$$R_{3,n} := \sqrt{n}[C(G_{n1}^{-1}, ..., G_{nd}^{-1}) - C] - \sum_{j=1}^{d} C_j \beta_{nj},$$

and C_j denotes the first order partial derivatives of C.

A.7 Useful Theorems, Facts, Inequalities, etc.

Proposition A.7.1 *For every copula function C and for $j = 1, ..., d$, the partial derivative C_j exist for almost all $\mathbf{u} \in [0,1]^d$ with respect to the Lebesgue-measure. For such $\mathbf{u}$, we have*

$$0 \leq C_j(\mathbf{u}) \leq 1, \ \mathbf{u} \in [0,1]^d.$$

Proof. See Theorem 2.2.7 of Nelsen (2006). □

Fact A.7.1 (Slutsky's theorem) *Let X_n and Y_n be sequences of random elements. If $X_n \rightsquigarrow X$, as $n \to \infty$, and $Y_n \xrightarrow{\mathbb{P}} c$, where c is constant, then*

(a) $X_n + Y_n \rightsquigarrow X + c$,

(b) $X_n \cdot Y_n \rightsquigarrow cX$,

Note that, in the statement of the theorem, the condition "Y_n converges in probability to a constant c" may be replaced with "Y_n converges in distribution to a constant c" these two requirements are equivalent according to this property. The requirement that Y_n converges to a constant is important if it were to converge to a non-degenerate random variable, the theorem would be no longer valid. The theorem remains valid if we replace all convergences in distribution with convergences in probability.

Fact A.7.2 (Hölder's Inequality) *Let $1 \leq p, q \leq \infty$ with $1/p + 1/q = 1$. Let $f \in L^p([0,1]^d, \lambda_d)$ and $g \in L^q([0,1]^d, \lambda_d)$. Then*

$$|fg|_1 \leq |f|_p \cdot |g|_q,$$

where, for $1 \leq r < \infty$, $|h|_r = \left(\int_{[0,1]^d} h^r \, d\lambda_d \right)^{1/r}$.

Fact A.7.3 *Let* $f \in L^2\left([0,1]^d, \lambda_d\right)$ *and* $g \in l^\infty\left([0,1]^d\right)$. *Then*

$$|fg|_2 \leq |f|_2 \cdot |g|_\infty.$$

Proof of Fact A.7.3.
We apply the Hölder's Inequality (see Fact A.7.2) with $p = 1$ and $q = \infty$. We obtain,

$$|fg|_2^2 = |f^2g^2|_1 \leq |f^2|_1 \cdot |g^2|_\infty = |f|_2^2 \cdot |g^2|_\infty.$$

Furthermore, $|g^2|_\infty \leq |g|_\infty^2$ ($|g(x)| \leq |g|_\infty$ implies that $|g(x)|^2 \leq |g|_\infty^2$ and taking supremum of both sides proves this claim).
Finally, we conclude

$$|fg|_2^2 \leq |f|_2^2 \cdot |g|_\infty^2.$$

Theorem A.7.1 (Second Fundamental Theorem of Calculus) *If* $f : [a,b] \to \mathbb{R}$ *is integrable and there exists a function* $g : [a,b] \to \mathbb{R}$ *such that* $g' = f$ *then:*

$$\int_a^b f = g(b) - g(a).$$

Note: f need not be continuous. An example of this would be $g(x) := x^2 \sin(1/x)$ which is differentiable on $[-1,1]$ but g' is not continuous at 0.

Appendix B

Data for Chapter 2 and Chapter 3

B.1 Run-off Triangles

Accident Year i	Development Year j								
	0	1	2	3	4	5	6	7	8
0	2'202'584	3'210'449	3'468'122	3'545'070	3'621'627	3'644'636	3'669'012	3'674'511	3'678'633
1	2'350'650	3'553'023	3'783'846	3'840'067	3'865'187	3'878'744	3'898'281	3'902'425	
2	2'321'885	3'424'190	3'700'876	3'798'198	3'854'755	3'878'993	3'898'825		
3	2'171'487	3'165'274	3'395'841	3'466'453	3'515'703	3'548'422			
4	2'140'328	3'157'079	3'399'262	3'500'520	3'585'812				
5	2'290'664	3'338'197	3'550'332	3'641'036					
6	2'148'216	3'219'775	3'428'335						
7	2'143'728	3'158'581							
8	2'144'738								

Table B.1: Run-off triangle: Triangle 1 (T1) (cumulative payments)

Accident Year i	Development Year j									
	0	1	2	3	4	5	6	7	8	9
0	5 012	8 269	10 907	11 805	13 569	16 211	18 039	18 638	18 692	18 864
1	106	4 285	5 396	10 666	13 782	15 599	15 496	16 169	16 704	
2	3 410	8 992	13 873	16 141	18 735	22 214	22 863	23 466		
3	5 655	11 555	15 766	21 266	23 425	26 083	27 067			
4	1 092	9 565	15 836	22 169	25 955	26 180				
5	1 513	6 445	11 702	12 935	15 852					
6	557	4 020	10 946	12 314						
7	1 351	6 947	13 112							
8	3 133	5 395								
9	2 063									

Table B.2: Run-off triangle: Triangle 2 (T2) (cumulative payments)

B.2 Individual link ratios $\{F_{i,j}\}$

Accident Year i	Development Year j 0	1	2	3	4	5	6	7
0	1,458	1,080	1,022	1,022	1,006	1,007	1,001	1,001
1	1,512	1,065	1,015	1,007	1,004	1,005	1,001	
2	1,475	1,081	1,026	1,015	1,006	1,005		
3	1,458	1,073	1,021	1,014	1,009			
4	1,475	1,077	1,030	1,024				
5	1,457	1,064	1,026					
6	1,499	1,065						
7	1,473							

Table B.3: Individual link ratios $F_{i,j}$ (age-to-age factors) of run-off triangle T1 defined in Table B.1

Accident Year i	Development Year j 0	1	2	3	4	5	6	7	8
0	1,650	1,319	1,082	1,149	1,195	1,113	1,033	1,003	1,009
1	40,425	1,259	1,977	1,292	1,132	0,993	1,043	1,033	
2	2,637	1,543	1,163	1,161	1,186	1,029	1,026		
3	2,043	1,364	1,349	1,102	1,113	1,038			
4	8,759	1,656	1,400	1,171	1,009				
5	4,260	1,816	1,105	1,226					
6	7,217	2,723	1,125						
7	5,142	1,887							
8	1,722								

Table B.4: Individual link ratios $F_{i,j}$ (age-to-age factors) of run-off triangle T2 defined in Table B.2

Bibliography

Anderson TW, Darling DA (1952) Asymptotic theory of certain “goodness of fit” criteria based on stochastic processes. Ann Math Statistics 23:193–212

Araujo A, Giné E (1980) The central limit theorem for real and Banach valued random variables. John Wiley & Sons, New York-Chichester-Brisbane, wiley Series in Probability and Mathematical Statistics

del Barrio E, Deheuvels P, van de Geer S (2007) Lectures on empirical processes. EMS Series of Lectures in Mathematics, European Mathematical Society (EMS), Zürich

Berg D, Quessy JF (2009) Local power analyses of goodness-of-fit tests for copulas. Scand J Stat 36(3):389–412

Billingsley P (1968) Convergence of probability measures. John Wiley & Sons Inc., New York

Billingsley P (1995) Probability and measure, 3rd edn. Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, a Wiley-Interscience Publication

Buchwalder M, Bühlmann H, Merz M, Wüthrich MV (2006) The mean square error of prediction in the chain ladder reserving method (Mack and Murphy revisited). Astin Bull 36(2):521–542

Bühlmann H, De Felice M, Gisler A, Moriconi F, Wüthrich MV (2009) Recursive credibility formula for chain ladder factors and the claims development result. Astin Bull 39(1):275–306

Charpentier A, Denuit M (2005) Mathématiques de l’assurance non-vie : Tome 2, Tarification et Provisionnement. Economica

Cochran WG (1952) The χ^2 test of goodness of fit. Ann Math Statistics 23:315–345

Cramér H (1928) On the composition of elementary errors. II. Statistical applications. Skand Aktuarietidskr 11:141–180

d'Agostino RB, Stephens MA (eds) (1986) Goodness-of-fit techniques, Statistics: Textbooks and Monographs, vol 68. Marcel Dekker Inc., New York

Darling DA (1955) The Cramér-Smirnov test in the parametric case. Ann Math Statist 26:1–20

Deheuvels P (1979) La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance. Acad Roy Belg Bull Cl Sci (5) 65(6):274–292

Deheuvels P (1981) A Kolmogorov-Smirnov type test for independence and multivariate samples. Rev Roumaine Math Pures Appl 26(2):213–226

Deheuvels P (2005) Weighted multivariate Cramér-von Mises-type statistics. Afr Stat 1(1):1–14

Deheuvels P (2009) A multivariate bahadur – kiefer representation for the empirical copula process. Zap Nauchn Sem POMI 364:120–147

Deheuvels P, Martynov G (2003) Karhunen-Loève expansions for weighted Wiener processes and Brownian bridges via Bessel functions. In: High dimensional probability, III (Sandjberg, 2002), Progr. Probab., vol 55, Birkhäuser, Basel, pp 57–93

Donsker MD (1951) An invariance principle for certain probability limit theorems. Mem Amer Math Soc, 1951(6):12

Donsker MD (1952) Justification and extension of Doob's heuristic approach to the Komogorov-Smirnov theorems. Ann Math Statistics 23:277–281

Doob JL (1949) Heuristic approach to the Kolmogorov-Smirnov theorems. Ann Math Statistics 20:393–403

Einmahl JHJ, Ruymgaart FH (1987) The order of magnitude of the moments of the modulus of continuity of multiparameter Poisson and empirical processes. J Multivariate Anal 21(2):263–273

Einmahl JHJ, Ruymgaart FH, Wellner JA (1988) A characterization of weak convergence of weighted multivariate empirical processes. Acta Sci Math (Szeged) 52(1-2):191–205

England PD, Verrall RJ (2002) Stochastic claims reserving in general insurance. Presented to the Institute of Actuaries, 28 January 2002

Fermanian JD, Radulović D, Wegkamp M (2004) Weak convergence of empirical copula processes. Bernoulli 10(5):847–860

Frees EW, Valdez EA (1998) Understanding relationships using copulas. N Am Actuar J 2(1):1–25

Gaenssler P, Stute W (1987) Seminar on empirical processes, DMV Seminar, vol 9. Birkhäuser Verlag, Basel

Genest C, Rémillard B (2008) Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. Ann Inst Henri Poincaré Probab Stat 44(6):1096–1127

Genest C, Ghoudi K, Rivest LP (1995) A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika 82(3):543–552

Genest C, Ghoudi K, Rivest LP (1998) "Understanding relationships using copulas," by Edward Frees and Emiliano Valdez, January 1998. N Am Actuar J 2(3):143–149

Genest C, Quessy JF, Rémillard B (2006) Goodness-of-fit procedures for copula models based on the probability integral transformation. Scand J Statist 33(2):337–366

Genest C, Rémillard B, Beaudoin D (2009) Goodness-of-fit tests for copulas: a review and a power study. Insurance Math Econom 44(2):199–213

Genest C, Huang W, Dufour JM (2013) A regularized goodness-of-fit test for copulas. J SFdS 154(1):64–77

Genz A (2004) Numerical computation of rectangular bivariate and trivariate normal and t probabilities. Stat Comput 14(3):251–260

Grimshaw SD (1989) A unified approach to estimating tail behavior. Technical Report 55(ARO 23010. 4-MA):1–116

Hoeffding W (1940) Scale-invariant correlation theory. In: N I Fisher and P K Sen (eds): The Collected Works of Wassily Hoeffding New York: Springer-Verlag pp 57–107

Hoeffding W (1941) Scale-invariant correlation measures for discontinuous distributions. In: N I Fisher and P K Sen (eds): The Collected Works of Wassily Hoeffding New York: Springer-Verlag pp 109–133

Joe H (1997) Multivariate models and dependence concepts, Monographs on Statistics and Applied Probability, vol 73. Chapman & Hall, London

de Jong P (2012) Modeling dependence between loss triangles. North American Actuarial Journal 16(1)

Kojadinovic I, Yan J, Holmes M (2010) Fast large-sample goodness-of-fit tests for copulas. Statistica Sinica, in press

Kosorok MR (2008) Introduction to empirical processes and semiparametric inference. Springer Series in Statistics, Springer, New York

Mack T (1993) Distribution-free calculation of the standard error of chain ladder reserve estimates. AST1N Bulletin, 23 pp 213–222

Mack T (1994) Measuring the variability of chain ladder reserve estimates. Casualty Actuarial Society Forum Casualty Actuarial Society - Arlington, Virginia, Spring, Vol 1 pp 101–182

Mack T (1999) The standard error of chain ladder reserve estimates recursive calculation and inclusion of tail factor. Astin Bulletin, Vol 29 (2):361–366

Mack T, Quarg G, Braun C (2006) The mean square error of prediction in the chain ladder reserving method—a comment. Astin Bull 36(2):543–552

Marshall AW, Olkin I (1967) A generalized bivariate exponential distribution. J Appl Probability 4:291–302

Mason DM (1984) Weak convergence of the weighted empirical quantile process in $L^2(0,1)$. Ann Probab 12(1):243–255

McNeil AJ, Nešlehová J (2009) Multivariate Archimedean copulas, d-monotone functions and l_1-norm symmetric distributions. Ann Statist 37(5B):3059–3097

Merz M, Wüthrich MV (2007) Prediction error of the expected claims development result in the chain ladder method. Schweiz Aktuarver Mitt (1):117–137

Merz M, Wüthrich MV (2008a) Modelling the claims development result for solvency purposes. Conference Paper, ASTIN Colloquium, July 2008, Manchester

Merz M, Wüthrich MV (2008b) Stochastic Claims Reserving Methods in Insurance. Wiley, 2008, volume 435 de The Wiley Finance Series

Merz M, Wüthrich MV, Lysenko N (2008) Uncertainty in the claims development result in the chain ladder method. Accepted for publication in Scand Act J

Nelsen RB (2006) An introduction to copulas, 2nd edn. Springer Series in Statistics, Springer, New York

Ohlsson E, Lauzeningks J (2009) The one-year non-life insurance risk. Insurance Math Econom 45(2):203–208

Omelka M, Gijbels I, Veraverbeke N (2009) Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing. Ann Statist 37(5B):3023–3058

Plackett RL (1954) A reduction formula for normal multivariate integrals. Biometrika 41:351–360

Prohorov YV (1956) Convergence of random processes and limit theorems in probability theory. Teor Veroyatnost i Primenen 1:177–238

Quessy JF (2005) Methodologie et application des copules: Tests d'adequation, tests d'independance, et bornes pour la valeur-a-risque. ProQuest LLC, Ann Arbor, MI, thesis (Ph.D.)–Universite Laval (Canada)

Rémillard B, Scaillet O (2009) Testing for equality between two copulas. J Multivariate Anal 100(3):377–386

Rüschendorf L (1976) Asymptotic distributions of multivariate rank order statistics. Ann Statist 4(5):912–923

Segers J (2012) Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. Bernoulli 18(3):764–782

Shao QM, Yu H (1996) Weak convergence for weighted empirical processes of dependent sequences. Ann Probab 24(4):2098–2127

Shorack GR (1982) Kiefer's theorem via the Hungarian construction. Z Wahrsch Verw Gebiete 61(3):369–373

Sklar A (1959) Fonctions de répartition à n dimensions et leurs marges. Publ Inst Statist Univ Paris 8:229–231

Smirnov N (1936) Sur la distribution de ω^2 (Critérium de M.R. v. Mises). CR Acad Sci Paris 202:449–452

Smirnov N (1937) Sur la distribution de ω^2 (Critérium de M.R. v. Mises) (Russian/French summary). Mat Sbornik (NS) 2:973–993

Smirnov NV (1944) Approximate laws of distribution of random variables from empirical data. Uspehi Matem Nauk 10:179–206

Sukhatme S (1972) Fredholm determinant of a positive definite kernel of a special type and its application. Ann Math Statist 43:1914–1926

Taylor G (2000) Loss Reserving-An Actuarial Perspective. Boston: Kluwer Academic Press.

Taylor GC, Ashe FR (1983) Second moments of estimates of outstanding claims. J Econometrics 23(1):37–61

Trivedi PK, Zimmer DM (2005) Copula modeling: An introduction for practitioners. Found Trends Econom 1(1):1–111

Tsukahara H (2000) Empirical copulas and some applications. Research Report No 27 The Institute for Economic Studies, Seijo University

Tsukahara H (2005) Semiparametric estimation in copula models. Canad J Statist 33(3):357–375

Tsukahara H (2011) Erratum to *Semiparametric estimation in copula models*. Canad J Statist 39(4):734–735

Urrleman VD, Ikeghbali AN, Roncalli T (2000) Copulas approximation and new families. Groupe de Recherche Opérationelle, Crédit Lyonnais, France, Working Paper

van der Vaart AW, Wellner JA (1996) Weak convergence and empirical processes. Springer Series in Statistics, Springer-Verlag, New York, with applications to statistics

von Mises R (1931) Vorlesungen aus dem Gebiete der angewandten Mathematik. Bd. I: Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik.