



SAPIENZA
UNIVERSITÀ DI ROMA

Scuola di Dottorato in Scienze Statistiche
Dottorato di Ricerca in Scienze Attuariali
XXXII Ciclo

La valutazione del Non-Life Reserve Risk tramite le Vine Copule e il Collective Risk Model

Relatore
Prof. Fabio Grasso

Candidato
Alessandro Ricotta
Matricola: 1741681

Tesi presentata per il Dottorato di Ricerca
Anno Accademico 2018/2019

Sommario

La tesi propone, nel contesto delle assicurazioni danni, un approccio volto a quantificare il rischio di riservazione considerando congiuntamente molteplici segmenti di attività. Coerentemente con il framework normativo europeo, definito *Solvency II*, la metodologia si qualifica come un possibile modello interno parziale per la quantificazione del requisito patrimoniale da detenere a fini di solvibilità. Il reserve risk è valutato analizzando separatamente sia la struttura di dipendenza esistente tra le differenti linee di business, sia i rischi marginali rappresentativi del rischio di riservazione dei singoli segmenti di attività.

La dipendenza implicita tra le fonti di rischio è valutata attraverso un approccio di ricampionamento congiunto applicato ai triangoli di run-off della riserva sinistri di 4 linee di business di una reale compagnia di assicurazione danni. In seguito, la dipendenza è modellizzata ricorrendo alla tipologia di copule multivariate definita *vine copule*. Tale costrutto metodologico possiede una notevole flessibilità in quanto consente di scomporre una struttura di dipendenza in n ($n > 2$) dimensioni tramite copule bivariate. Nel caso studio preso in esame, in particolare, la dipendenza multivariata è stata studiata ricorrendo alla classe di vine copule definita D-vine.

Di contro, le variabili aleatorie unidimensionali rappresentative dei singoli rischi marginali di riservazione sono analizzate attraverso una metodologia Collective Risk Model; in tale contesto, la variabilità e la dipendenza dei parametri del modello sono quantificate ricorrendo ad un approccio bayesiano.

Infine, i rischi marginali dei singoli segmenti di attività sono aggregati tramite la vine copula stimata, permettendo di determinare il requisito patrimoniale di solvibilità afferente al reserve risk complessivo.

Dichiarazione

La presente tesi è oggetto del lavoro esclusivo dello scrivente e non contiene risultati svolti in collaborazione tranne come specificato nel testo. L'elaborato, o sue parti, non sono state presentate contemporaneamente presso l'Università di Roma "La Sapienza" o altra Università per conseguire altri titoli. Dichiaro, inoltre, di avere consultato personalmente tutte le fonti citate.

In fede

Alessandro Ricotta

Roma, 31 ottobre 2019

Ringraziamenti

Il presente lavoro è il traguardo di un percorso accademico raggiunto anche grazie al prezioso aiuto di persone conosciute sia nel mondo accademico che in quello lavorativo.

Desidero ringraziare il prof. Fabio Grasso per la disponibilità e il sostegno che mi ha sempre profuso durante il corso di dottorato e, in modo particolare, nella redazione della tesi. Ringrazio il prof. Nino Savelli per avermi incoraggiato, spronato e supportato ad intraprendere tale percorso. Egual pensiero va al prof. Gian Paolo Clemente.

Ringrazio Stefano Hajek, Claudio Chiani, Guerino Marini, Riccardo Casalini e Benoit Payre per la disponibilità che mi hanno dimostrato, nonché per il loro interesse nei confronti del mio lavoro. Ringrazio Michele Ottini per i costruttivi confronti che mi ha dato modo di condividere con lui.

Infine, un pensiero speciale va a Edoardo Luini, instancabile e sempre presente compagno di dottorato, con cui ho avuto il piacere e l'onore di collaborare.

Indice

	Pagina
1 Introduzione	7
1.1 Il reserve risk in Solvency II	8
1.2 Obiettivo e contributi della tesi	19
2 Dipendenza e copule tra due variabili aleatorie	22
2.1 La dipendenza tra due variabili casuali	22
2.2 Le copule	31
2.3 Distribuzioni ellittiche e relative copule	40
2.3.1 Distribuzione e copula Normale	43
2.3.2 Distribuzione e copula t-student	45
2.3.3 Distribuzione e copula di Kotz	48
2.3.4 Distribuzione e copula di Pearson Tipo II	49
2.3.5 Distribuzione e copula di Pearson Tipo VII	51
2.3.6 Distribuzione e copula Exponential Power	52
2.3.7 Distribuzione e copula di Laplace	54
2.4 Le copule archimedee	56
2.4.1 Gumbel-Hougaard	57
2.4.2 Ali-Mikhail-Haq	58
2.4.3 Clayton	59
2.4.4 Frank	60
2.4.5 Joe	61
2.4.6 BB1	62
2.4.7 BB6	63
2.4.8 BB7	64
2.4.9 BB8	64
2.5 Altre famiglie di copule	66
2.5.1 Fréchet	66
2.5.2 Marshall-Olkin	68
2.5.3 Farlie-Gumbel-Morgenstern	69
2.5.4 Plackett	70
3 La pair-copula construction e le vine copule	72
3.1 La pair-copula construction	73
3.2 Le vine copule	81

3.2.1	Il processo di stima delle vine copule	89
4	Caso studio: la quantificazione del non-life reserve risk	104
4.1	Il data set di riferimento	105
4.2	La modellizzazione della dipendenza tramite vine copule	107
4.2.1	Stima del D-vine	113
4.2.2	Stima del C-vine	129
4.2.3	Selezione della vine copula	142
4.3	La modellizzazione dei rischi marginali tramite Collective Risk Model	143
4.3.1	Il Collective Risk Model	144
4.3.1.1	La stima bayesiana della variabilità dei fattori di disturbo	146
4.3.1.2	La stima bayesiana della correlazione lineare tra i fattori di disturbo	155
4.3.1.3	Le distribuzioni marginali del reserve risk	158
4.4	La stima del requisito patrimoniale di solvibilità complessivo per il reserve risk	165
5	Conclusioni	176
	Bibliografia	179
A	Triangoli di run-off	185
B	MCMC: distribuzioni a posteriori	190
C	Gamma misturata	202
	Notazione	205

Capitolo 1

Introduzione

L'entrata in vigore della nuova disciplina comunitaria in tema di solvibilità delle imprese di assicurazione, rappresentata dal progetto *Solvency II*, ha determinato nel corso degli ultimi anni un particolare interesse verso modelli volti a valutare i rischi prevalenti a cui è sottoposta l'attività assicurativa. Il nuovo contesto normativo, infatti, pone particolare attenzione alla misurazione quantitativa dei rischi con lo scopo di valutarne gli effetti sul livello di solvibilità degli assicuratori. A fronte dell'attenzione posta dal legislatore sul profilo di rischio delle compagnie di assicurazione, la recente letteratura ha visto un notevole sviluppo di modelli stocastici, sia in ambito prettamente attuariale che di risk management, al fine di potere valutare e gestire nel modo più appropriato possibile la rischiosità a cui sono esposte le imprese di assicurazione.

Il rischio di sottoscrizione (c.d. *underwriting risk*) rappresenta il rischio caratteristico dell'attività assicurativa; quest'ultimo, nel contesto delle assicurazioni danni, è costituito dal rischio di tariffazione (*premium risk*), dal rischio di riserva (*reserve risk*), dal rischio di catastrofe (*catastrophe risk*) e, infine, dal rischio di riscatto (*lapse risk*). Il premium risk quantifica la rischiosità associata ai sinistri futuri dovuta alla variabilità nella distribuzione temporale, nella frequenza e nell'importo degli eventi assicurati. Il premium risk, quindi, misura il rischio di non adeguatezza dei premi rispetto ai sinistri futuri a cui sarà esposto l'assicuratore. Il reserve risk, di contro, afferisce alla rischiosità insita nei sinistri già accaduti che sorge dalle fluttuazioni casuali dell'andamento temporale e dell'ammontare dei sinistri liquidati. In particolare, la riserva sinistri è esposta ad una duplice fonte di aleatorietà; quella imputabile al processo stocastico liquidativo dei sinistri, nonché la variabilità di stima associata alla valutazione della riserva sinistri. Il rischio di catastrofe, invece, deriva da eventi estremi ed irregolari che non possono essere adeguatamente considerati nella quantificazione del premium e reserve risk. Infine, il rischio di estinzione anticipata, che usualmente risulta contenuto nel contesto delle assicurazioni danni, afferisce solo a obbligazioni assicurative contenenti opzioni contrattuali di riscatto; in tale contesto, la rischiosità deriva dalla possibile errata specificazione delle ipotesi adottate per valutare tali opzioni contrattuali, esponendo quindi l'assicuratore a potenziali perdite.

1.1 Il reserve risk in Solvency II

L'attuale framework normativo europeo in materia di accesso ed esercizio delle attività di assicurazione e riassicurazione è definito dal Regolamento Delegato UE 2015/35 della Commissione, nel seguito definito come Atti Delegati (Commissione Europea (2015)), parzialmente modificato dal Regolamento Delegato UE 2019/981 (Commissione Europea (2019)), che integrano la Direttiva 2009/138/CE del Parlamento Europeo e del Consiglio, denominata Solvency II (Parlamento Europeo (2009)).

La normativa definisce il requisito patrimoniale di solvibilità (*Solvency Capital Requirement*, SCR) che deve detenere un'impresa di assicurazione per potere esercitare l'attività assicurativa, come l'ammontare delle perdite potenziali a cui è esposto l'assicuratore, stimate ad un livello di confidenza del 99,5% e valutate su un orizzonte temporale annuo. Il SCR può essere stimato facendo uso dell'approccio metodologico, denominato *standard formula*, espressamente previsto dal legislatore, oppure ricorrendo ad un modello interno, anche parziale, sviluppato dall'impresa di assicurazione e approvato dall'autorità di vigilanza. Gli Atti Delegati individuano, attraverso la standard formula, la metodologia di calcolo del requisito patrimoniale da detenere a fini di solvibilità. Tale metodo prevede di quantificare il SCR seguendo un approccio modulare, identificando le fonti di rischio, oggetto di valutazione quantitativa, a cui sono esposti gli assicuratori. La formula standard prevede che il SCR sia dato dalla somma algebrica di tre componenti:

- il requisito patrimoniale di solvibilità di base (*Basic Solvency Capital Requirement*, BSCR);
- il requisito patrimoniale di solvibilità per il rischio operativo;
- la capacità di assorbimento di perdite delle riserve tecniche e delle imposte differite.

Il BSCR, a sua volta, è dato dall'*aggregazione* di sei macromoduli di rischio (Atti Delegati, art. 87):

- rischio di sottoscrizione dell'assicurazione non vita;
- rischio di sottoscrizione dell'assicurazione vita;
- rischio di sottoscrizione dell'assicurazione malattia;
- rischio di mercato;
- rischio di inadempimento della controparte (*default*);
- rischio relativo alle attività immateriali (tale rischio viene sommato ai precedenti).

In formule, il Basic Solvency Capital Requirement è pari a:

$$BasicSCR = \sqrt{\sum_{i,j} Corr_{i,j} \cdot SCR_i \cdot SCR_j} + SCR_{intagibles}$$

dove i pedici i, j individuano i 5 macromoduli relativi ai rischi di sottoscrizione non vita, vita, malattia, rischi di mercato e controparte.

La standard formula, aggregando le varie fonti di rischio, ammette un effetto di diversificazione tra le stesse; la normativa, tuttavia, si basa sulla semplificazione che prevede di aggregare i rischi come se fossero degli scarti quadratici medi. Come noto, tale approssimazione non è necessariamente soddisfatta nella pratica. I coefficienti di correlazione lineare da adottare in tale processo sono i seguenti:

Modulo i \ Modulo j	Modulo j				
	Mercato	Default	Vita	Malattia	Non vita
Mercato	1				
Default	0,25	1			
Vita	0,25	0,25	1		
Malattia	0,25	0,25	0,25	1	
Non vita	0,25	0,5	0	0	1

Tabella 1.1: Direttiva 2009/138/CE (allegato IV) - Coefficienti di correlazione lineare per i 5 macromoduli di rischio.

Ponendo l'attenzione sul rischio di sottoscrizione non vita, noto in letteratura come *non-life underwriting risk*, la normativa fissa il coefficiente di correlazione tra tale rischio e quello di controparte pari allo 0,5. Tale valore di correlazione è il più elevato tra tutte le combinazioni dei 5 moduli costituenti il BSCR ed evidenzia il notevole impatto a livello di solvibilità sul rischio di sottoscrizione non vita che si avrebbe in caso di default dei riassicuratori. La correlazione tra non-life underwriting risk e rischio di mercato, invece, è posta pari allo 0,25; infine, il rischio di sottoscrizione non vita e i rischi di sottoscrizione vita e malattia sono ipotizzati tra loro incorrelati.

Il rischio di sottoscrizione non vita, a sua volta, è definito tramite un approccio modulare volto a valutarne le singole componenti.

L'articolo 114 degli Atti Delegati individua tre sottomoduli:

- rischio di tariffazione e riservazione;
- rischio di catastrofe;

- rischio di estinzione anticipata;

In formule, il SCR del non-life underwriting risk risulta:

$$SCR_{non-life} = \sqrt{\sum_{i,j} CorrNL_{i,j} \cdot SCR_i \cdot SCR_j}$$

dove i termini SCR_i e SCR_j individuano i requisiti patrimoniali di solvibilità per i sottomoduli di cui sopra, mentre $CorrNL_{i,j}$ rappresenta i coefficienti di correlazione lineare definiti per aggregare le fonti di rischio. La correlazione tra i rischi di tariffazione e riservazione e quello di catastrofe è posta pari a 0,25, mentre il rischio di riscatto è ipotizzato essere incorrelato rispetto ai precedenti due sottomoduli.

Sottomodulo i \ Sottomodulo j	Tariffazione e riservazione	Catastrofe	Estinzione anticipata
Tariffazione e riservazione	1		
Catastrofe	0,25	1	
Estinzione anticipata	0	0	1

Tabella 1.2: Atti Delegati (art. 114) - Coefficienti di correlazione lineare del modulo di rischio di sottoscrizione non vita.

In particolare, come descritto precedentemente, la standard formula accorpa i rischi di tariffazione e riservazione. Il SCR associato a tale sottomodulo è definito come segue (art. 115 degli Atti Delegati):

$$SCR_{nl\ prem\ res} = 3 \cdot \sigma_{nl} \cdot V_{nl}$$

dove σ_{nl} è lo scostamento standard¹ associato al rischio di tariffazione e riservazione non vita, mentre V_{nl} è la rispettiva misura di volume. Quest'ultima è pari alla somma delle misure di volume per il rischio di tariffazione e riservazione delle singole linee di attività (*lines of business*, Lob) previste dal legislatore europeo per le assicurazioni danni:

- Segmento 1: Assicurazione e riassicurazione proporzionale sulla responsabilità civile risultante dalla circolazione di autoveicoli;
- Segmento 2: Altre assicurazioni e riassicurazioni proporzionali auto;

¹La grandezza σ_{nl} è una misura di variabilità non assoluta bensì relativa, in quanto adimensionale; tale parametro, sebbene definito comunemente come deviazione standard, identifica più propriamente un coefficiente di variazione.

- Segmento 3: Assicurazione e riassicurazione proporzionale marittima, aeronautica e trasporti;
- Segmento 4: Assicurazione e riassicurazione proporzionale contro l'incendio e altri danni ai beni;
- Segmento 5: Assicurazione e riassicurazione proporzionale sulla responsabilità civile generale;
- Segmento 6: Assicurazione e riassicurazione proporzionale di credito e di cauzione;
- Segmento 7: Assicurazione e riassicurazione proporzionale di tutela giudiziaria;
- Segmento 8: Assicurazione e riassicurazione proporzionale di assistenza;
- Segmento 9: Assicurazione e riassicurazione proporzionale di perdite pecuniarie di vario genere;
- Segmento 10: Riassicurazione non proporzionale danni a beni RC;
- Segmento 11: Riassicurazione non proporzionale marittima, aeronautica e trasporti;
- Segmento 12: Riassicurazione non proporzionale danni ai beni.

Per la generica Lob s , il parametro di volume è così ottenuto (art. 116 degli Atti Delegati),

$$V_s = (V_{(prem,s)} + V_{(res,s)}) \cdot (0,75 + 0,25 \cdot DIV_s)$$

dove

- $V_{(s,prem)}$ è la misura di volume, intesa come misura di esposizione al rischio, del premium risk per la Lob s , definita dalla seguente combinazione lineare:

$$V_{(prem,s)} = \max[P_s; P_{last,s}] + FP_{existing,s} + FP_{future,s}$$

in cui

- (i) P_s individua la stima dei premi che l'impresa acquisirà per l'esercizio successivo;
- (ii) $P_{last,s}$ individua i premi acquisiti nell'esercizio precedente;
- (iii) $FP_{existing,s}$ rappresenta il valore attuale atteso dei premi che verranno acquisiti dopo il primo anno in relazione ai contratti già esistenti;

- (iv) $FP_{future,s}$ rappresenta, per tutti i contratti la cui durata iniziale è inferiore o pari a un anno, il valore attuale atteso dei premi associati a polizze che verranno sottoscritte nei 12 mesi futuri, escludendo tuttavia i premi afferenti al primo anno di vita dei contratti. Di contro, per tutti i contratti la cui durata iniziale è superiore a un anno, l'ammontare è pari al 30% del valore attuale atteso dei premi che l'impresa acquisirà dopo i 12 mesi successivi.²

Tutti i volumi dei premi sono al netto delle cessioni in riassicurazione.

In alternativa, gli assicuratori possono determinare la misura di volume afferente al premium risk come segue:

$$V_{(prem,s)} = P_s + FP_{existing,s} + FP_{future,s}$$

purché, come prescritto dall'articolo 116 degli Atti Delegati, siano verificate tutte le seguenti condizioni,

- l'organo amministrativo, direttivo o di vigilanza dell'impresa di assicurazione o di riassicurazione ha deciso che i premi acquisiti dall'impresa nel segmento s nei 12 mesi successivi non eccederanno P_s ;
 - l'impresa di assicurazione ha stabilito efficaci meccanismi di controllo per garantire il rispetto dei limiti per i premi acquisiti di cui al punto precedente;
 - l'impresa di assicurazione ha comunicato all'autorità di vigilanza la decisione, motivandola opportunamente.
- $V_{(s,res)}$ è la misura di volume, intesa come misura di esposizione al rischio, del reserve risk del ramo s , definita come l'ammontare delle *best estimate* sinistri al netto degli importi recuperabili dai riassicuratori.
 - DIV_s è il fattore di diversificazione geografica del segmento s , definito dall'indice di Herfindhal:

$$DIV_s = \frac{\sum_r (V_{(prem,r,s)} + V_{(res,r,s)})^2}{(\sum_r V_{(prem,r,s)} + V_{(res,r,s)})^2}$$

dove r individua le diverse aree geografiche elencate nell'allegato III degli Atti Delegati.

²La riduzione del 70% del valore attuale atteso dei premi associati a polizze pluriennali è stata introdotta dal Regolamento Delegato 2019/981.

	Regione
1	Europa settentrionale
2	Europa occidentale
3	Europa orientale
4	Europa meridionale
5	Asia centrale e occidentale
6	Asia orientale
7	Asia meridionale e sudorientale
8	Oceania
9	Africa settentrionale
10	Africa meridionale
11	America settentrionale esclusi gli USA
12	America caraibica e centrale
13	Sudamerica orientale
14	Sudamerica settentrionale, meridionale e occidentale
15	Stati Uniti d'America nordorientali
16	Stati Uniti d'America sudorientali
17	Stati Uniti d'America centroccidentali
18	Stati Uniti d'America occidentali

Tabella 1.3: Atti Delegati (allegato III) - Regioni per il calcolo del fattore di diversificazione geografica per il premium e reserve risk.

Il beneficio derivante dalla diversificazione geografica può essere al più pari al 25% ed è nullo nel caso in cui l'assicuratore operi in una sola area geografica. Il legislatore, inoltre, prevede che l'effetto di diversificazione geografica non debba essere considerato in merito alle Lob 6, 10, 11 e 12, che individuano rispettivamente l'assicurazione e riassicurazione proporzionale relativa a credito e cauzione e le linee di business associate alla riassicurazione non proporzionale esercitata dall'impresa. Il beneficio di diversificazione geografica, infine, non deve essere considerato nel caso in cui l'impresa stimi i valori dello scostamento standard ricorrendo a parametri specifici interni (*undertaking-specific parameters*, USP).

Lo scostamento standard attinente al premium e reserve risk, σ_{nl} , è definito dall'articolo 117 degli Atti Delegati:

$$\sigma_{nl} = \frac{1}{V_{nl}} \cdot \sqrt{\sum_{s,t} CorrS_{s,t} \cdot \sigma_s \cdot V_s \cdot \sigma_t \cdot V_t}$$

dove σ_s , σ_t , V_s e V_t rappresentano la variabilità e il parametro di volume del premium e reserve risk per le varie linee di business s e t .

I parametri di correlazione lineare, $CorrS_{s,t}$, previsti dalla normativa per aggregare il rischio di tariffazione e riservazione delle differenti Lob, riportati nella seguente tabella, assumono tutti valori pari a 0,25 o 0,50.

$s \backslash t$	1	2	3	4	5	6	7	8	9	10	11	12
1	1											
2	0,5	1										
3	0,5	0,25	1									
4	0,25	0,25	0,25	1								
5	0,5	0,25	0,25	0,25	1							
6	0,25	0,25	0,25	0,25	0,5	1						
7	0,5	0,5	0,25	0,25	0,5	0,5	1					
8	0,25	0,5	0,5	0,5	0,25	0,25	0,25	1				
9	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	1			
10	0,25	0,25	0,25	0,25	0,5	0,5	0,5	0,25	0,25	1		
11	0,25	0,25	0,5	0,5	0,25	0,25	0,25	0,25	0,5	0,25	1	
12	0,25	0,25	0,25	0,5	0,25	0,25	0,25	0,5	0,25	0,25	0,25	1

Tabella 1.4: Atti Delegati (allegato IV) - Coefficienti di correlazione lineare delle Lob per il rischio di tariffazione e riservazione dell'assicurazione non vita.

Infine, per potere valutare σ_{nl} occorre stimare lo scostamento standard, σ_s , del premium e reserve risk della singola linea di business s . Il legislatore stabilisce che tale variabilità sia definita aggregando i singoli scostamenti standard del rischio di tariffazione e di riservazione³:

$$\sigma_s = \frac{\sqrt{\sigma_{(prem,s)}^2 \cdot V_{(prem,s)}^2 + \sigma_{(res,s)}^2 \cdot V_{(res,s)}^2 + 2 \cdot \sigma_{(prem,s)} \cdot V_{(prem,s)} \cdot \sigma_{(res,s)} \cdot V_{(res,s)}}}{V_{(prem,s)} + V_{(res,s)}}$$

La normativa fornisce i seguenti valori di variabilità per il premium ($\sigma_{(prem,s)}$) e reserve risk ($\sigma_{(res,s)}$):

³Si noti che lo Studio di Impatto Quantitativo 2, QIS 2 (CEIOPS (2006)), prevedeva di valutare separatamente il reserve risk e premium risk a livello complessivo di impresa (tramite aggregazione delle singole Lob) e successivamente di aggregare tali due fonti di rischio.

Segmento	Scostamento standard per il rischio di tariffazione	Scostamento standard per il rischio di riservazione
1	10%	9%
2	8%	8%
3	15%	11%
4	8%	10%
5	14%	11%
6	19%	17,2%
7	8,3%	5,5%
8	6,4%	22%
9	13%	20%
10	17%	20%
11	17%	20%
12	17%	20%

Tabella 1.5: Atti Delegati (allegato II) e successive integrazioni del Regolamento Delegato 2019/981 (allegato I) - Scostamento standard delle singole Lob per il sottomodulo di rischio di tariffazione e riservazione per l'assicurazione non vita.

Gli Atti Delegati quantificano lo scostamento standard per il rischio di tariffazione compreso tra il 6,4% e il 19%. Inoltre, in merito ai segmenti relativi all'assicurazione e riassicurazione proporzionale sulla responsabilità civile risultante dalla circolazione di autoveicoli, assicurazione incendio e altri danni ai beni e responsabilità civile generale (Lob 1, 4 e 5), il legislatore prevede che in caso di riassicurazione passiva non proporzionale di tipo excess of loss, lo scostamento standard possa essere ridotto in misura del 20%. Considerando il reserve risk, invece, il parametro di variabilità per le diverse Lob varia tra il 5,5% e il 22%, risultando più elevato rispetto alla variabilità del premium risk, ad eccezione dei segmenti 3, 5, 6 e 7. In tale contesto, inoltre, non è contemplato alcun aggiustamento in funzione della struttura riassicurativa dell'impresa.

In definitiva, la variabilità associata al rischio di tariffazione e riservazione per la singola linea di business è ottenuta aggregando, secondo la logica della correlazione lineare e con coefficiente pari a 0,5, gli scostamenti standard del reserve e premium risk. Al fine di fare emergere l'equivalenza tra i parametri di variabilità adottati nella standard formula, denominati dalla normativa con σ_s , e il coefficiente di variazione della somma di due variabili casuali, si può osservare quanto segue.

Il coefficiente di variazione di una variabile aleatoria (v.a.⁴) è definito come il

⁴Nel seguito, nelle formule la tilde denoterà le variabili aleatorie.

rapporto tra la deviazione standard e il valore atteso della v.a.:

$$CV(\tilde{X}) = \frac{std(\tilde{X})}{E(\tilde{X})} \quad \text{con } E(\tilde{X}) \neq 0$$

dove std e E definiscono rispettivamente gli operatori standard deviation e valore atteso della variabile casuale. Il coefficiente di variazione è una misura di variabilità *relativa* in quanto grandezza adimensionale. Considerando la somma di due variabili casuali, vale:

$$\begin{aligned} CV(\tilde{X} + \tilde{Y}) &= \frac{std(\tilde{X} + \tilde{Y})}{E(\tilde{X} + \tilde{Y})} = \frac{\sqrt{std(\tilde{X})^2 + std(\tilde{Y})^2 + 2 \cdot \rho \cdot std(\tilde{X}) \cdot std(\tilde{Y})}}{E(\tilde{X}) + E(\tilde{Y})} \\ &= \frac{\sqrt{CV(\tilde{X})^2 E(\tilde{X})^2 + CV(\tilde{Y})^2 E(\tilde{Y})^2 + 2\rho CV(\tilde{X})E(\tilde{X})CV(\tilde{Y})E(\tilde{Y})}}{E(\tilde{X}) + E(\tilde{Y})} \end{aligned}$$

dove ρ rappresenta il coefficiente di correlazione lineare di Pearson, definito per due generiche v.a. \tilde{X}_1 e \tilde{X}_2 , con media e varianze finite, come rapporto tra la covarianza delle v.a. e il prodotto delle rispettive standard deviation:

$$\rho(\tilde{X}_1; \tilde{X}_2) = \frac{cov(\tilde{X}_1; \tilde{X}_2)}{std(\tilde{X}_1)std(\tilde{X}_2)}$$

Modificando la notazione in coerenza con quella adottata dagli Atti Delegati,

- $CV(\tilde{X}) = \sigma_{prem}, \quad CV(\tilde{Y}) = \sigma_{res};$
- $E(\tilde{X}) = V_{prem}, \quad E(\tilde{Y}) = V_{res};$
- $\rho = 0,5$

è possibile riscrivere la formula del coefficiente di variazione della somma di due v.a. come segue:

$$\sigma = \frac{\sqrt{\sigma_{(prem)}^2 \cdot V_{(prem)}^2 + \sigma_{(prem)} \cdot V_{(prem)} \cdot \sigma_{(res)} \cdot V_{(res)} + \sigma_{(res)}^2 \cdot V_{(res)}^2}}{V_{(prem)} + V_{(res)}}$$

che coincide con quella prescritta dalla normativa comunitaria per definire lo scostamento standard del rischio di tariffazione e riservazione per la generica Lob s .

Da quanto esposto precedentemente, emerge che nel calcolo del premium e reserve risk, come definito dalla standard formula, gli unici parametri caratteristici dell'impresa risultano essere le misure di volume associate rispettivamente ai rischi di tariffazione e riservazione. Gli scostamenti standard delle singole linee di business, nonché i coefficienti di correlazione tra le varie Lob, infatti, sono

predeterminati dalla normativa e unici a livello di mercato europeo.

Il legislatore, al fine di rendere la standard formula più aderente alle specificità della singola impresa, fornisce agli assicuratori la facoltà di stimare con dati propri alcuni parametri rientranti nel calcolo del SCR (si vedano l'articolo 218 e l'allegato XVII degli Atti Delegati). Tale metodo di stima è definito *undertaking specific parameters* (USP) e la sua adozione da parte delle imprese è sottoposta, come per il modello interno, ad approvazione dell'autorità di vigilanza. L'approccio USP comprende anche gli scostamenti standard dei rischi di tariffazione e riservazione delle assicurazioni non vita. Tale metodologia prevede che i parametri di variabilità del premium e reserve risk da utilizzare nel calcolo dalla standard formula per il SCR siano determinati come media ponderata tra i parametri stimati tramite i dati interni dell'impresa e gli scostamenti standard definiti negli Atti Delegati:

$$\sigma_{(s,USP)} = c \cdot \hat{\sigma}_s + (1 - c) \cdot \sigma_s$$

dove $\hat{\sigma}_s$ individua la variabilità (per il premium o reserve risk) della Lob s definita in base ai dati interni della compagnia, σ_s è il rispettivo scostamento standard definito negli Atti Delegati e c costituisce il parametro di ponderazione tra le due grandezze precedenti. Il parametro di credibilità utilizzato per ponderare la *sigma* di mercato con quella specifica dell'impresa varia in funzione della profondità temporale (N) della serie storica di dati utilizzati nel processo di stima. Per le linee di business 1, 5 e 6, relative alle assicurazioni sulla responsabilità civile risultante dalla circolazione di autoveicoli, responsabilità civile generale e credito e cauzione, il parametro di credibilità associato alla variabilità specifica dell'impresa è il seguente:

N	5	6	7	8	9	10	11	12	13	14	≥ 15
c	34%	43%	51%	59%	67%	74%	81%	87%	92%	96%	100%

Per le restanti linee di attività, invece, il parametro di ponderazione risulta essere,

N	5	6	7	8	9	≥ 10
c	34%	51%	67%	81%	92%	100%

A fronte di una adeguata profondità storica di dati, quindi, il metodo di stima USP permette di utilizzare nel calcolo del SCR il parametro di variabilità stimato esclusivamente con i dati dell'impresa. In merito al reserve risk, la normativa prevede due possibili approcci di stima basati sui dati interni dell'assicuratore⁵. La prima, che di fatto risulta metodologicamente equivalente a quella utilizzata per il premium risk, individua un modello in cui la variabile indipendente (ipotizzata

⁵Per una accurata trattazione della modellistica e delle ipotesi sottostanti alle metodologie USP si rimanda a De Felice e Moriconi (2016).

essere lognormale) è il costo ultimo, stimato all'inizio dell'anno, dei sinistri accaduti negli esercizi precedenti, mentre la variabile dipendente risulta essere il costo ultimo degli stessi sinistri stimato alla fine dell'esercizio. L'obiettivo è quello di quantificare lo scarto quadratico medio della variabile dipendente sulla base delle relazioni funzionali definite dal modello rispetto alla variabile indipendente.

La seconda metodologia per calcolare lo scostamento standard del reserve risk utilizza il modello sviluppato da Wüthrich e Merz (2008). In tale contesto la variabile aleatoria oggetto di studio è il run-off della riserva sinistri, denominato dagli autori come *claims development result* (CDR). Tale grandezza è definita come il saldo tra la riserva sinistri stimata all'inizio dell'esercizio, i pagamenti liquidati nell'anno e la nuova stima della riserva residua effettuata al termine dell'esercizio. Gli autori quantificano in forma analitica, tramite la formula che prende il loro nome, la variabilità su un orizzonte temporale annuo associata al claims development result rispondente al metodo deterministico Chain-Ladder di stima della riserva sinistri. La varianza della v.a. CDR, definita *mean square error of prediction* (MSEP), consta sia della variabilità di processo sottostante allo sviluppo stocastico dei pagamenti, che della variabilità di parametro, associata all'incertezza di stima dei parametri del modello. La normativa definisce lo scostamento standard del reserve risk, calibrato sui dati interni dell'impresa, come il rapporto tra la radice quadrata del MSEP e il valore atteso della riserva sinistri determinata tramite metodo Chain-Ladder; il parametro $\hat{\sigma}_s$ risulta, quindi, a tutti gli effetti un coefficiente di variazione.

Le principali limitazioni di cui soffre l'impostazione di calcolo del requisito patrimoniale di solvibilità per il premium e reserve risk, così come definito dalla standard formula, sono le seguenti.

- L'aggregazione delle tipologie di rischio (premium e reserve) e delle fonti di rischio (Lob) è definita sulla base del concetto di *correlazione lineare*, senza considerare altre possibili tipologie di dipendenza esistenti tra i rischi.
- Il requisito patrimoniale di solvibilità è quantificato tramite un moltiplicatore predeterminato e fisso (pari a 3) che viene applicato al parametro di variabilità e alla misura di esposizione del rischio in esame.
- Gli scostamenti standard sottostanti le fonti di rischio, nonché i parametri di correlazione tra i rischi, sono individuati dal legislatore e determinati a livello di mercato europeo.

In merito al primo punto, l'approccio metodologico previsto dalla standard formula per aggregare i rischi è valido e coerente solo nel contesto di v.a. multivariate ellittiche. In tale ambito, infatti, le v.a. sono univocamente definite dal vettore delle medie, dalla matrice delle covarianze e dalla funzione caratteristica. I valori attesi e le varianze sono grandezze specifiche delle singole v.a., ossia delle variabili stocastiche unidimensionali; la dipendenza sottostante alla distribuzione multivariata, invece, è determinata dalla matrice delle covarianze e dalla funzione

generatrice. In tali modelli, quindi, la matrice di correlazione assume in modo naturale l'accezione di parametro, peculiarità non valida in altre strutture di modelli multivariati. La limitazione individuata dal secondo punto, invece, deriva dalla rigidità nell'adottare un predeterminato moltiplicatore per individuare il *worst case scenario*, definito come differenza tra il quantile 99,5% ed il valore atteso del rischio in esame a cui è esposto l'assicuratore. Come noto, infatti, il valore del quantile ad un predeterminato livello di confidenza e, quindi, anche la distanza tra quest'ultimo e il valore atteso, varia in funzione della famiglia distributiva e delle specificità della variabile aleatoria considerata, quali ad esempio l'asimmetria e la leptocurticità della distribuzione. Infine, l'ultima limitazione riguarda i parametri di variabilità; questi ultimi sono identici per tutte le imprese, non rispecchiando, quindi, le specificità del singolo assicuratore. L'approccio USP previsto dal legislatore sopperisce, in parte, a tale esigenza; la stima basata sui parametri specifici dell'impresa, tuttavia, risulta essere predeterminata dal punto di vista metodologico, non lasciando, dunque, spazio alle imprese di valutare il modello che meglio si adatta a rappresentare il proprio profilo di rischio. In aggiunta, anche i coefficienti di correlazione lineare adottati nella standard formula risultano essere non modificabili dagli assicuratori. A livello di singola impresa, dunque, non vi è garanzia che la dipendenza, seppur lineare, definita negli Atti Delegati sia effettivamente rappresentativa di quella esistente tra le varie fonti di rischio.

1.2 Obiettivo e contributi della tesi

Lo scopo della tesi è quello di proporre, nel contesto delle assicurazioni danni, un approccio volto a quantificare il rischio di riservazione considerando congiuntamente molteplici linee di business. Coerentemente con il framework normativo previsto da Solvency II, la metodologia si qualifica come un possibile modello interno parziale. Il reserve risk è valutato studiando separatamente sia la struttura di dipendenza tra le varie linee di attività, sia modellizzando le variabili aleatorie marginali rappresentative della distribuzione del rischio di riservazione dei singoli rami.

La dipendenza esistente tra le linee di business è analizzata ricorrendo ad una particolare tipologia di copule multivariate, definite *vine copule*. Nell'ultima decade, la teoria e l'applicazione delle vine copule ha riscontrato notevole interesse soprattutto nel mondo finanziario⁶; di contro, le applicazioni in ambito attuariale sono state particolarmente ridotte e quasi assenti, da ciò che è noto allo scrivente, nel contesto del reserve risk. Obiettivo primario del presente elaborato è quello di declinare il costruito metodologico delle vine copule al fine di quantificare il rischio di riservazione delle assicurazioni danni. Le vine copule, inizialmente proposte da Joe (1996), Bedford e Cooke (2001, 2002), permettono di modellizzare

⁶In Brechmann e Schepsmeier (2013) è riportata una estesa rassegna bibliografica in merito.

in modo flessibile la dipendenza esistente tra n ($n > 2$) fonti di rischio; la dipendenza multivariata, infatti, è definita tramite fattorizzazione in copule bivariate (definita *pair-copula construction*, PCC). La scomposizione tramite copule bidimensionali permette, quindi, di costruire una copula multivariata estremamente flessibile in grado di cogliere le differenti tipologie di dipendenza esistenti tra i vari rischi marginali. I vine, inoltre, possiedono il beneficio delle copule che permette di modellizzare separatamente la struttura di dipendenza esistente tra le fonti di rischio e i modelli sottostanti ai singoli rischi marginali.

Nella tesi, il data set di partenza utilizzato per valutare il reserve risk è costituito dai triangoli di run-off⁷ del numero e ammontare dei sinistri pagati afferenti a 4 linee di business di una impresa di assicurazione danni. In particolare, i segmenti di business considerati riguardano la responsabilità civile derivante dalla circolazione di autoveicoli, la responsabilità civile generale, l'assicurazione incendio ed eventi naturali e l'assicurazione altri danni ai beni. Ricorrendo ad un approccio *bootstrap dipendente* è stato possibile estrarre la dipendenza implicita tra i triangoli di run-off delle differenti Lob. In seguito, ricorrendo alla classe delle vine copule definita D-vine (*drawable vine*), è stata costruita la copula multivariata che modella la dipendenza tra le singole linee di business. Successivamente, l'analisi si è spostata sulla modellizzazione delle variabili aleatorie marginali rappresentative del reserve risk dei differenti segmenti di attività. Queste ultime sono state definite seguendo l'approccio Collective Risk Model (CRM) come sviluppato in Ricotta e Luini (2019). In tale modello, la componente che quantifica la variabilità insita nella stima dei parametri, definita in letteratura come *parameter risk*, è calibrata ricorrendo ad un approccio bayesiano; l'obiettivo è quello di tenere in debita considerazione l'informazione storica associata all'incertezza di parametro implicita nei triangoli di run-off. In modo analogo, inoltre, la metodologia bayesiana è utilizzata per quantificare il parametro di correlazione lineare tra le variabili del numero (*frequency*) e importo (*severity*) del modello CRM. Infine, il SCR complessivo afferente al reserve risk dell'impresa di assicurazione è stato ottenuto aggregando i rischi marginali delle singole Lob tramite la struttura di dipendenza multivariata definita dal D-vine precedentemente stimato.

La tesi è organizzata come segue. Il capitolo 2 espone i concetti di dipendenza tra due variabili aleatorie; in seguito, si introduce la definizione di copula, ponendo l'attenzione sulle principali famiglie bivariate esistenti in letteratura. Il capitolo 3 presenta da un punto di vista teorico la *pair-copula construction*, le vine copule e i metodi di stima. Nel capitolo 4 si sviluppa il caso studio costruito considerando i triangoli di run-off di 4 Lob provenienti da una reale impresa di assicurazione. Il costruito metodologico delle vine copule è utilizzato per modellizzare la dipendenza insita nelle linee di business. Le *pair-copule* sono stimate e selezionate ricorrendo al metodo della massima verosimiglianza e al criterio di informazione di Akaike (AIC); quest'ultimo, inoltre, è utilizzato per valutare la vine copula più appropriata tra le $n!$ possibili configurazioni esistenti di *regular*

⁷Il data set deriva da dati reali opportunamente modificati al fine di preservarne la discrezionalità.

vine. Infine, i rischi marginali sono modellizzati ricorrendo ad un approccio CRM. Questi ultimi sono aggregati tra loro tramite la vine copula stimata, che risulta essere un D-vine, permettendo così di quantificare il requisito patrimoniale di solvibilità complessivo afferente al reserve risk. Le conclusioni chiudono la tesi.

I principali contributi e aspetti innovativi del presente elaborato sono i seguenti.

- A differenza della letteratura prevalente, che ha visto una notevole applicazione delle vine copule in ambito finanziario, nel presente lavoro le vine copule sono utilizzate in ambito attuariale con lo scopo di aggregare il reserve risk afferente a molteplici linee di business.
- La vine copula utilizzata per aggregare il rischio di riservazione delle varie Lob è calibrata su dati ottenuti tramite i triangoli di run-off dei rami considerati. In particolare, la dipendenza insita in tale data set è estratta facendo uso di un approccio bootstrap dipendente. Il campionamento congiunto dei triangoli di run-off ha permesso, quindi, di valutare la dipendenza implicitamente contenuta nei dati di partenza. Tale metodologia permette di ottenere una matrice di dati dipendenti tra loro di numerosità arbitraria, facendo venire meno il problema, tipico del contesto in esame, della limitata numerosità delle informazioni (si pensi ad esempio alla serie storica dei *run-off ratio*⁸ di una Lob). Nella tesi, quindi, a differenza degli approcci usualmente utilizzati nella pratica attuariale, la dipendenza tra il reserve risk di più Lob è valutata utilizzando direttamente i triangoli di run-off, senza ricorrere a indicatori sintetici rappresentativi del rischio di riservazione sottostante, quale ad esempio il già citato run-off ratio.
- Le variabili aleatorie marginali rappresentative del reserve risk delle singole Lob sono modellizzate ricorrendo ad un approccio CRM in cui il parameter risk è calibrato tramite una metodologia bayesiana. Tale metodo permette di stimare la variabilità dei parametri del modello con l'obiettivo di considerare l'intera informazione storica presente nei triangoli di run-off. Il metodo bayesiano, inoltre, consente di stimare la dipendenza lineare esistente tra i parametri caratteristici del modello CRM, dati dalla frequency e severity dei sinistri.

⁸Per run-off ratio si intende il rapporto tra il run-off, ossia lo smontamento della riserva, e il valore di riserva iniziale.

Capitolo 2

Dipendenza e copule tra due variabili aleatorie

Il capitolo ha lo scopo di presentare le misure di dipendenza, tra due variabili aleatorie, maggiormente utilizzate nella pratica attuariale, quali il coefficiente di correlazione lineare di Pearson, le misure di concordanza e quelle di dipendenza di coda. Successivamente, viene esposto il concetto di copula, seguito dalle principali famiglie bivariate esistenti in letteratura, quali le copule ellittiche e quelle archimedee.

2.1 La dipendenza tra due variabili casuali

Una misura di dipendenza ha l'obiettivo di cogliere, con determinate specificità, la relazione esistente tra due variabili aleatorie. Le misure presentate nel seguito afferiscono a tre diverse tipologie di legami che possono sussistere tra una coppia di variabili casuali, quali il grado di dipendenza lineare, quello di concordanza e la dipendenza mostrata nei valori estremi delle variabili aleatorie considerate.

Preliminarmente è opportuno richiamare i concetti di indipendenza e perfetta dipendenza positiva e negativa.

Due variabili aleatorie sono tra loro indipendenti se e solo se vale la seguente relazione:

$$H_{\tilde{X}_1, \tilde{X}_2}(x_1, x_2) = F_{\tilde{X}_1}(x_1)F_{\tilde{X}_2}(x_2) \quad \forall (x_1, x_2) \in \mathbb{R}^2$$

dove $H_{\tilde{X}_1, \tilde{X}_2}(x_1, x_2)$ rappresenta la funzione di ripartizione congiunta, mentre $F_{\tilde{X}_1}(x_1)$ e $F_{\tilde{X}_2}(x_2)$ individuano le funzioni di ripartizione marginali¹. In modo analogo, la relazione precedente può essere espressa in termini di funzione di densità:

¹Al fine di semplificare la notazione, nel seguito si adotterà la seguente dicitura per le funzioni di ripartizione e densità: $F_{\tilde{X}}(x) = F(x)$, $f_{\tilde{X}}(x) = f(x)$, restando inteso, se non diversamente specificato, che la variabile casuale considerata risulta essere \tilde{X} , la cui funzione di ripartizione o densità è calcolata nel punto x .

$$h(x_1, x_2) = f(x_1)f(x_2) \quad \forall (x_1, x_2) \in \mathbb{R}^2$$

Di contro, il concetto di perfetta dipendenza positiva tra le v.a. \tilde{X}_1 e \tilde{X}_2 è espresso dalla proprietà di *comonotonicità*. In particolare, due variabili aleatorie sono dette comonotone se una è funzione crescente dell'altra:

$$\tilde{X}_1 = \psi(\tilde{X}_2)$$

dove $\psi(\cdot)$ rappresenta una funzione crescente. Una importante conseguenza della comonotonicità risulta essere l'additività dei quantili delle variabili casuali. Dato il quantile di ordine α , q_α , vale:

$$q_\alpha(\tilde{X}_1 + \tilde{X}_2) = q_\alpha(\tilde{X}_1) + q_\alpha(\tilde{X}_2)$$

Situazione speculare rispetto al caso di perfetta dipendenza positiva è quella definita di *contromonotonicità*; in tale circostanza, la v.a. \tilde{X}_1 risulta essere funzione decrescente di \tilde{X}_2 :

$$\tilde{X}_1 = \psi(\tilde{X}_2)$$

dove $\psi(\cdot)$ individua una funzione decrescente.

Il coefficiente di correlazione lineare di Pearson (ρ) rappresenta una delle misure di dipendenza più note e permette di quantificare il grado di relazione lineare esistente tra due variabili casuali. Date \tilde{X}_1 e \tilde{X}_2 , dotate di media e varianza finite, il coefficiente di correlazione lineare è definito come segue:

$$\rho(\tilde{X}_1; \tilde{X}_2) = \frac{\text{cov}(\tilde{X}_1; \tilde{X}_2)}{\sigma(\tilde{X}_1)\sigma(\tilde{X}_2)} \quad (2.1.1)$$

dove $\text{cov}(\tilde{X}_1; \tilde{X}_2) = E[(\tilde{X}_1 - E[\tilde{X}_1])(\tilde{X}_2 - E[\tilde{X}_2])]$ rappresenta la covarianza tra le due v.a., mentre σ individua lo scarto quadratico medio della variabile considerata.

Il valore di ρ è compreso, per costruzione, nell'intervallo $[-1, 1]$. In particolare, $|\rho(\tilde{X}_1; \tilde{X}_2)| = 1$ se le v.a. sono perfettamente dipendenti tra loro in modo lineare, quindi $\tilde{X}_1 = a + b\tilde{X}_2$ con $a, b \in \mathbb{R}$ e $b > 0$ nel caso di dipendenza positiva mentre $b < 0$ nel caso di dipendenza negativa. Inoltre, per $b_1 > 0$ e $b_2 > 0$ vale,

$$\rho(a_1 + b_1\tilde{X}_1; a_2 + b_2\tilde{X}_2) = \rho(\tilde{X}_1; \tilde{X}_2)$$

Il coefficiente di correlazione lineare, quindi, è invariante rispetto a trasformazioni lineari strettamente crescenti delle variabili aleatorie. Il valore di ρ , tuttavia, non rispetta la proprietà di invarianza rispetto a trasformazioni non lineari strettamente crescenti delle v.a.:

$$\rho(\psi(\tilde{X}_1); \psi(\tilde{X}_2)) \neq \rho(\tilde{X}_1; \tilde{X}_2)$$

dove $\psi(\cdot)$ individua una funzione non lineare strettamente monotona crescente.

Esempio. Sia $H(x_1, x_2)$ la funzione di ripartizione della variabile bidimensionale Normale Standard di parametro $\rho = 0,5$ e siano $F(x_1)$ e $F(x_2)$ le funzioni di ripartizione delle rispettive variabili marginali Normali Standard. Per costruzione \tilde{X}_1 e \tilde{X}_2 risultano correlate tra loro con $\rho = 0,5$. Si consideri come funzione strettamente monotona crescente su tutto il dominio \mathbb{R} la funzione cubica $\psi(x) = x^3$. È agevole mostrare che il coefficiente di correlazione lineare calcolato sulle variabili trasformate, $\tilde{Y}_1 = \tilde{X}_1^3$ e $\tilde{Y}_2 = \tilde{X}_2^3$, non coincide con il valore originario di 0,5.

$$\rho(\psi(\tilde{Y}_1); \psi(\tilde{Y}_2)) = \rho(\tilde{X}_1^3; \tilde{X}_2^3) = \frac{\text{cov}(\tilde{X}_1^3; \tilde{X}_2^3)}{\sigma(\tilde{X}_1^3)\sigma(\tilde{X}_2^3)} = 0,35$$

dove $\text{cov}(\tilde{X}_1^3; \tilde{X}_2^3) = \int_{\mathbb{R}^2} (H(x_1, x_2) - F(x_1)F(x_2)) 3x_1^2 3x_2^2 dx_1 dx_2$ (Cuadras (2002)).

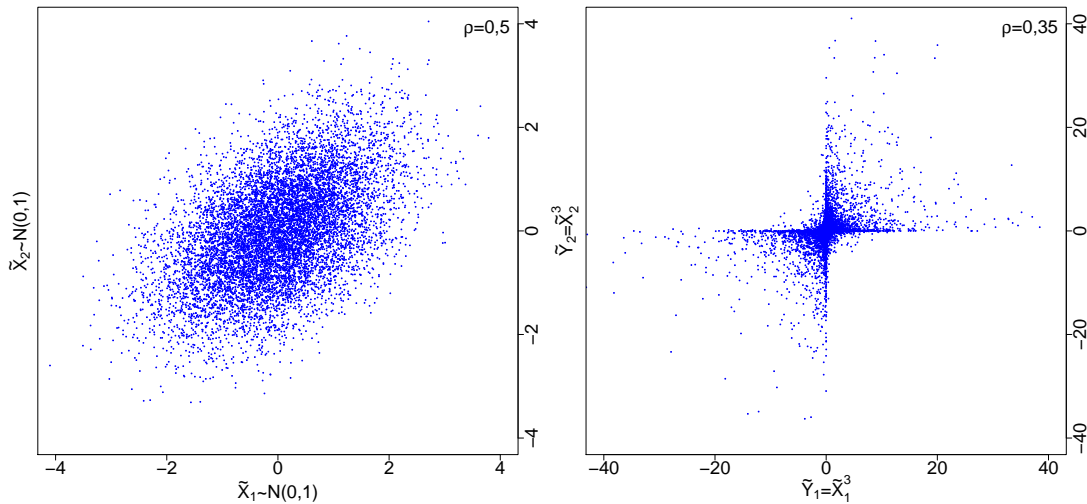


Figura 2.1: Scatterplot ottenuto con 10.000 realizzazioni della v.a. bidimensionale Normale Standard di parametro $\rho = 0,5$ (grafico di sinistra) e relativo scatterplot ottenuto con i valori trasformati tramite la funzione cubica $\psi(x) = x^3$ (grafico di destra).

■

In generale non è sempre possibile ottenere un coefficiente di correlazione lineare pari a -1 e +1; i valori estremi ottenibili di ρ , infatti, possono essere compresi nell'intervallo $-1 \leq \rho_{min} \leq 0 \leq \rho_{max} \leq 1$. L'ampiezza dell'intervallo $[\rho_{min}, \rho_{max}]$ dipende dalle specificità delle v.a. in esame; il coefficiente di correlazione, quindi, dipende dalle caratteristiche distributive delle variabili aleatorie marginali. La correlazione tra due v.a. è pari a ρ_{min} e ρ_{max} se le variabili sono rispettivamente

contromonotone e comonotone; la relazione inversa, tuttavia, non è necessariamente soddisfatta. In situazioni di perfetta dipendenza, inoltre, il coefficiente di correlazione lineare non assume necessariamente valori elevati; tanto più la relazione tra due variabili è distante dalla linearità e minore sarà la capacità del coefficiente di correlazione di cogliere la dipendenza tra le variabili aleatorie stesse.

Esempio. Siano \tilde{X} e \tilde{Y} due v.a. costruite come segue: $\tilde{X} \sim N(0, 1)$ e $\tilde{Y} = \tilde{X}^3$. Risulta evidente che le due v.a. siano tra loro comonotone in quanto la variabile \tilde{Y} è funzione strettamente crescente di \tilde{X} . Per determinare il coefficiente di correlazione di Pearson tra le due v.a. è necessario quantificare la varianza della v.a. \tilde{Y} e successivamente il termine di covarianza tra le due v.a. In particolare, la varianza del cubo della variabile Normale Standard è pari a:

$$\begin{aligned} \text{var}(\tilde{Y}) &= E(\tilde{Y}^2) - E(\tilde{Y})^2 = E[(\tilde{X}^3)^2] - [E(\tilde{X}^3)]^2 \\ &= E(\tilde{X}^6) - 0 = 15 \end{aligned}$$

in quanto $E(\tilde{X}^n) = \sigma^n(n-1)!!$, da cui $E(\tilde{X}^6) = \sigma^6(5)!! = 15$, dove $c!!$ rappresenta il doppio fattoriale² della grandezza c .

Il termine di covarianza, invece, risulta essere il seguente:

$$\text{cov}(\tilde{X}; \tilde{Y}) = E(\tilde{X}\tilde{Y}) - E(\tilde{X})E(\tilde{Y}) = E(\tilde{X}\tilde{X}^3) - 0 = E(\tilde{X}^4) = \sigma^4(3)!! = 3$$

Si noti che in termini generali, stante $\tilde{X} \sim N(0, 1)$ e $\tilde{Y} = \tilde{X}^n$, per il coefficiente di correlazione vale la seguente formula:

$$\rho(\tilde{X}; \tilde{X}^n) = \frac{n!!}{\sqrt{(2n-1)!!}}$$

Nel caso in esame, posto $n = 3$, si ottiene:

$$\rho(\tilde{X}; \tilde{Y}) = \frac{\text{cov}(\tilde{X}; \tilde{Y})}{\sigma(\tilde{X})\sigma(\tilde{Y})} = \frac{3}{\sqrt{15}} = 0,78$$

Le v.a. oggetto di studio, quindi, pur essendo tra loro comonotone, possiedono coefficiente di correlazione lineare minore di 1.

²Il doppio fattoriale di un numero n è definito come segue: se n è pari si ha $n!! = \prod_{k=1}^{\frac{n}{2}} 2k$, mentre se n è dispari vale $n!! = \prod_{k=1}^{\frac{n+1}{2}} (2k-1)$.

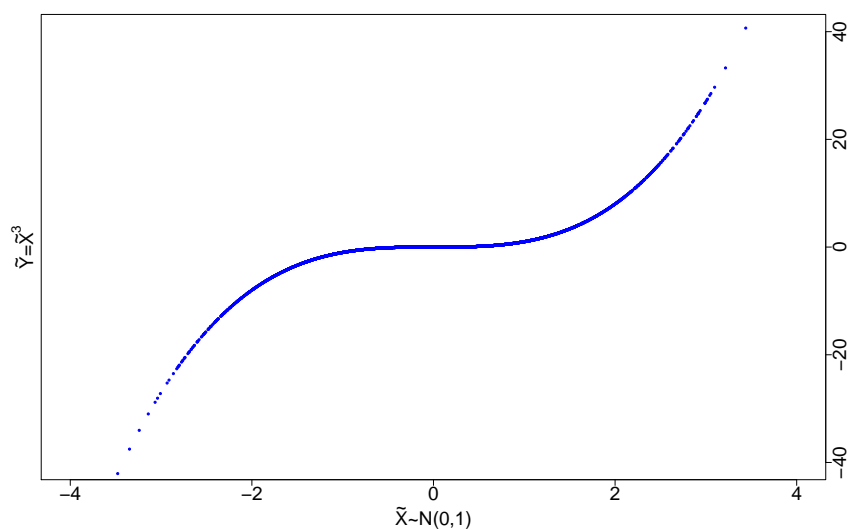


Figura 2.2: Scatterplot ottenuto tramite simulazione di 10.000 punti relativo alla relazione tra la v.a. $\tilde{X} \sim N(0, 1)$ e $\tilde{Y} = \tilde{X}^3$. Il valore di ρ è pari a 0,78.

Si osservi, inoltre, che al crescere della relazione non lineare esistente tra le v.a., pur mantenendo queste ultime comonotone, il valore di ρ assume un comportamento decrescente. In particolare, si consideri a titolo esemplificativo la funzione monotona crescente $\psi(x) = x^n$ con $n = 3, 5, 7, 9, 11, 13, 15$; come mostrato nel grafico seguente, al crescere della potenza considerata e quindi della non linearità della funzione, la correlazione lineare tra le due v.a. tende a decrescere.

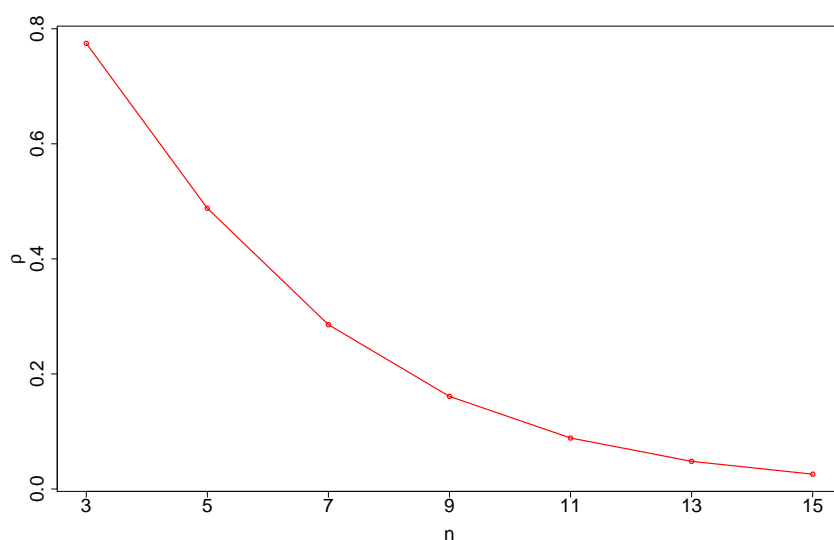


Figura 2.3: Coefficiente di correlazione lineare calcolato tra le v.a. $\tilde{X} \sim N(0, 1)$ e $\tilde{Y} = \tilde{X}^n$ al variare della potenza n .



Il coefficiente di correlazione lineare calcolato tra due v.a. indipendenti assume valore nullo. Un valore di $\rho = 0$, tuttavia, non implica l'indipendenza tra le variabili stesse; la relazione inversa è valida solo nel caso di variabili gaussiane.

Esempio. Si considerino la variabile uniforme continua nell'intervallo $[-1, 1]$ e la variabile definita dalla funzione quadratica $\psi(x) = x^2$, $\tilde{X} \sim U(-1, 1)$ e $\tilde{Y} = \tilde{X}^2$. Le due v.a. seppur dipendenti tra loro, in quanto \tilde{Y} funzione di \tilde{X} , possiedono coefficiente di correlazione nullo. La covarianza tra le due v.a., infatti, è pari a zero:

$$\text{cov}(\tilde{X}; \tilde{Y}) = E(\tilde{X}\tilde{X}^2) - E(\tilde{X})E(\tilde{X}^2) = 0$$

in quanto i momenti semplici di ordine r della variabile continua uniforme definita sul dominio $[-a; a]$ sono nulli per r dispari. Si noti che il risultato dell'esempio precedente è valido per ogni variabile aleatoria dotata di media e momento semplice di ordine 3 pari a zero. ■

In definitiva, le principali limitazioni nell'uso del coefficiente di correlazione lineare per descrivere la dipendenza tra due v.a. sono le seguenti:

- ρ è definito solo per v.a. con media e varianza finite;
- $\rho = 0$ implica l'indipendenza tra le v.a. solo se le variabili sono gaussiane;
- il valore di ρ dipende dalle caratteristiche distributive delle v.a. marginali;
- $\rho = \rho_{min}$ e $\rho = \rho_{max}$ se le v.a. sono perfettamente dipendenti tra loro in senso negativo e positivo, ma non vale il viceversa;
- il coefficiente di correlazione lineare non è invariante rispetto a trasformazioni non lineari strettamente crescenti;
- le v.a. marginali e il valore di ρ determinano univocamente la distribuzione di probabilità congiunta solo nel caso di distribuzioni ellittiche.

Per quantificare la dipendenza tra due v.a., superando le limitazioni del coefficiente di correlazione lineare, è possibile ricorrere a misure di dipendenza definite in letteratura come misure di concordanza. In particolare, le più utilizzate sono il τ di Kendall e il ρ di Spearman.

Il τ di Kendall è una misura di concordanza tra due variabili casuali; in termini intuitivi una coppia di v.a. è concordante se valori elevati di una variabile tendono ad essere associati a valori elevati dell'altra e, analogamente, a valori contenuti di una variabile corrispondono valori contenuti dell'altra. La situazione speculare, invece, determina il caso di discordanza. In termini formali, due punti (x_1, x_2) e (x'_1, x'_2) sono definiti concordanti se $(x_1 - x_2)(x'_1 - x'_2) > 0$, mentre si dicono

discordanti se $(x_1 - x_2)(x'_1 - x'_2) < 0$. Considerando le coppie di v.a. $(\tilde{X}_1, \tilde{X}_2)$ e $(\tilde{X}'_1, \tilde{X}'_2)$ tra loro indipendenti, con x_1 e x_2 copie³ delle variabili x'_1 e x'_2 , il τ di Kendall è definito come la differenza tra le probabilità di concordanza e discordanza delle coppie considerate:

$$\tau(\tilde{X}_1; \tilde{X}_2) = P((\tilde{X}_1 - \tilde{X}'_1)(\tilde{X}_2 - \tilde{X}'_2) > 0) - P((\tilde{X}_1 - \tilde{X}'_1)(\tilde{X}_2 - \tilde{X}'_2) < 0) \quad (2.1.2)$$

Il τ di Kendall, inoltre, può essere espresso con la seguente relazione:

$$\tau(\tilde{X}_1; \tilde{X}_2) = 4 \iint H(x_1, x_2)h(x_1, x_2) dx_1 dx_2 - 1$$

La formulazione campionaria, di contro, è definita come:

$$\hat{\tau} = \frac{c - d}{c + d} = \frac{c - d}{\binom{n}{2}}$$

dove c e d individuano rispettivamente il numero delle concordanze e discordanze tra le coppie di osservazioni ed n la numerosità campionaria. La statistica $\hat{\tau}$ fornisce una stima non distorta del valore del τ di Kendall della popolazione in esame.

Il ρ di Spearman è un'altra misura di concordanza tra due v.a. ed è definita come segue. Siano $(\tilde{X}_1, \tilde{Y}_1)$, $(\tilde{X}_2, \tilde{Y}_2)$ e $(\tilde{X}_3, \tilde{Y}_3)$ tre coppie indipendenti e identicamente distribuite della v.a. bidimensionale (\tilde{X}, \tilde{Y}) ; il ρ di Spearman tra le v.a. \tilde{X} e \tilde{Y} è definito come,

$$\rho_S(\tilde{X}; \tilde{Y}) = 3P((\tilde{X}_1 - \tilde{X}_2)(\tilde{Y}_1 - \tilde{Y}_3) > 0) - P((\tilde{X}_1 - \tilde{X}_2)(\tilde{Y}_1 - \tilde{Y}_3) < 0) \quad (2.1.3)$$

Dalla formula precedente risulta che la grandezza ρ_S individua, a meno di una costante moltiplicativa, la differenza tra la probabilità delle concordanze e quella delle discordanze tra le coppie $(\tilde{X}_1, \tilde{Y}_1)$ e $(\tilde{X}_2, \tilde{Y}_3)$. La relazione precedente può essere formulata come segue:

$$\rho_S(\tilde{X}; \tilde{Y}) = 12 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [H(x, y) - F(x)F(y)] dF(x) dF(y)$$

Il ρ di Spearman, inoltre, può essere espresso come:

$$\rho_S(\tilde{X}; \tilde{Y}) = \rho(F(x), F(y))$$

mostrando che ρ_S è equivalente al coefficiente di correlazione lineare tra le funzioni di ripartizione delle due variabili casuali in esame. La statistica campionaria di ρ_S calcolata considerando un campione di dimensione n è pari a:

³La v.a. \tilde{X}_1 è definita *copia* della v.a. \tilde{X}_2 se $F(x_1) = F(x_2)$.

$$\hat{\rho}_S(\tilde{X}; \tilde{Y}) = \frac{12}{n(n^2 - 1)} \sum_i \left(r_i - \frac{n+1}{2} \right) \left(s_i - \frac{n+1}{2} \right)$$

dove $r_i = \text{rango}(x_i)$ e $s_i = \text{rango}(y_i)$. La statistica $\hat{\rho}_S$ è uno stimatore distorto di ρ_S , infatti $E(\hat{\rho}_S) = \frac{(n-2)\rho_S + 3\tau}{n+1}$ che tende a ρ_S quando la numerosità campionaria tende a infinito (Balakrishnan e Lai (2009, p. 157)).

Per le misure di concordanza definite dal τ di Kendall e dal ρ di Spearman, valgono i seguenti risultati:

- l'indipendenza tra \tilde{X} e \tilde{Y} implica che $\tau = \rho_S = 0$, tuttavia non vale la relazione inversa;
- τ e $\rho_S \in [-1, 1]$, inoltre i valori di -1 e 1 sono raggiunti solo quando le v.a. sono tra loro contromonotone e comonotone rispettivamente e vale anche la relazione inversa;
- τ e ρ_S sono invarianti rispetto a trasformazioni strettamente monotone crescenti delle variabili aleatorie.
- la distribuzione congiunta di una variabile aleatoria $H(x, y)$ non è univocamente determinata dalle distribuzioni marginali $F(x)$ e $F(y)$ e dalla misura di concordanza prescelta.

Si noti che, a differenza del coefficiente lineare di Pearson, date due variabili aleatorie marginali continue, è sempre possibile costruire una v.a. congiunta con valore di concordanza arbitrario nell'intervallo $[-1, 1]$.

Sebbene il τ di Kendall e il ρ di Spearman siano entrambe delle misure di concordanza, i valori che assumono possono essere alquanto differenti. In particolare, la relazione che lega τ e ρ_S data una v.a. bidimensionale $H(x, y)$ dotata di marginali continue \tilde{X} e \tilde{Y} è la seguente (Nelsen (2006, p. 174)):

$$-1 \leq 3\tau - 2\rho_S \leq 1$$

Altre diseuguaglianze valide risultano essere,

$$\frac{1 + \rho_S}{2} \geq \left(\frac{1 + \tau}{2} \right)^2$$

e

$$\frac{1 - \rho_S}{2} \geq \left(\frac{1 - \tau}{2} \right)^2$$

Le precedenti tre relazioni possono essere combinate tra loro per giungere al seguente risultato che individua la regione di valori che possono assumere le grandezze τ e ρ_S , come illustrato nella figura seguente:

$$\frac{3\tau - 1}{2} \leq \rho_S \leq \frac{1 + 2\tau - \tau^2}{2} \quad \tau \geq 0$$

e

$$\frac{\tau^2 + 2\tau - 1}{2} \leq \rho_S \leq \frac{1 + 3\tau}{2} \quad \tau \leq 0$$

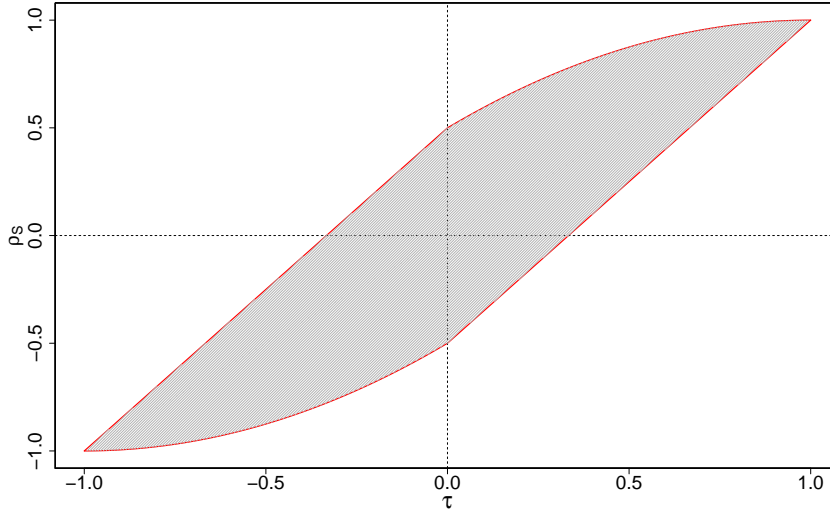


Figura 2.4: Evidenza dei valori che possono assumere le misure di concordanza τ e ρ_S per coppie di variabili aleatorie continue.

I coefficienti di dipendenza di coda misurano anch'essi il legame tra coppie di v.a. e quantificano, in particolare, la dipendenza esistente nelle code della distribuzione bivariata oggetto di studio. In particolare, il coefficiente di dipendenza di coda superiore guarda alla probabilità che la variabile \tilde{Y} ecceda il quantile di ordine α , $q_\alpha(\tilde{Y})$, condizionatamente a valori di \tilde{X} che eccedono il rispettivo quantile $q_\alpha(\tilde{X})$. Il coefficiente è così definito:

$$\lambda_U(\tilde{X}; \tilde{Y}) = \lim_{\alpha \rightarrow 1^-} P[\tilde{Y} > F_{\tilde{Y}}^{-1}(\alpha) | \tilde{X} > F_{\tilde{X}}^{-1}(\alpha)] \quad (2.1.4)$$

e, se esiste il limite di cui sopra, vale che $\lambda_U \in [0, 1]$. Se $\lambda_U = 0$, le v.a. \tilde{X} e \tilde{Y} sono asintoticamente indipendenti nella coda superiore, viceversa le v.a. mostreranno dipendenza di coda superiore. Analogamente, il coefficiente di coda inferiore è definito come segue:

$$\lambda_L(\tilde{X}; \tilde{Y}) = \lim_{\alpha \rightarrow 0^+} P[\tilde{Y} \leq F_{\tilde{Y}}^{-1}(\alpha) | \tilde{X} \leq F_{\tilde{X}}^{-1}(\alpha)] \quad (2.1.5)$$

dove, se il limite esiste, $\lambda_L \in [0, 1]$. Se $\lambda_L = 0$, le v.a. non mostrano asintoticamente alcuna dipendenza di coda inferiore, in caso contrario saranno tra loro dipendenti sulla coda inferiore.

2.2 Le copule

Nello studio della dipendenza tra v.a. le copule rappresentano uno strumento metodologico fondamentale in quanto, per variabili aleatorie continue multivariate, permettono di separare le distribuzioni marginali dalla rispettiva struttura di dipendenza, la quale può essere rappresentata univocamente tramite una copula.

Prima di definire la funzione copula e al fine di rendere la trattazione più agevole, è opportuno richiamare i seguenti concetti di base. Considerando la v.a. \tilde{X} definita sull'asse reale, la funzione di ripartizione, $F(x) = P[\tilde{X} \leq x]$, è una funzione F di dominio \mathbb{R} e codominio $[0, 1]$ tale per cui:

- F è non decrescente;
- $F(-\infty) = 0$ e $F(\infty) = 1$

In merito ad una variabile casuale bidimensionale (\tilde{X}, \tilde{Y}) con marginali definite sull'asse reale, la funzione di ripartizione, $H(x, y) = P[\tilde{X} \leq x, \tilde{Y} \leq y]$, è una funzione H di dominio \mathbb{R}^2 e codominio $[0, 1]$ tale per cui:

- H è non decrescente rispetto ad ognuno dei suoi argomenti;
- $H(x, -\infty) = H(-\infty, y) = 0$ e $H(\infty, \infty) = 1$

Le funzioni di ripartizione marginali sono date rispettivamente da $F(x) = H(x, \infty)$ e $G(y) = H(\infty, y)$.

Si consideri il vettore aleatorio $(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_d)$ con marginali continue e dotate di funzione di ripartizione $F_i(x)$; il vettore aleatorio

$$(U_1, U_2, \dots, U_d) = (F_1(x_1), F_2(x_2), \dots, F_d(x_d)) \quad (2.2.1)$$

possiede marginali uniformemente distribuite in $[0, 1]$. La copula associata a $(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_d)$ è definita come la funzione di ripartizione di (U_1, U_2, \dots, U_d) . Una copula, quindi, è una funzione di ripartizione multivariata con distribuzioni marginali uniformi $U(0, 1)$. Considerando il generico caso n -dimensionale, la copula è quindi definita sul dominio individuato dall'ipercubo di lato unitario \mathbb{I}^n , dove $\mathbb{I} = [0, 1]$, mentre il codominio risulta essere \mathbb{I} . Considerando il caso bidimensionale, senza alcuna perdita di generalità, una copula può essere definita formalmente come segue.

Definizione. Una copula è una funzione $C : \mathbb{I}^2 \rightarrow \mathbb{I}$ che gode delle seguenti proprietà

$$C(u, 0) = 0 = C(0, v) \quad u, v \in [0, 1] \quad (2.2.2)$$

$$C(u, 1) = u \quad e \quad C(1, v) = v \quad (2.2.3)$$

Per ogni u_1, u_2, v_1 e v_2 tali che $u_1 \leq u_2$ e $v_1 \leq v_2$,

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0 \quad (2.2.4)$$

□

La relazione (2.2.4) individua il volume della funzione copula C sul rettangolo di vertici $[u_1, u_2] \times [v_1, v_2]$. La copula, quindi, assegna un numero compreso in $[0, 1]$ ad ogni rettangolo del dominio; la relazione precedente, inoltre, garantisce che tale numero non possa essere negativo.

Il teorema di Sklar fornisce evidenza di come separare la struttura di dipendenza contenuta in una distribuzione di probabilità congiunta dalle rispettive variabili marginali.

Teorema di Sklar. Sia H una funzione di ripartizione bidimensionale e siano F e G le funzioni di ripartizioni delle variabili marginali. Allora, esiste una funzione copula C tale per cui per ogni $(x, y) \in \mathbb{R}^2$

$$H(x, y) = C(F(x), G(y)) \quad (2.2.5)$$

Se F e G sono continue, la funzione copula C è unica; in caso contrario, C è univocamente definita su $Immagine(F) \times Immagine(G)$. Di contro, se C è una copula e F e G sono le funzioni di ripartizioni marginali, allora la funzione H definita nella (2.2.5) è una funzione di ripartizione bivariata di marginali F e G .

□

Il teorema di Sklar, quindi, fornisce una rappresentazione della copula in termini di funzione di ripartizione congiunta. È possibile, inoltre, invertire la relazione precedente ed esprimere la copula in termini di distribuzione congiunta e di funzione inversa delle funzioni di ripartizione marginali. Si noti che se le variabili marginali non sono strettamente crescenti, queste non possiedono funzione inversa; in tale circostanza, quindi, si deve ricorrere alla funzione quasi-inversa definita come,

$$F^{(-1)}(t) = \inf(\{x | F(x) \geq t\}) = \sup\{x | F(x) \leq t\}$$

Se F è strettamente crescente la funzione quasi-inversa coincide con la funzione inversa, $F^{(-1)} = F^{-1}$. La funzione copula C , quindi, può essere ottenuta come,

$$C(u, v) = H(F^{(-1)}(u), G^{(-1)}(v)) \quad u, v \in [0, 1] \quad (2.2.6)$$

La relazione precedente mostra, analogamente al teorema di Sklar, l'evidenza tale per cui sottostante ad ogni distribuzione congiunta esista una particolare struttura di dipendenza la quale, inoltre, può essere estratta tramite la funzione copula dalla distribuzione bivariata in esame.

Esempio. Si consideri la seguente funzione di ripartizione congiunta, definita in letteratura come *distribuzione logistica bivariata di Gumbel* (Kotz, Balakrishnan e Johnson (2000, p. 551)),

$$H(x, y) = (1 + e^{-x} + e^{-y})^{-1}$$

con $x, y \in \mathbb{R}$. Tramite metodo dell'inversione è possibile esplicitare la funzione copula che lega le variabili marginali \tilde{X} e \tilde{Y} . Le funzioni di ripartizione marginali sono ottenibili come segue,

$$\begin{aligned} H(x, \infty) &= F(x) = (1 + e^{-x})^{-1} \\ H(\infty, y) &= G(y) = (1 + e^{-y})^{-1} \end{aligned}$$

mostrando che $F(x)$ e $G(y)$ individuano la funzione di ripartizione della variabile logistica standard. Dopo avere calcolato le distribuzioni marginali, è possibile esplicitare la copula implicitamente contenuta nella distribuzione congiunta H . Per il teorema di Sklar, infatti, vale la relazione $H(x, y) = C(F(x), G(y))$, da cui segue che $C(u, v) = H(F^{-1}(u), G^{-1}(v))$. Le funzioni F e G possiedono la seguente funzione inversa:

$$\begin{aligned} F^{-1}(u) &= -\ln\left(\frac{1-u}{u}\right) \\ G^{-1}(v) &= -\ln\left(\frac{1-v}{v}\right) \end{aligned}$$

da cui è possibile trovare la funzione copula:

$$\begin{aligned} C(u, v) &= H(F^{-1}(u), G^{-1}(v)) \\ &= \left[1 + \exp\left(-\left(-\ln\left(\frac{1-u}{u}\right)\right)\right) + \exp\left(-\left(-\ln\left(\frac{1-v}{v}\right)\right)\right) \right]^{-1} \\ &= \left[1 + \frac{1-u}{u} + \frac{1-v}{v} \right]^{-1} \\ &= \frac{uv}{v+u-uv} \end{aligned}$$

La copula così ottenuta appartiene alla famiglia di *Ali-Mikhail-Haq*; in particolare, in tale circostanza il parametro caratteristico della famiglia considerata è pari all'unità, $\theta = 1$. ■

L'esempio precedente, quindi, mostra come sia possibile estrarre la funzione copula dalla distribuzione congiunta; di contro, stante una determinata copula, è sempre possibile costruire una distribuzione multivariata con marginali arbitrarie.

Esempio. Riprendendo l'esempio di cui sopra, è possibile mostrare come costruire una distribuzione bivariata con marginali arbitrarie e funzione copula prefissata. Si consideri la copula di Ali-Mikhail-Haq con parametro unitario, $C(u, v) = \frac{uv}{v+u-uv}$ e siano $F(x) = (1 + e^{-x})^{-1}$ e $G(y) = (1 + e^{-y})^{-1}$ le funzioni di ripartizione delle due variabili marginali. Ricorrendo al teorema di Sklar è possibile costruire la funzione di ripartizione congiunta:

$$\begin{aligned} H(x, y) &= C(F(x), G(y)) \\ &= \frac{(1 + e^{-x})^{-1}(1 + e^{-y})^{-1}}{(1 + e^{-x})^{-1} + (1 + e^{-y})^{-1} - (1 + e^{-x})^{-1}(1 + e^{-y})^{-1}} \\ &= (1 + e^{-x} + e^{-y})^{-1} \end{aligned}$$

che risulta essere la distribuzione logistica bivariata di Gumbel dell'esempio precedente. ■

La relazione (2.2.6) mostra come una copula sia una funzione di ripartizione bivariata con variabili marginali uniformi. Una funzione copula, quindi, può essere pensata come la rappresentazione uniforme di una variabile aleatoria bivariata. In termini di funzioni di densità, una variabile bidimensionale può essere scritta esplicitandone la copula come segue,

$$h(x, y) = c(F(x), G(y))f(x)g(y) \quad (2.2.7)$$

dove h , c , f e g individuano le funzioni di densità della variabile (\tilde{X}, \tilde{Y}) , della copula C e delle rispettive distribuzioni marginali.

In particolare, la densità della funzione copula è definita come,

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v} \quad (2.2.8)$$

La funzione di ripartizione della copula, inoltre, può essere convenientemente scritta come:

$$C(u, v) = A_C(u, v) + S_C(u, v)$$

dove $A_C(u, v) = \int_0^u \int_0^v \frac{\partial^2 C}{\partial s \partial t}(s, t) dt ds$ e $S_C(u, v) = C(u, v) - A_C(u, v)$. Se $C \equiv A_C$, allora C è assolutamente continua e ha densità data da $\frac{\partial^2 C}{\partial s \partial t}(s, t)$, mentre se $C \equiv S_C$, C è detta singolare. In tutti gli altri casi intermedi, C avrà sia una componente assolutamente continua che una componente singolare. Si noti che se C è definita su \mathbb{I}^2 allora possiede supporto completo; di contro, quando la copula è singolare il supporto ha misura zero.

Infine, ricorrendo alle copule, per le funzioni di ripartizione e densità condizionate di due generiche variabili casuali valgono le seguenti relazioni:

$$F(x|y) = \frac{\partial H(x, y)}{\partial y} = \frac{\partial C(F(x), G(y))}{\partial G(y)} \quad (2.2.9)$$

$$f(x|y) = \frac{f(x, y)}{g(y)} = \frac{c(F(x), G(y))f(x)g(y)}{g(y)} = c(F(x), G(y))f(x) \quad (2.2.10)$$

Esempio. Si consideri la copula Clayton dotata di seguente funzione di ripartizione⁴,

$$C(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}} \quad \text{con} \quad \theta \in (0, \infty)$$

E' immediato osservare che le variabili marginali u e v sono distribuite uniformemente su $[0, 1]$; la funzione di ripartizione delle marginali, infatti, può essere

⁴La copula Clayton è definita come $C(u, v) = \max[(u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}, 0]$ con $\theta \in [-1, 0) \cup (0, \infty)$; ai fini dell'esempio, tuttavia, si considera $\theta \in (0, \infty)$, da cui la funzione copula può essere semplificata come $C(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}$.

ottenuta come $F(u) = C(u, \infty) = (u^{-\theta})^{-\theta^{-1}} = u$ che coincide con la funzione di ripartizione della variabile uniforme continua nell'intervallo $[0, 1]$. Relazione analoga sussiste per la marginale v .

La funzione di ripartizione della variabile u condizionata v risulta essere:

$$F(u|v) = \frac{\partial C(u, v)}{\partial v} = (u^{-\theta} + v^{-\theta} - 1)^{-1 - \frac{1}{\theta}} v^{-1 - \theta}$$

La funzione di densità, di contro, è pari a,

$$f(u|v) = c(F(u), F(v)) f(u) = \frac{\partial^2 C(u, v)}{\partial u \partial v} = (1 + \theta)(uv)^{-1 - \theta} (u^{-\theta} + v^{-\theta} - 1)^{-2 - \frac{1}{\theta}}$$

Si noti che nel caso di marginali uniformi, l'espressione della densità non condizionata e condizionata coincidono in quanto nella relazione (2.2.10) si ha $f(x) = 1$. Le funzioni di densità e di ripartizione condizionate, tuttavia, sono funzioni unidimensionali e quindi in tale contesto la grandezza condizionante, v , assume significato di parametro e non di variabile. Fissando, quindi, il valore della variabile condizionata v è possibile ottenere le specifiche funzioni condizionate di ripartizione e di densità. I grafici seguenti riportano le funzioni di ripartizione e densità condizionate rispetto al valore $v = 0, 2$ per la copula Clayton di parametro $\theta = 2$.

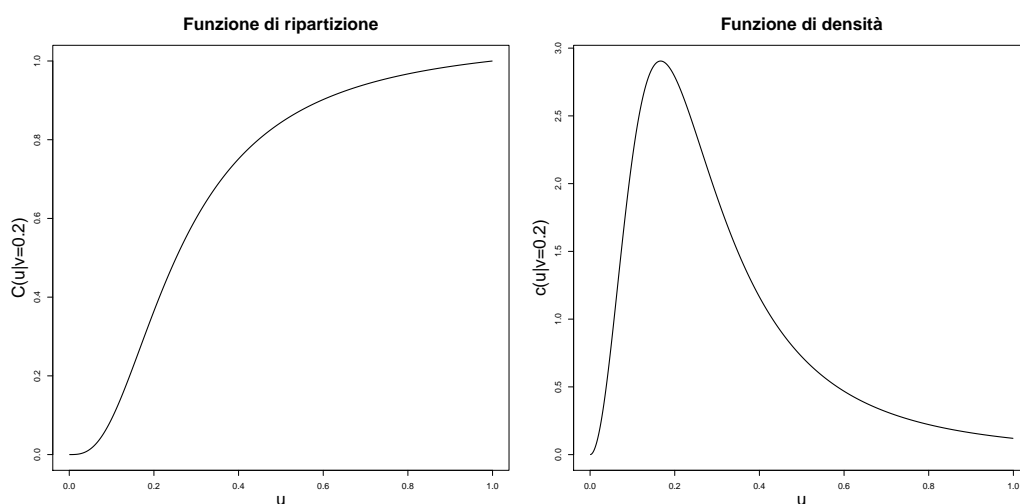


Figura 2.5: Funzioni di ripartizione e di densità condizionate rispetto alla variabile $v = 0, 2$ per la copula Clayton di parametro $\theta = 2$.

I grafici seguenti, invece, evidenziano sulla densità bivariata della copula il condizionamento al valore di $v = 0, 2$.

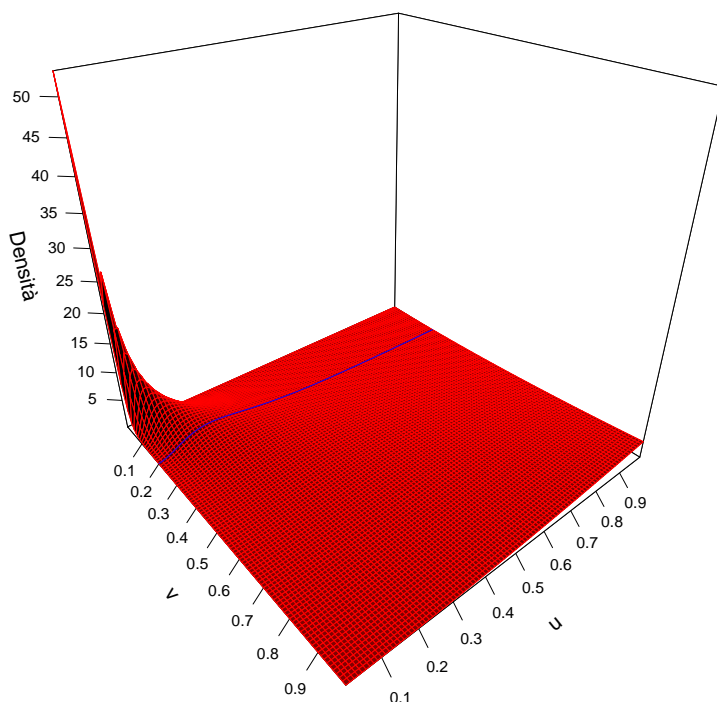


Figura 2.6: Densità della copula Clayton di parametro $\theta = 2$ con evidenza della densità condizionata rispetto al valore $v = 0, 2$.

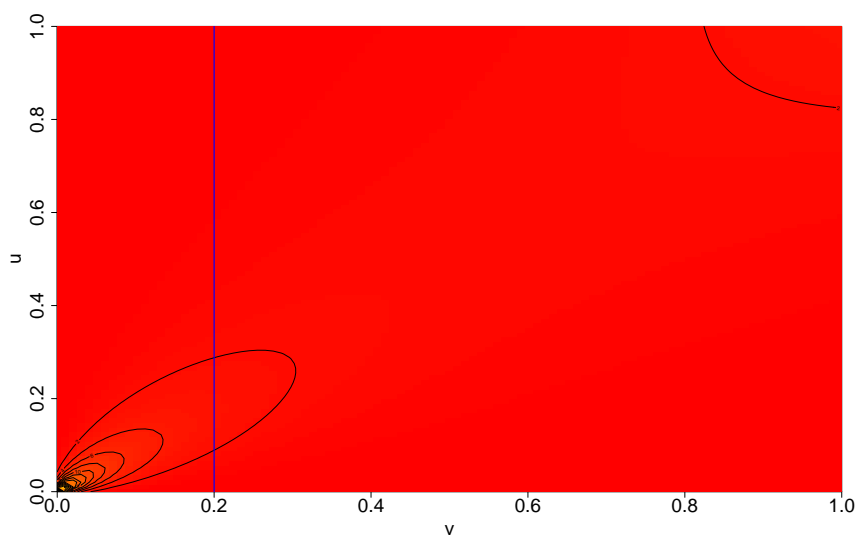


Figura 2.7: Curve di livello della funzione di densità della copula Clayton di parametro $\theta = 2$ con evidenza del valore della variabile condizionante pari a $v = 0, 2$.

■

Le funzioni copula godono della proprietà di invarianza rispetto a trasformazioni strettamente crescenti delle variabili. Siano \tilde{X} e \tilde{Y} due v.a. continue dotate di copula $C_{\tilde{X},\tilde{Y}}$. Se α e β sono funzioni strettamente crescenti allora, $C_{\alpha(\tilde{X}),\beta(\tilde{Y})} = C_{\tilde{X},\tilde{Y}}$. Inoltre, considerando le funzioni α e β strettamente monotone, valgono le seguenti relazioni.

- Se α è strettamente decrescente e β è strettamente crescente:

$$C_{\alpha(\tilde{X}),\beta(\tilde{Y})}(u, v) = v - C_{\tilde{X},\tilde{Y}}(1 - u, v)$$

che individua la copula C ruotata di 90° , $C_{90^\circ}(u, v) = v - C(1 - u, v)$.

- Se α e β sono strettamente decrescenti:

$$C_{\alpha(\tilde{X}),\beta(\tilde{Y})}(u, v) = u + v - 1 + C_{\tilde{X},\tilde{Y}}(1 - u, 1 - v)$$

che definisce la copula C ruotata di 180° , $C_{180^\circ}(u, v) = u + v - 1 + C(1 - u, 1 - v)$, detta anche copula di sopravvivenza.

- Se α è strettamente crescente e β è strettamente decrescente:

$$C_{\alpha(\tilde{X}),\beta(\tilde{Y})}(u, v) = u - C_{\tilde{X},\tilde{Y}}(u, 1 - v)$$

la quale identifica la copula C ruotata di 270° , $C_{270^\circ}(u, v) = u - C(u, 1 - v)$. Quest'ultima, inoltre, può essere ottenuta come rotazione di 180° della copula C_{90° ; ruotare C_{90° di 180° , infatti, porta alla seguente copula $C(u, v) = u + v - 1 + [1 - v - C(u, 1 - v)] = u - C(u, 1 - v) = C_{270^\circ}$.

Si noti che, nelle relazioni precedenti, la copula C è indipendente dalla forma funzionale di α e β .

Una copula gode della proprietà di *simmetria radiale* se la copula di sopravvivenza coincide con la rispettiva copula non ruotata, $C = C_{180^\circ}$. Altra caratteristica di simmetria delle copule è data dalla proprietà di *interscambiabilità*; una copula gode di tale proprietà quando $C(u, v) = C(v, u)$ per ogni $(u, v) \in \mathbb{I}^2$. In tale circostanza la copula è anche definita semplicemente *simmetrica* (Nelsen (2006, p. 37)).

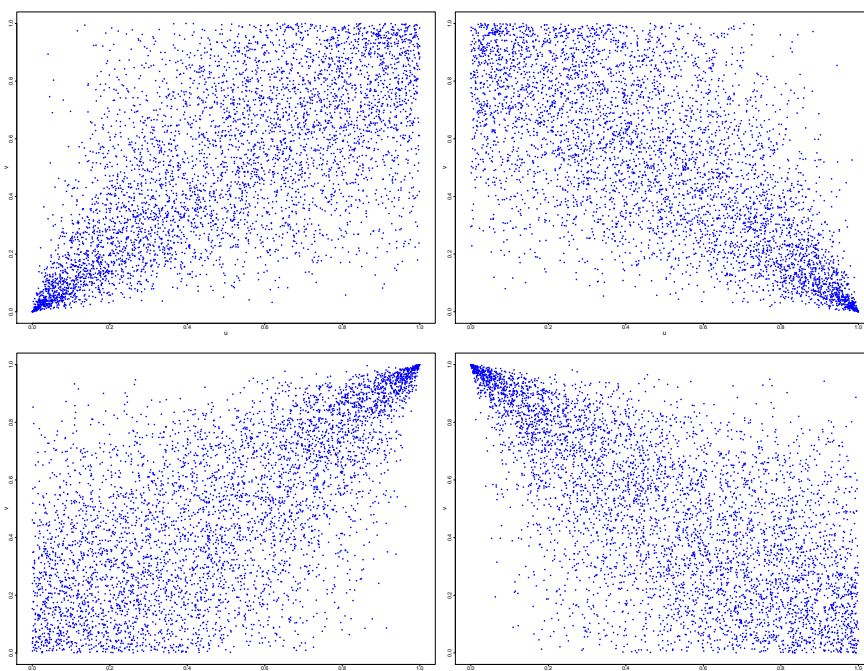


Figura 2.8: 5.000 realizzazioni casuali della copula Clayton di parametro $\theta = 2$ (grafico in alto a sinistra), della copula ruotata di 90° (grafico in alto a destra), della copula ruotata di 180° (grafico in basso a sinistra) e della copula ruotata di 270° (grafico in basso a destra).

Le funzioni copula sono superiormente e inferiormente limitate dai cosiddetti *limiti di Fréchet* che individuano le situazioni estreme di dipendenza negativa e positiva. In particolare, vale il seguente teorema.

Teorema: limiti di Fréchet. Per ogni copula C si hanno i seguenti limiti

$$\max(u + v - 1, 0) \leq C(u, v) \leq \min(u, v) \tag{2.2.11}$$

□

Dove

$$W(u, v) = \max(u + v - 1, 0) \tag{2.2.12}$$

individua il limite inferiore di Fréchet dato dalla copula contromonota, mentre

$$M(u, v) = \min(u, v) \tag{2.2.13}$$

rappresenta il limite superiore dato dalla copula comonotona. Le funzioni copula, quindi, permettono di costruire strutture di dipendenza che variano tra i casi di perfetta dipendenza negativa e positiva. Il caso di indipendenza, invece, è dato dalla copula prodotto:

$$\Pi(u, v) = uv \tag{2.2.14}$$

Una famiglia di copule è qualificata come *comprendiva* quando permette di ottenere le copule W , Π e M .

Le misure di concordanza del τ di Kendall (2.1.2) e del ρ di Spearman (2.1.3) dipendono, nel caso di variabili marginali aleatorie continue, unicamente dalla copula sottostante alle variabili marginali. In particolare, tali misure di concordanza possono essere formulate come segue (Balakrishnan e Lai (2009, p. 155)),

$$\tau = 4 \int_0^1 \int_0^1 C(u, v) c(u, v) du dv - 1 = 4E(C(u, v)) - 1 \quad (2.2.15)$$

$$\rho_S = 12 \int_0^1 \int_0^1 C(u, v) du dv - 3 = 12E(\tilde{U}\tilde{V}) - 3 \quad (2.2.16)$$

Valgono, inoltre, le seguenti relazioni:

- $\tau(\tilde{X}, \tilde{Y}) = \rho_S(\tilde{X}, \tilde{Y}) = -1 \iff C(x, y) = W$;
- $\tau(\tilde{X}, \tilde{Y}) = \rho_S(\tilde{X}, \tilde{Y}) = 0 \implies C(x, y) = \Pi$;
- $\tau(\tilde{X}, \tilde{Y}) = \rho_S(\tilde{X}, \tilde{Y}) = 1 \iff C(x, y) = M$.

In aggiunta, il τ di Kendall calcolato tra le v.a. \tilde{X} e \tilde{Y} dotate di copula C coincide con quello della copula ruotata di 180° , $\tau(C(x, y)) = \tau(C_{180^\circ}(x, y))$ (Embrechts, Lindskog e McNeil (2003)). Analogamente, vale lo stesso risultato per le copule ruotate di 90° e 270° : $\tau(C_{90^\circ}(x, y)) = \tau(C_{270^\circ}(x, y))$, dove $\tau(C(x, y)) = -\tau(C_{90^\circ}(x, y))$.

Anche i coefficienti di dipendenza di coda possono essere riformulati ricorrendo alla funzione copula. In particolare, le relazioni (2.1.4) e (2.1.5) possono essere espresse come segue:

$$\lambda_U(\tilde{X}; \tilde{Y}) = 2 - \lim_{v \rightarrow 1^-} \frac{1 - C(v, v)}{1 - v} \quad (2.2.17)$$

$$\lambda_L(\tilde{X}; \tilde{Y}) = \lim_{v \rightarrow 0^+} \frac{C(v, v)}{v} \quad (2.2.18)$$

Se le v.a. \tilde{X} e \tilde{Y} sono comonotone, vale che $\lambda_U = \lambda_L = 1$; di contro, se le v.a. sono contromonotone, $\lambda_U = \lambda_L = 0$. Si noti che i coefficienti di dipendenza di coda superiore e inferiore risultano nulli anche nel caso di v.a. indipendenti. In merito alla copula di sopravvivenza, inoltre, il coefficiente di dipendenza di coda superiore coincide con il coefficiente di dipendenza di coda inferiore della copula non ruotata e viceversa per il coefficiente di coda inferiore:

$$\begin{aligned} \lambda_L^{C_{180^\circ}} &= \lambda_U^C \\ \lambda_U^{C_{180^\circ}} &= \lambda_L^C \end{aligned}$$

Per le copule che godono della proprietà di simmetria radiale, infine, il coefficiente di dipendenza di coda superiore e inferiore coincidono (McNeil, Frey e Embrechts (2005, p. 209)).

Da un punto di vista pratico risulta di primaria importanza potere generare delle realizzazioni casuali da una prescelta copula. Un algoritmo di carattere generale che risponde a tale esigenza è quello denominato come *metodo della distribuzione condizionata* (Nelsen (2006, p. 40)). In particolare, in base a tale algoritmo, per simulare dalla copula $C(u, v)$ è necessario eseguire i seguenti passaggi:

- 1) generare due variabili aleatorie u e t indipendenti e distribuite uniformemente in $[0, 1]$;
- 2) porre $v = F^{(-1)}(u|t)$, dove $F^{(-1)}$ rappresenta la funzione quasi-inversa di F con $F(u|t) = \frac{\partial C(u,t)}{\partial t}$;
- 3) le realizzazioni distribuite secondo la copula considerata C saranno (u, v) .

2.3 Distribuzioni ellittiche e relative copule

Le distribuzioni multivariate sferiche ed ellittiche⁵ sono rispettivamente delle estensioni delle distribuzioni multidimensionali Normale Standard, $N(\mathbf{0}, \mathbf{I})$, e Normale, $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Una variabile aleatoria $\tilde{\mathbf{x}}$ in n dimensioni è definita *sferica* se per ogni $\Gamma \in \mathcal{O}(\mathbf{n})$,

$$\Gamma \tilde{\mathbf{x}} \stackrel{d}{=} \tilde{\mathbf{x}}$$

dove $\stackrel{d}{=}$ individua che due v.a. hanno la stessa distribuzione di probabilità, mentre $\mathcal{O}(\mathbf{n})$ rappresenta l'insieme delle matrici $n \times n$ ortogonali. Equivalentemente, la v.a. $\tilde{\mathbf{x}}$ può essere descritta ricorrendo alla seguente rappresentazione stocastica:

$$\tilde{\mathbf{x}} \stackrel{d}{=} \tilde{r} \tilde{\mathbf{u}} \tag{2.3.1}$$

dove $\tilde{r} \geq 0$ è una v.a. unidimensionale indipendente da $\tilde{\mathbf{u}}$, mentre quest'ultima rappresenta la v.a. n -dimensionale distribuita uniformemente sulla superficie dell'ipersfera di raggio unitario in \mathbb{R}^n . Le grandezze \tilde{r} e $\tilde{\mathbf{u}}$ sono definite rispettivamente *variabile generatrice* e *base uniforme* della distribuzione sferica. Si noti, inoltre, che data la variabile $\tilde{\mathbf{x}}$, vale che $\tilde{\mathbf{x}} / \|\tilde{\mathbf{x}}\| \stackrel{d}{=} \tilde{\mathbf{u}}$.

Una variabile aleatoria $\tilde{\mathbf{x}}$ possiede distribuzione *ellittica* di parametri $\boldsymbol{\mu}$, di dimensione $(n \times 1)$, e $\boldsymbol{\Sigma}$, matrice simmetrica e semidefinita positiva di dimensione $(n \times n)$, se

⁵Per una trattazione approfondita del tema si rimanda a Fang, Kotz e Ng (1989).

$$\tilde{\mathbf{x}} \stackrel{d}{=} \boldsymbol{\mu} + \mathbf{A}'\tilde{\mathbf{y}}$$

dove $\tilde{\mathbf{y}}$ è una distribuzione sferica, mentre \mathbf{A} è una matrice $n \times n$ tale per cui $\mathbf{A}'\mathbf{A} = \boldsymbol{\Sigma}$ con $\text{rango}(\boldsymbol{\Sigma}) = n$. I parametri $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ rappresentano rispettivamente i parametri di posizione e di dispersione; in letteratura, la matrice $\boldsymbol{\Sigma}$ è anche definita matrice caratteristica.

Data la matrice $\boldsymbol{\Sigma}$ si ha che,

$$\rho_{i,j} = \Sigma_{i,j} / \sqrt{\Sigma_{i,i}\Sigma_{j,j}} \quad (2.3.2)$$

Se le variabili aleatorie \tilde{X}_i e \tilde{X}_j possiedono media e varianza finite, vale che $\rho_{i,j} = \text{cov}(\tilde{X}_i; \tilde{X}_j) / \sqrt{\text{var}(\tilde{X}_i)\text{var}(\tilde{X}_j)}$ il quale coincide con il coefficiente di correlazione lineare di Pearson definito dalla (2.1.1). Il concetto di correlazione lineare, quindi, nel contesto delle variabili ellittiche risulta essere una naturale misura di dipendenza tra le variabili marginali coinvolte in quanto svolge un ruolo di parametro caratteristico della distribuzione multivariata stessa nonché della rispettiva copula sottostante. La matrice di elementi $\rho_{i,j}$ è individuata con \mathbf{P} e prende il nome di matrice di correlazione.

La v.a. $\tilde{\mathbf{x}}$ possiede la seguente rappresentazione stocastica:

$$\tilde{\mathbf{x}} \stackrel{d}{=} \boldsymbol{\mu} + \tilde{r}\mathbf{A}'\tilde{\mathbf{u}} \quad (2.3.3)$$

Si noti che la conoscenza della variabile unidimensionale \tilde{r} nelle relazioni (2.3.1) e (2.3.3) permette di generare delle realizzazioni casuali dalla variabile sferica o ellittica prescelta. La variabile ellittica $\tilde{\mathbf{x}}$ possiede funzione di densità della seguente forma,

$$C_n |\boldsymbol{\Sigma}|^{-\frac{1}{2}} g((\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})) \quad (2.3.4)$$

dove $g(\cdot)$ è una funzione univocamente definita dalla variabile \tilde{r} e C_n è la costante di normalizzazione definita come,

$$C_n = \frac{\Gamma(n/2)}{2\pi^{n/2} \int_0^\infty r^{n-1} g(r^2) dr} \quad (2.3.5)$$

La funzione di densità espressa dalla (2.3.4) possiede la caratteristica di avere curve livello ellittiche, da cui deriva la qualifica di distribuzione *ellittica*, definita in letteratura come *Elliptically Contoured, EC*; la variabile $\tilde{\mathbf{x}}$ viene sinteticamente individuata come $\tilde{\mathbf{x}} \sim EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$. Si noti che la variabile $\tilde{\mathbf{x}}$ è sferica se $\boldsymbol{\mu} = \mathbf{0}$ e $\boldsymbol{\Sigma} = \mathbf{I}$.

Le distribuzioni ellittiche sono distribuzioni multivariate che definiscono implicitamente la famiglia di copule che ne prende il nome. In Fang, Fang e Kotz (2002) si fornisce un'espressione generale per le copule ellittiche. In particolare, considerando, senza perdita di generalità, il caso di una variabile aleatoria bivariata ($n = 2$), in cui $\boldsymbol{\mu} = (0, 0)$ e la matrice $\boldsymbol{\Sigma}$ sia definita come⁶,

⁶Si sta quindi assumendo $\Sigma_{1,1} = \Sigma_{2,2} = 1$. In tale circostanza, la matrice di dispersione $\boldsymbol{\Sigma}$ coincide con la matrice di correlazione \mathbf{P} .

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

si ha che la funzione di densità della variabile $\tilde{\mathbf{x}} \sim EC(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ diviene:

$$f(x_1, x_2; \rho; g) = \frac{1}{\sqrt{1-\rho^2}} g\left(\frac{x_1^2 + x_2^2 - 2\rho x_1 x_2}{\sqrt{1-\rho^2}}\right)$$

La funzione di ripartizione della copula sottostante alla variabile aleatoria $\tilde{\mathbf{x}}$ è data dalla seguente espressione:

$$C(u, v) = \frac{1}{\sqrt{1-\rho^2}} \int_{-\infty}^{Q_g^{-1}(u)} \int_{-\infty}^{Q_g^{-1}(v)} g\left(\frac{x_1^2 + x_2^2 - 2\rho x_1 x_2}{\sqrt{1-\rho^2}}\right) dx_1 dx_2 \quad (2.3.6)$$

dove $Q_g(\cdot)$ individua la funzione di ripartizione delle variabili marginali. La densità della (2.3.6) risulta,

$$c(u, v) = \phi(Q_g^{-1}(u), Q_g^{-1}(v); \rho; g) \quad (2.3.7)$$

dove $\phi(x, y; \rho; g) = f(x, y; \rho; g)/[q_g(x)q_g(y)]$, in cui $q_g(\cdot)$ è la funzione di densità marginale.

Per tutte le copule ellittiche il τ di Kendall è definito come,

$$\tau(\tilde{X}_1, \tilde{X}_2) = \frac{2}{\pi} \arcsin(\rho) \quad (2.3.8)$$

dove ρ è dato dalla (2.3.2).

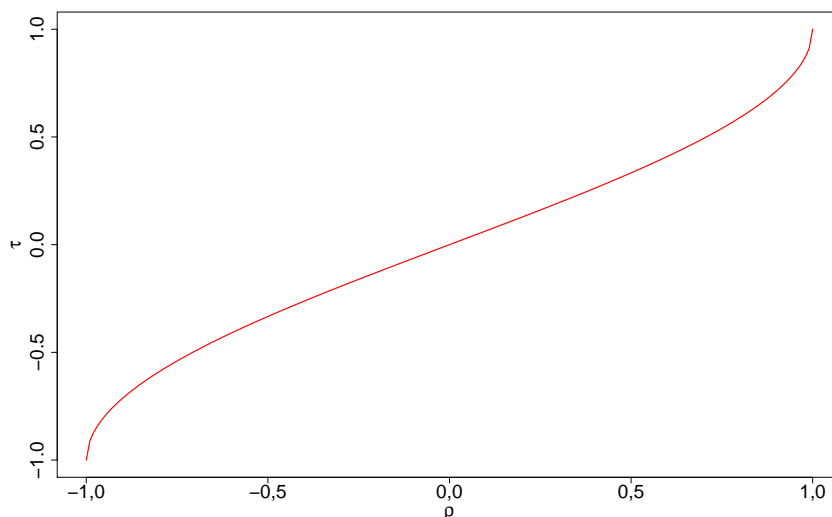


Figura 2.9: τ di Kendall per le copule ellittiche al variare del coefficiente di correlazione lineare tra le v.a. \tilde{X}_1 e \tilde{X}_2 .

I coefficienti di dipendenza di coda superiore e inferiore coincidono tra loro e sono definiti come (Schmidt (2002)):

$$\lambda_U(\tilde{X}_1, \tilde{X}_2) = \lambda_L(\tilde{X}_1, \tilde{X}_2) = \frac{\int_0^{h(\rho)} \frac{u^\alpha}{\sqrt{1-u^2}} du}{\int_0^1 \frac{u^\alpha}{\sqrt{1-u^2}} du}$$

dove ρ è definito dalla (2.3.2), $h(\rho) = \left(1 + \frac{(1-\rho)^2}{1-\rho^2}\right)^{-\frac{1}{2}}$, mentre $\alpha < 0$ è l'indice di variazione regolare⁷ (*regularly varying index*) della v.a. \tilde{r} che rappresenta la variabile generatrice della distribuzione ellittica in esame. Analogamente α può essere anche determinato in funzione dell'indice di variazione regolare della funzione generatrice della densità $g(\cdot)$, che deve possedere indice di variazione pari a $-\alpha/2 - 1 < 0$. Per valutare la dipendenza di coda di una variabile ellittica, quindi, è sufficiente quantificare se $g(\cdot)$ o $f_{\tilde{r}}(r)$ sono funzioni a variazione regolare. Analogamente, in Demarta e McNeil (2005) è fornita un'ulteriore formula per definire la dipendenza di coda, nello specifico si ha che:

$$\lambda_U(\tilde{X}_1, \tilde{X}_2) = \lambda_L(\tilde{X}_1, \tilde{X}_2) = \frac{\int_{(\pi/2 - \arcsin \rho)/2}^{\pi/2} \cos^\alpha t dt}{\int_0^{\pi/2} \cos^\alpha t dt}$$

2.3.1 Distribuzione e copula Normale

La variabile aleatoria Normale bidimensionale, $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, è una variabile continua definita nel piano \mathbb{R}^2 ed è caratterizzata da due parametri, il vettore delle medie, $\boldsymbol{\mu} = (\mu_1, \mu_2)$, con μ_1 e $\mu_2 \in \mathbb{R}$, e la matrice di dispersione $\boldsymbol{\Sigma}$, che rappresenta la matrice delle varianze-covarianze, pari a:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & cov_{1,2} \\ cov_{2,1} & \sigma_2^2 \end{pmatrix}$$

dove $cov_{1,2} = cov_{2,1}$ individua la covarianza tra le variabili \tilde{X}_1 e \tilde{X}_2 , mentre σ^2 individua la varianza della v.a. considerata. La matrice di correlazione \mathbf{P} associata alla matrice di dispersione $\boldsymbol{\Sigma}$ è definita come,

$$\mathbf{P} = \begin{pmatrix} 1 & \rho_{1,2} = \frac{cov_{1,2}}{\sigma_1 \sigma_2} \\ \rho_{2,1} = \frac{cov_{2,1}}{\sigma_2 \sigma_1} & 1 \end{pmatrix}$$

in cui $\rho_{1,2} = \rho_{2,1}$ rappresenta il coefficiente di correlazione lineare di Pearson definito nella (2.1.1).

La densità della variabile Normale bidimensionale risulta,

$$f(x_1, x_2) = \frac{1}{2\pi} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}$$

⁷Una funzione $f : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ è detta *regularly varying* di indice $\alpha \in \mathbb{R}$ se per ogni $t > 0$ si ha: $\lim_{x \rightarrow \infty} \frac{f(tx)}{f(x)} = t^\alpha$.

dove la funzione $g(\cdot)$ della relazione (2.3.4) e la costante di normalizzazione data dalla (2.3.5) risultano rispettivamente,

$$g(t) = \exp\left(-\frac{1}{2}t\right) \quad C_n = \frac{1}{\pi^{n/2}2^{n/2}}$$

con n numero di dimensioni considerate, che nel caso in esame risulta essere pari a 2.

La rappresentazione stocastica della variabile Gaussiana, data da $\tilde{\mathbf{x}} \stackrel{d}{=} \boldsymbol{\mu} + \tilde{r}\mathbf{A}'\tilde{\mathbf{u}}$, nel caso bidimensionale risulta essere caratterizzata dalla variabile generatrice distribuita secondo una v.a. *Chi* con 2 gradi di libertà, $\tilde{r} \sim \chi_2$.

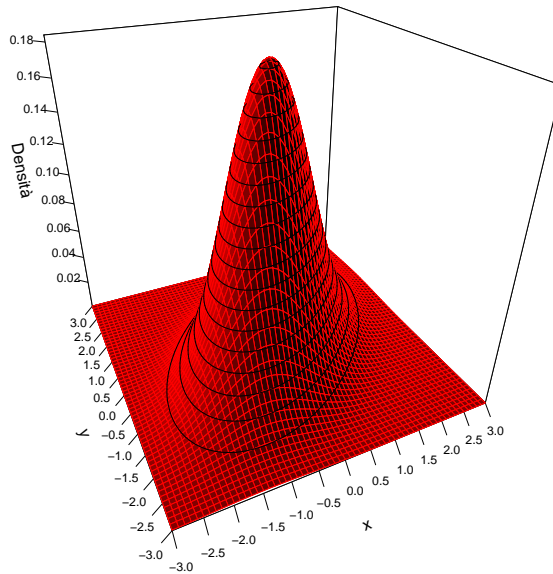


Figura 2.10: Funzione di densità della distribuzione Gaussiana bivariata di parametri $\boldsymbol{\mu} = \mathbf{0}$ e matrice di dispersione $\boldsymbol{\Sigma}$ di elementi $\Sigma_{1,1} = \Sigma_{2,2} = 1$ e $\Sigma_{1,2} = \Sigma_{2,1} = 0,5$.

La copula implicitamente definita dalla distribuzione Gaussiana può essere definita come segue:

$$C(u, v) = \Phi_{\mathbf{P}}(\Phi^{-1}(u), \Phi^{-1}(v))$$

dove $\Phi_{\mathbf{P}}$ rappresenta la funzione di ripartizione della variabile Normale bidimensionale $N(\mathbf{0}, \mathbf{P})$, mentre Φ individua la funzione di ripartizione della variabile unidimensionale Normale Standard⁸. La copula Gaussiana è una famiglia comprensiva in quanto, sotto opportune scelte del parametro ρ , coincide con la copula prodotto, con la copula comonotona e contromonotona:

- se $\rho = -1$, $C = W$;

⁸A fronte della proprietà delle copule di invarianza a trasformazioni strettamente crescenti delle v.a. marginali, segue che la copula della distribuzione $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ coincide con la copula della variabile $N(\mathbf{0}, \mathbf{P})$, ossia la distribuzione congiunta con variabili marginali standardizzate.

- se $\rho = 0$, $C = \Pi$;
- se $\rho = 1$, $C = M$.

Tale copula, inoltre, non possiede dipendenza di coda in quanto $\lambda_U = \lambda_L = 0$. La misura di concordanza definita dal ρ di Spearman, infine, è pari a,

$$\rho_S = \frac{6}{\pi} \arcsin\left(\frac{1}{2}\rho\right)$$

mentre il τ di Kendall è definito, come per tutte le copule ellittiche, dalla (2.3.8).

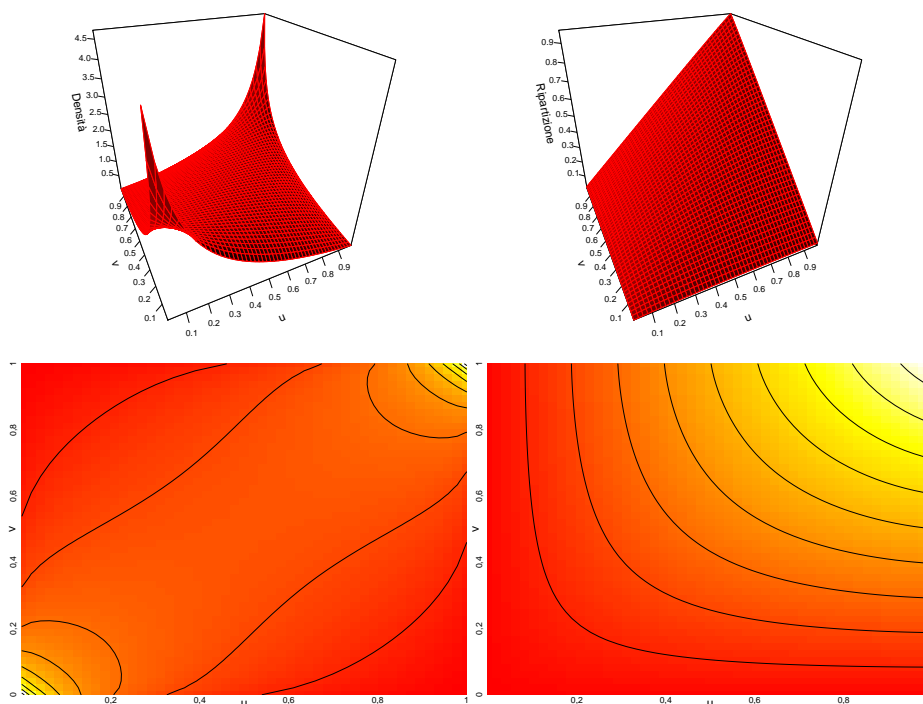


Figura 2.11: Densità e funzione di ripartizione, con rispettive curve di livello, della copula Gaussiana di parametro $\rho = 0,5$.

2.3.2 Distribuzione e copula t-student

La variabile casuale bidimensionale t-student è una distribuzione continua, definita su tutto \mathbb{R}^2 e possiede quale parametro caratteristico $m \in \mathbb{R}^+$ che rappresenta i gradi di libertà della distribuzione⁹. La densità della variabile t-student, $t(\boldsymbol{\mu}, \boldsymbol{\Sigma}, m)$ è la seguente:

$$f(x_1, x_2) = \frac{1}{\pi m} \frac{\Gamma(1 + \frac{m}{2})}{\Gamma(\frac{m}{2})} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \left\{ 1 + \frac{(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})}{m} \right\}^{-(1+m/2)}$$

⁹Nel caso particolare in cui $m = 1$ si ottiene la distribuzione Cauchy bivariata.

dove la funzione $g(\cdot)$ e la costante di normalizzazione sono rispettivamente,

$$g(t) = \left(1 + \frac{t}{m}\right)^{-(1+m/2)} \quad C_n = (\pi m)^{-\frac{n}{2}} \frac{\Gamma\left(\frac{1}{2}(n+m)\right)}{\Gamma\left(\frac{m}{2}\right)}$$

con n numero di dimensioni considerate. La rappresentazione stocastica della variabile t-student, data da $\tilde{\mathbf{x}} \stackrel{d}{=} \boldsymbol{\mu} + \tilde{r}\mathbf{A}'\tilde{\mathbf{u}}$, nel caso bidimensionale è definita tramite la variabile generatrice \tilde{r} distribuita come:

$$\frac{\tilde{r}^2}{m} \sim BetaType_{II}\left(1, \frac{m}{2}\right)$$

da cui,

$$\tilde{r} \sim \sqrt{BetaType_{II}\left(1, \frac{m}{2}\right) m}$$

dove $BetaType_{II}$ individua la distribuzione Beta del secondo tipo¹⁰. Si noti, inoltre, che nel caso in esame, il quadrato della variabile \tilde{r} è distribuito come segue $\tilde{r}^2 \sim 2F(2, m)$, dove F individua la variabile casuale F di Fisher; vale, infatti, la seguente relazione, $\frac{2}{m}F(2, m) = BetaType_{II}(1, \frac{m}{2})$.

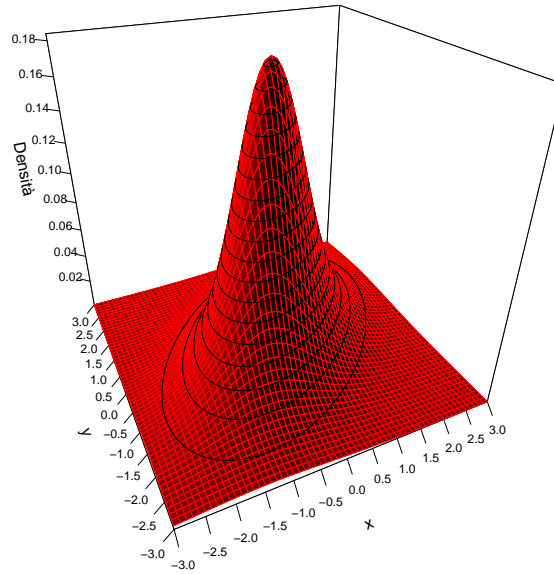


Figura 2.12: Funzione di densità della distribuzione t-student biviata di parametri $\boldsymbol{\mu} = \mathbf{0}$, $m = 4$ e matrice di dispersione $\boldsymbol{\Sigma}$ di elementi $\Sigma_{1,1} = \Sigma_{2,2} = 1$ e $\Sigma_{1,2} = \Sigma_{2,1} = 0$, 5.

La copula definita dalla distribuzione t-student possiede la seguente funzione di ripartizione:

¹⁰La variabile Beta del secondo tipo di parametri α e β è definita come $\frac{\tilde{X}}{1-\tilde{X}}$, dove \tilde{X} è la variabile Beta di parametri α e β .

$$C(u, v) = \mathbf{t}_{m, \mathbf{P}}(t_m^{-1}(u), t_m^{-1}(v))$$

dove $\mathbf{t}_{m, \mathbf{P}}$ individua la funzione di ripartizione della variabile t-student $t(\mathbf{0}, \mathbf{P}, m)$, con \mathbf{P} matrice di correlazione, mentre t_m è la funzione di ripartizione della variabile unidimensionale t-student standard.

La copula t-student, in analogia con quella Gaussiana, coincide con la copula contromonotona e comonotona per valori di ρ pari a -1 e 1 rispettivamente. Si noti, tuttavia che un valore di ρ pari a zero (con $m < \infty$) non determina la copula prodotto. Per $m \rightarrow \infty$, infine, la copula t-student coincide con quella Normale.

- se $\rho = -1$, $C = W$;
- se $\rho = 1$, $C = M$;
- se $m \rightarrow \infty$, $C = C^{Gaussiana}$.

La copula t-student possiede, per $\rho \neq -1$, dipendenza di coda; in particolare, il coefficiente di dipendenza di coda superiore e inferiore coincidono e sono pari a,

$$\lambda_U = \lambda_L = 2t_{m+1} \left(-\sqrt{\frac{(m+1)(1-\rho)}{1+\rho}} \right)$$

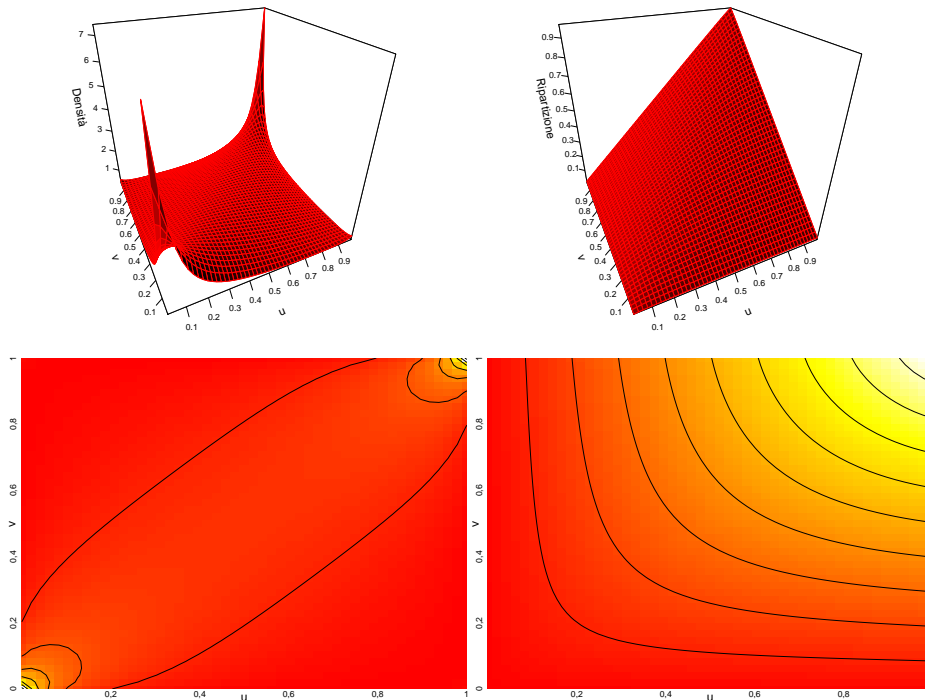


Figura 2.13: Densità e funzione di ripartizione, con rispettive curve di livello, della copula t-student di parametri $m = 4$ e $\rho = 0,5$.

2.3.3 Distribuzione e copula di Kotz

La variabile aleatoria di Kotz¹¹ è una distribuzione continua e definita su tutto il piano reale \mathbb{R}^2 . I parametri caratteristici della distribuzione sono $p, s \in \mathbb{R}^+$ e N tale per cui $2N + n > 2$, con n numero di dimensioni considerate. Considerando il caso in cui $n = 2$, si ha la seguente densità:

$$f(x_1, x_2) = \frac{s\Gamma\left(\frac{1}{2}\right)}{\pi^{1/2}\Gamma\left(\frac{N}{s}\right)} p^{N/s} |\Sigma|^{-\frac{1}{2}} [(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})]^{N-1} \exp\{-p[(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})]^s\}$$

dove la funzione $g(\cdot)$ e la costante di normalizzazione sono pari a,

$$g(t) = t^{N-1} \exp(-pt^s) \quad C_n = \frac{s\Gamma\left(\frac{n}{2}\right)}{\pi^{n/2}\Gamma\left(\frac{2N+n-2}{2s}\right)} p^{(2N+n-2)/2s}$$

Nel caso particolare in cui $N = s = 1$ e $p = \frac{1}{2}$, la distribuzione di Kotz corrisponde con la distribuzione Gaussiana. Nel caso in cui $s = 1$, che coincide con la distribuzione originaria formulata da Kotz, la variabile generatrice \tilde{r} della rappresentazione stocastica $\tilde{\mathbf{x}} \stackrel{d}{=} \boldsymbol{\mu} + \tilde{r}\mathbf{A}'\tilde{\mathbf{u}}$ risulta essere una variabile aleatoria Gamma, $\tilde{r} \sim \text{Gamma}(N, p)$.

Il grafico seguente riporta la funzione di densità della distribuzione bivariata; si noti che il parametro N determina la presenza o meno della concavità all'interno della distribuzione.

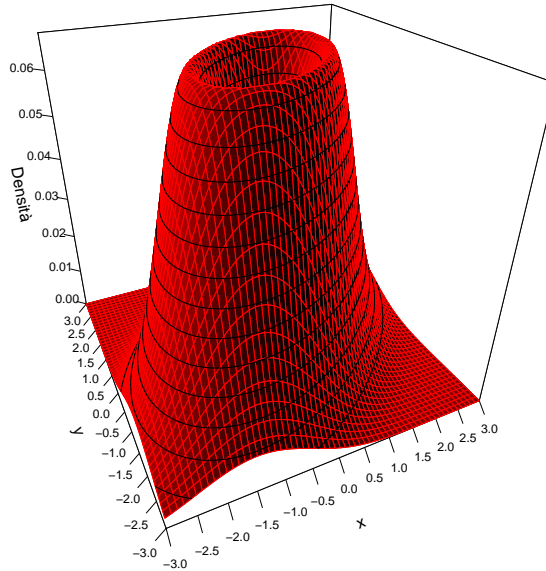


Figura 2.14: Distribuzione di Kotz bivariata di parametri $\boldsymbol{\mu} = \mathbf{0}$, $s = 1$, $N = 2$, $p = 0,5$ e matrice di dispersione Σ di elementi $\Sigma_{1,1} = \Sigma_{2,2} = 1$ e $\Sigma_{1,2} = \Sigma_{2,1} = 0,5$.

¹¹In Kollo e Roos (2003) è riportata una esaustiva trattazione della distribuzione di Kotz.

Il grafico seguente riporta la funzione di densità e di ripartizione della copula estratta, tramite simulazioni, dalla distribuzione della variabile in esame. Si noti che la copula implicitamente definita dalla distribuzione di Kotz non possiede, come tutte le copule ellittiche, forma analitica chiusa; la funzione di ripartizione tuttavia può essere espressa in termini generali ricorrendo alla relazione (2.3.6). Tale copula, infine, non possiede dipendenza di coda.

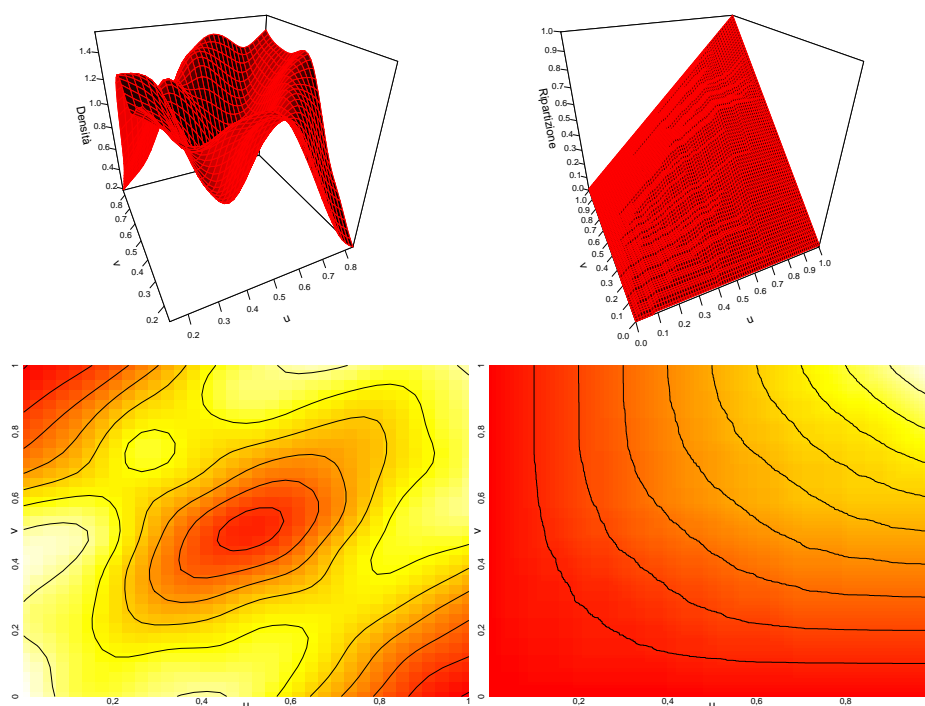


Figura 2.15: Densità e funzione di ripartizione, con rispettive curve di livello, della copula Kotz associata alla distribuzione di parametri $s = N = 1$, $p = 0,5$ e $\rho = 0,5$.

2.3.4 Distribuzione e copula di Pearson Tipo II

La variabile casuale Pearson Tipo II è una variabile continua definita sullo spazio di piano che soddisfa la relazione $(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \leq 1$. Il parametro caratteristico della distribuzione è $m > 1$. Nel caso bivariato, $n = 2$, la densità, nel suo campo di esistenza, risulta essere:

$$f(x_1, x_2) = \frac{m+1}{\pi} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} [1 - (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})]^m$$

dove la funzione $g(\cdot)$ e la costante di normalizzazione, nel caso generico n -dimensionale, sono pari a:

$$g(t) = (1-t)^m \quad C_n = \frac{\Gamma(n/2 + m + 1)}{\Gamma(m+1) \pi^{n/2}}$$

La variabile generatrice \tilde{r} che compare nella rappresentazione stocastica della v.a., $\tilde{\mathbf{x}} \stackrel{d}{=} \boldsymbol{\mu} + \tilde{r} \mathbf{A}' \tilde{\mathbf{u}}$, è distribuita come la radice quadrata della variabile Beta di parametri $n/2$ e $m+1$, in particolare: $\tilde{r} \sim \sqrt{\text{Beta}(n/2, m+1)}$.

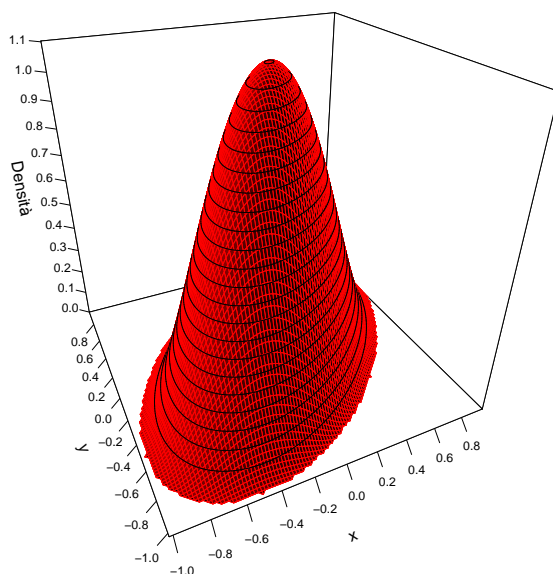


Figura 2.16: Funzione di densità della distribuzione Pearson Tipo II di parametri $\mu = \mathbf{0}$, $m = 2$, e matrice di dispersione Σ di elementi $\Sigma_{1,1} = \Sigma_{2,2} = 1$ e $\Sigma_{1,2} = \Sigma_{2,1} = 0,5$.

I grafici seguenti riportano le funzioni di densità e di ripartizione, ottenuti tramite simulazioni, della copula implicitamente definita dalla distribuzione Pearson Tipo II; tale copula, inoltre, non possiede dipendenza di coda.

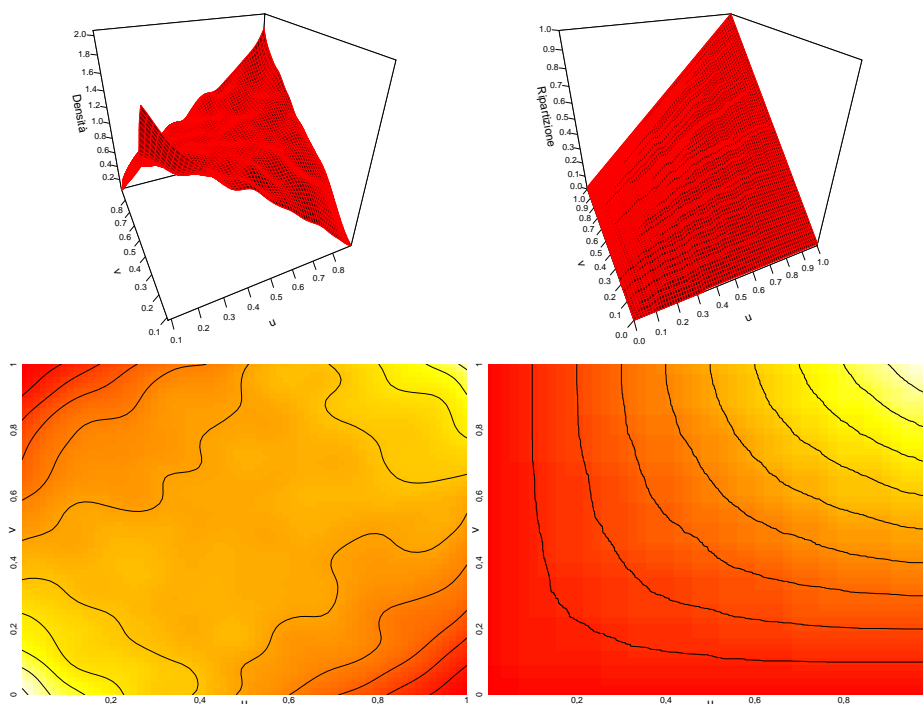


Figura 2.17: Densità, funzione di ripartizione e curve di livello della copula Pearson Tipo II associata alla rispettiva distribuzione di parametri $m = 2$ e $\rho = 0,5$.

2.3.5 Distribuzione e copula di Pearson Tipo VII

La distribuzione Pearson Tipo VII è una variabile casuale continua definita su tutto il piano \mathbb{R}^2 . Parametri caratteristici della distribuzione sono N , con $N > \frac{n}{2}$, e $m \in \mathbb{R}^+$.

Considerando il caso in cui $n = 2$, la densità della variabile dotata di vettore delle medie $\boldsymbol{\mu}$ e matrice di dispersione $\boldsymbol{\Sigma}$, assume la seguente forma:

$$f(x_1, x_2) = \frac{1}{\pi m} \frac{\Gamma(N)}{\Gamma(N-1)} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \left[1 + \frac{(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})}{m} \right]^{-N}$$

dove la funzione $g(\cdot)$ e la costante di normalizzazione risultano essere per $n \geq 2$:

$$g(t) = \left(1 + \frac{t}{m} \right)^{-N} \quad C_n = \frac{1}{(\pi m)^{-n/2}} \frac{\Gamma(N)}{\Gamma(N - n/2)}$$

La variabile generatrice della distribuzione ellittica di Pearson Tipo VII è legata alla variabile casuale Beta di secondo tipo, in particolare:

$$\frac{\tilde{r}^2}{m} \sim \text{BetaTypeII} \left(\frac{n}{2}, N - \frac{n}{2} \right)$$

$$\tilde{r} \sim \sqrt{\text{BetaTypeII} \left(\frac{n}{2}, N - \frac{n}{2} \right) m}$$

La distribuzione Pearson Tipo VII include, come casi particolari, la distribuzione t-student e la Cauchy. Per $N = \frac{1}{2}(m + n)$, infatti, si ottiene la t-student n -dimensionale, mentre per $N = \frac{1}{2}(n + 1)$ si determina la distribuzione Cauchy in n dimensioni.

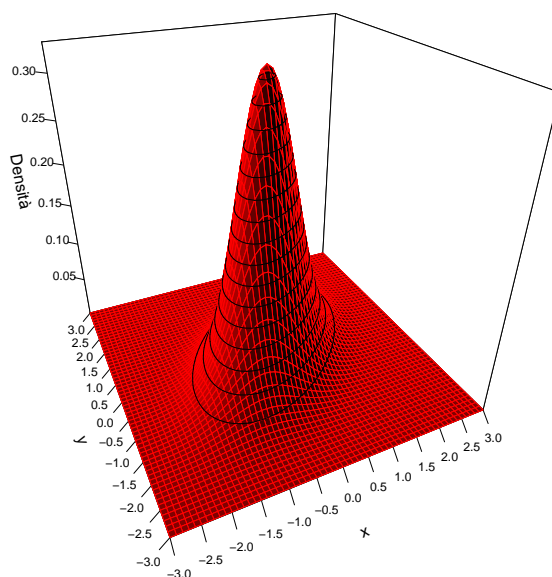


Figura 2.18: Densità della distribuzione di Pearson Tipo VII di parametri $\boldsymbol{\mu} = \mathbf{0}$, $N = 10$, $m = 10$, e matrice di dispersione $\boldsymbol{\Sigma}$ di elementi $\Sigma_{1,1} = \Sigma_{2,2} = 1$ e $\Sigma_{1,2} = \Sigma_{2,1} = 0,5$.

Di seguito si riportano i grafici caratteristici, ottenuti tramite simulazioni, della copula definita dalla distribuzione ellittica in esame; si noti che tale copula possiede dipendenza di coda.

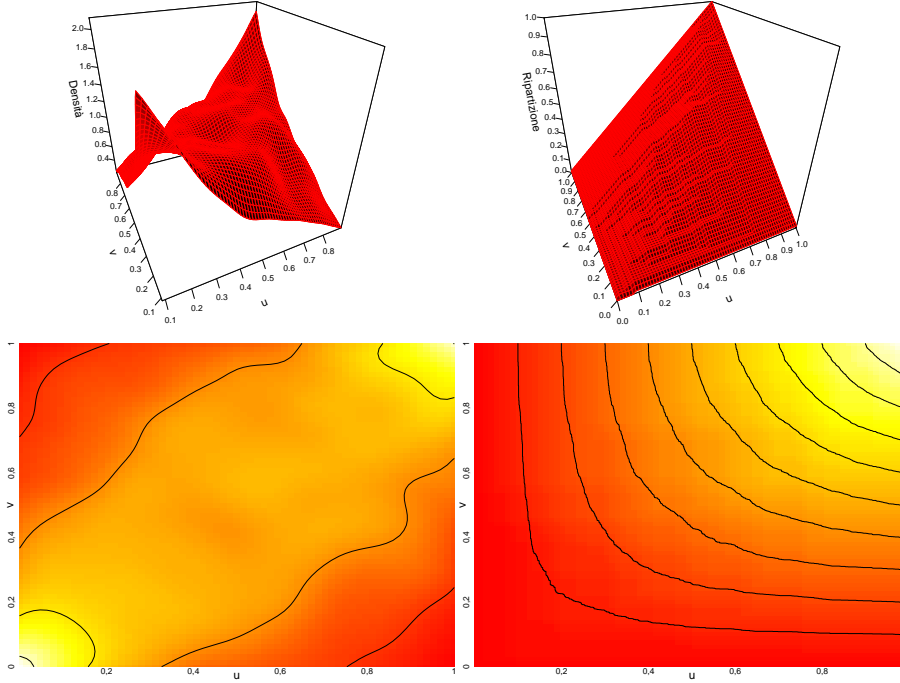


Figura 2.19: Densità, funzione di ripartizione e curve di livello, della copula Pearson Tipo VII estratta dalla rispettiva distribuzione di parametri $N = 10$, $m = 10$ e $\rho = 0, 5$.

2.3.6 Distribuzione e copula Exponential Power

La distribuzione Exponential Power è definita su tutto il piano \mathbb{R}^2 e dotata di parametro caratteristico $k \in \mathbb{R}^+$. La densità bivariata della distribuzione dotata di vettore delle medie $\boldsymbol{\mu}$ e matrice caratteristica $\boldsymbol{\Sigma}$ è la seguente:

$$f(x_1, x_2) = \frac{2}{\pi \Gamma\left(1 + \frac{2}{k}\right) 2^{1+2/k}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} [(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})]^{k/2}\right\}$$

dove la funzione $g(\cdot)$ e la costante di normalizzazione sono date da,

$$g(t) = \exp\left(-\frac{1}{2} t^{k/2}\right) \quad C_n = \frac{n \Gamma\left(\frac{n}{2}\right)}{\pi^{n/2} \Gamma\left(1 + \frac{n}{k}\right) 2^{1+n/k}}$$

Per $n = k = 2$ si ottiene la distribuzione Normale bivariata. Inoltre, la distribuzione Exponential Power è un caso particolare della distribuzione di Kotz quando in quest'ultima si hanno $N = 1$, $r = 1/2$ e $s = k/2$. Nella rappresentazione stocastica $\tilde{\mathbf{x}} \stackrel{d}{=} \boldsymbol{\mu} + \tilde{r} \mathbf{A}' \tilde{\mathbf{u}}$, la variabile generatrice \tilde{r} risulta essere una trasformazione della variabile Gamma (Solaro (2004)), in particolare: $\tilde{r} \sim [\text{Gamma}(1/k, 1/2)]^{1/k}$.

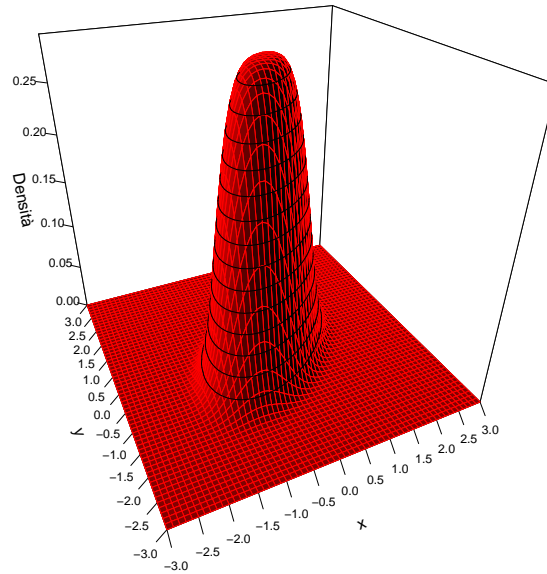


Figura 2.20: Funzione di densità della distribuzione Exponential Power di parametro $\mu = \mathbf{0}$, $k = 4$ e matrice di dispersione Σ di elementi $\Sigma_{1,1} = \Sigma_{2,2} = 1$ e $\Sigma_{1,2} = \Sigma_{2,1} = 0,5$.

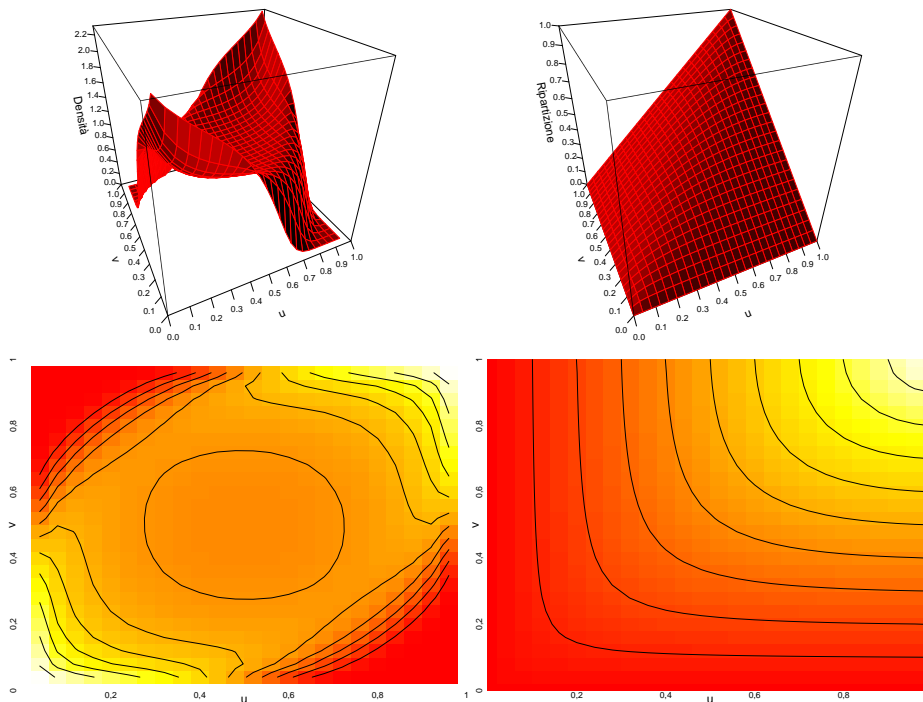


Figura 2.21: Densità, funzione di ripartizione e curve di livello, della copula Exponential Power definita dalla rispettiva distribuzione di parametri $k = 4$ e $\rho = 0,5$.

La copula Exponential Power, i cui grafici della funzione di densità e ripartizione sono riportati nella figura precedente, non possiede dipendenza di coda;

ne sono un esempio, infatti, la copula Gaussiana e di Kotz di cui la Exponential Power ne è un caso particolare.

2.3.7 Distribuzione e copula di Laplace

La distribuzione bivariata di Laplace¹² è una variabile continua definita su tutto il piano \mathbb{R}^2 i cui parametri risultano essere il vettore delle medie $\boldsymbol{\mu}$ e la matrice di dispersione $\boldsymbol{\Sigma}$. La densità risulta essere,

$$f(x_1, x_2) = \frac{1}{\pi} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} K_0 \left(\sqrt{2(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})} \right)$$

dove K_0 è la funzione di Bessel modificata del terzo tipo¹³. La funzione $g(\cdot)$ e la costante di normalizzazione, per il caso generico n -dimensionale, risultano essere:

$$g(t) = K_0(\sqrt{2t}) \quad C_n = \frac{1}{2^{n/2-1} \pi^{n/2} \Gamma(n/2)}$$

Considerando la rappresentazione stocastica $\tilde{\mathbf{x}} \stackrel{d}{=} \boldsymbol{\mu} + \tilde{r} \mathbf{A}' \tilde{\mathbf{u}}$, nel caso bidimensionale, la variabile \tilde{r} possiede la seguente densità:

$$f(r) = \frac{2r K_0(\sqrt{2}r)}{\sqrt{\pi}}, \quad r > 0$$

Per generare dei numeri casuali distribuiti secondo una Laplace bidimensionale risulta, tuttavia, più agevole sfruttare la relazione,

$$\tilde{\mathbf{x}} \stackrel{d}{=} \tilde{Z}^{\frac{1}{2}} \tilde{\mathbf{Y}}$$

dove \tilde{Z} individua la variabile casuale unidimensionale esponenziale di media unitaria, $\tilde{Z} \sim Exp(1)$, mentre $\tilde{\mathbf{Y}}$ la variabile aleatoria bidimensionale Normale dotata di vettore delle medie nullo e matrice di covarianza $\boldsymbol{\Sigma}$, $\tilde{\mathbf{Y}} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$.

I due grafici seguenti mostrano la funzione di densità della variabile bidimensionale in esame, nonché la rispettiva copula ellittica; si noti che quest'ultima non possiede dipendenza di coda.

¹²Per una trattazione approfondita si rimanda a Kotz, Kozubowski e Podgórski (2001a,b).

¹³La funzione di Bessel modificata del terzo tipo è la seguente: $K_\lambda(x) = \frac{1}{2} \left(\frac{x}{2}\right)^\lambda \int_0^\infty t^{-\lambda-1} \exp\left(-t - \frac{x^2}{4t}\right) dt$ con $x > 0$.

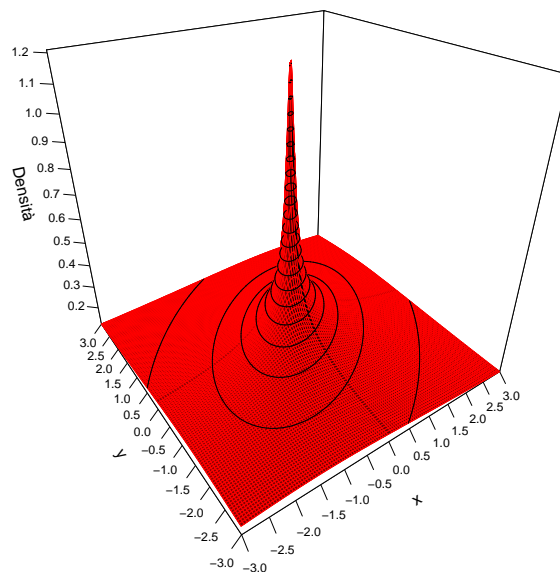


Figura 2.22: Funzione di densità della distribuzione Laplace con matrice di dispersione Σ di elementi $\Sigma_{1,1} = \Sigma_{2,2} = 1$ e $\Sigma_{1,2} = \Sigma_{2,1} = 0,5$.

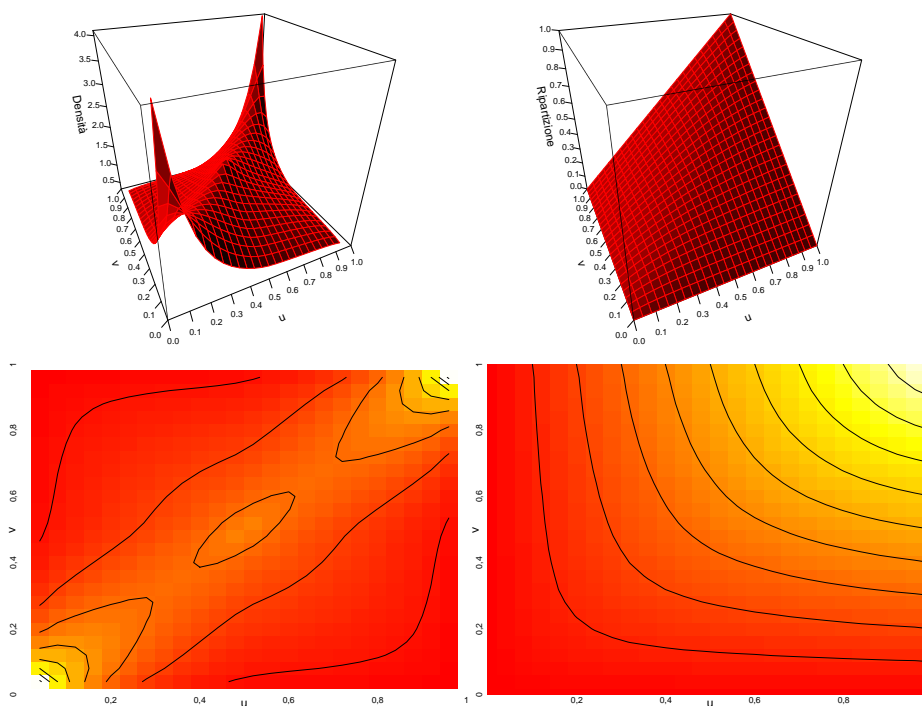


Figura 2.23: Densità e funzione di ripartizione, con curve di livello, della copula Laplace associata alla rispettiva distribuzione dotata di $\rho = 0,5$.

2.4 Le copule archimedee

Le copule archimedee costituiscono una classe di copule ampiamente studiata in letteratura e utilizzata nella pratica attuariale in quanto permettono di modellizzare molteplici strutture di dipendenza e, inoltre, sono di agevole costruzione. La definizione di tali copule, infatti, si basa su una sola funzione definita *generatore* della copula. In particolare, sia φ una funzione continua e strettamente decrescente da $[0, 1] \rightarrow [0, \infty]$, tale che $\varphi(1) = 0$ e $\varphi(0) \leq \infty$. La pseudo-inversa di φ è la funzione $\varphi^{[-1]}$ di dominio $[0, \infty]$ data da,

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{[-1]}(t), & 0 \leq t \leq \varphi(0) \\ 0, & \varphi(0) \leq t \leq \infty \end{cases}$$

Inoltre, se $\varphi(0) = \infty$, vale che $\varphi^{[-1]} = \varphi^{-1}$. Se la funzione φ è convessa allora,

$$C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v)) \quad (2.4.1)$$

è una copula bivariata. Tutte le copule che soddisfano la relazione precedente sono dette *archimedee*.

Le copule archimedee, quindi, possono essere costruite sfruttando la relazione (2.4.1), in quanto è sufficiente trovare una opportuna funzione φ che sia continua, decrescente e convessa da $[0, 1]$ a $[0, \infty]$ con $\varphi(1) = 0$. Le copule archimedee C di generatore φ soddisfano le seguenti proprietà:

- C è simmetrica, $C(u, v) = C(v, u) \quad \forall u, v \in \mathbb{I}$;
- C è associativa, $C[C(u, v), w] = C[u, C(v, w)] \quad \forall u, v, w \in \mathbb{I}$;
- se $c > 0$ è una costante, allora $c\varphi$ è ancora un generatore di C .

La funzione di ripartizione della copula archimedea $C(u, v)$, inoltre, può essere espressa come,

$$K_C(t) = t - \frac{\varphi(t)}{\varphi'(t^+)} \quad (2.4.2)$$

Il τ di Kendall e la dipendenza di coda di tale classe di copule possono essere formulati in termini della funzione generatrice φ :

$$\tau = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt$$

mentre (Nelsen (2006, p. 215)),

$$\lambda_L = \lim_{x \rightarrow \infty} \frac{\varphi^{[-1]}(2x)}{\varphi^{[-1]}(x)} \quad \lambda_U = 2 - \lim_{x \rightarrow 0^+} \frac{1 - \varphi^{[-1]}(2x)}{1 - \varphi^{[-1]}(x)}$$

Nel seguito si riportano le principale famiglie di copule archimedee uniparametriche e biparametriche studiate in letteratura e utilizzate a fini pratici¹⁴; in tale contesto i parametri caratteristici delle famiglie di copule considerate sono individuati con θ e δ .

2.4.1 Gumbel-Hougaard

Tale famiglia di copule è caratterizzata dal parametro $\theta \in (1, \infty)$ ed è dotata della seguente funzione di ripartizione,

$$C(u, v) = \exp \left\{ - \left[(-\ln u)^\theta + (-\ln v)^\theta \right]^{\frac{1}{\theta}} \right\}$$

con generatore $\varphi(t) = (-\ln t)^\theta$. Il τ di Kendall è definito come:

$$\tau = \frac{\theta - 1}{\theta}$$

La copula non possiede dipendenza di coda inferiore, $\lambda_L = 0$, mentre il coefficiente di dipendenza di coda superiore è pari a $\lambda_U = 2 - 2^{\frac{1}{\theta}}$. La famiglia di copule di Gumbel-Hougaard gode delle seguenti proprietà:

- se $\theta = 1$, $C = \Pi$;
- se $\theta \rightarrow \infty$, $C = M$.

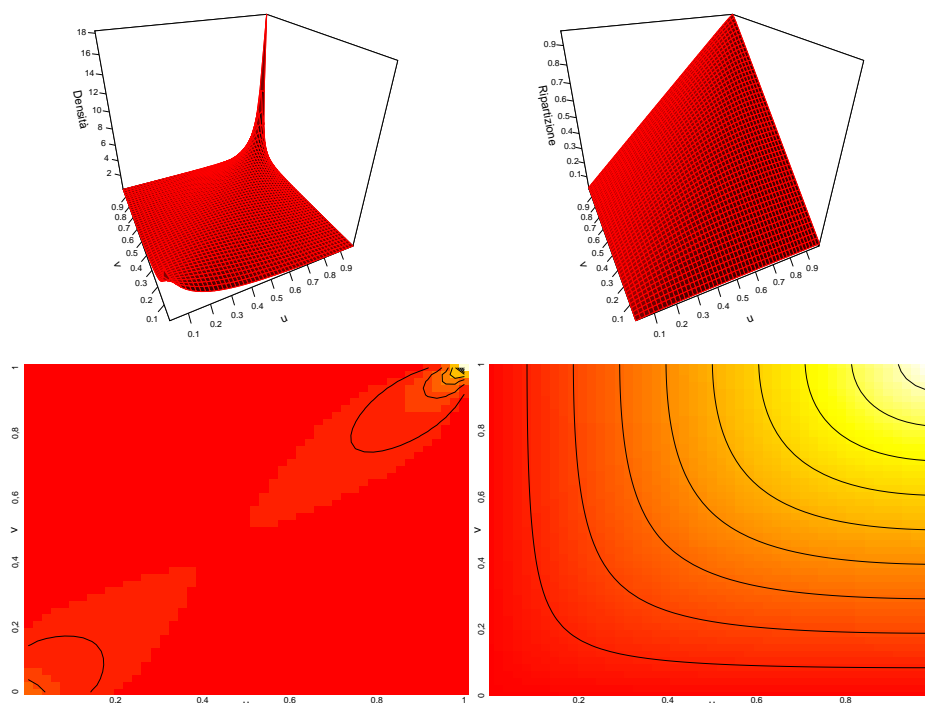


Figura 2.24: Densità e funzione di ripartizione, con rispettive curve di livello, della copula Gumbel-Hougaard di parametro $\theta = 2$.

¹⁴Per una completa trattazione del tema si rimanda a Joe (1997) e Nelsen (2006).

2.4.2 Ali-Mikhail-Haq

La famiglia di copule Ali-Mikhail-Haq dipende dal parametro θ definito nell'intervallo $[-1, 1]$. La funzione di ripartizione della copula è la seguente,

$$C(u, v) = \frac{uv}{1 - \theta(1-u)(1-v)}$$

il cui generatore risulta essere $\varphi(t) = \ln \frac{1-\theta(1-t)}{t}$. Il τ di Kendall è pari a,

$$\tau = \frac{3\theta - 2}{3\theta} - \frac{2(1-\theta)^2}{3\theta^2} \ln(1-\theta)$$

ed è compreso nell'intervallo $[\frac{5-8\ln 2}{3}, \frac{1}{3}] \cong [-0, 1817; 0, 3333]$; il grado di dipendenza che è possibile modellizzare con tale copula, quindi, risulta contenuto. La copula in esame, inoltre, non possiede dipendenza di coda, $\lambda_L = \lambda_U = 0$. La copula Ali-Mikhail-Haq, infine, gode delle seguenti proprietà:

- se $\theta = 0$, $C = \Pi$;
- se $\theta = 1$

$$C = \frac{uv}{u+v-uv} = \frac{\Pi}{\Sigma - \Pi}$$

con $\Sigma = u + v$.

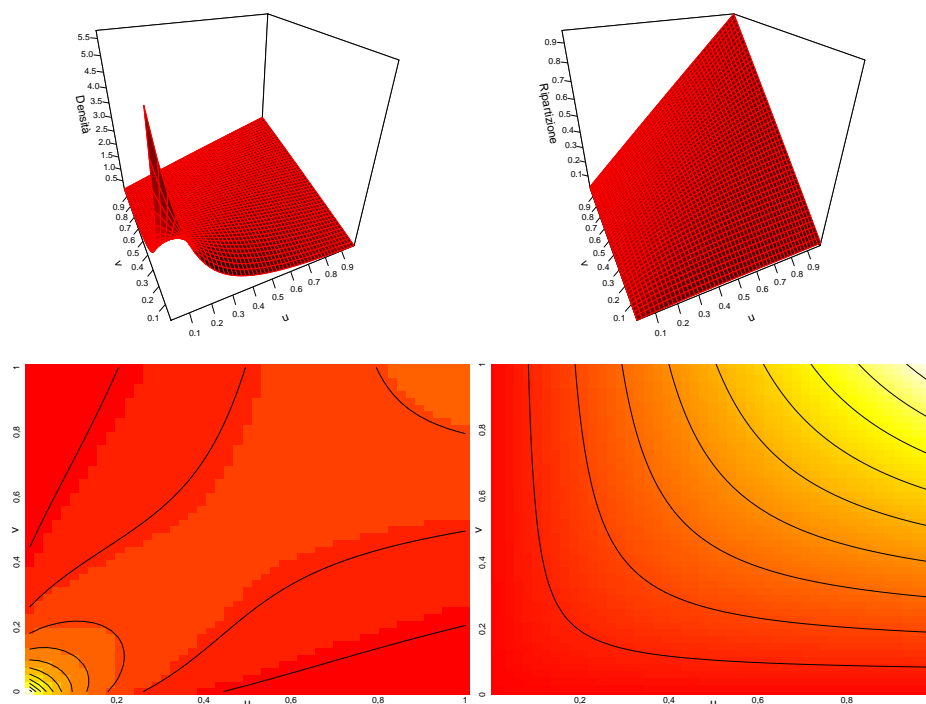


Figura 2.25: Densità e funzione di ripartizione, con rispettive curve di livello, della copula Ali-Mikhail-Haq di parametro $\theta = 0, 9$.

2.4.3 Clayton

La copula Clayton possiede parametro θ , dotato di dominio $[-1, 0) \cup (0, \infty)$, e seguente funzione di ripartizione:

$$C(u, v) = [\max(u^{-\theta} + v^{-\theta} - 1; 0)]^{-\frac{1}{\theta}}$$

Il generatore della copula è $\varphi(t) = \frac{1}{\theta}(t^{-\theta} - 1)$, mentre il τ di Kendall è definito come,

$$\tau = \frac{\theta}{\theta + 2}$$

La copula, inoltre, possiede dipendenza di coda inferiore pari a $\lambda_L = 2^{-\frac{1}{\theta}}$ per $\theta \geq 0$. La famiglia di copule Clayton è comprensiva e gode delle seguenti proprietà:

- se $\theta = -1$, $C = W$;
- se $\theta \rightarrow 0$, $C = \Pi$;
- se $\theta = 1$, $C = \frac{\Pi}{\Sigma - \Pi}$;
- se $\theta \rightarrow \infty$, $C = M$.

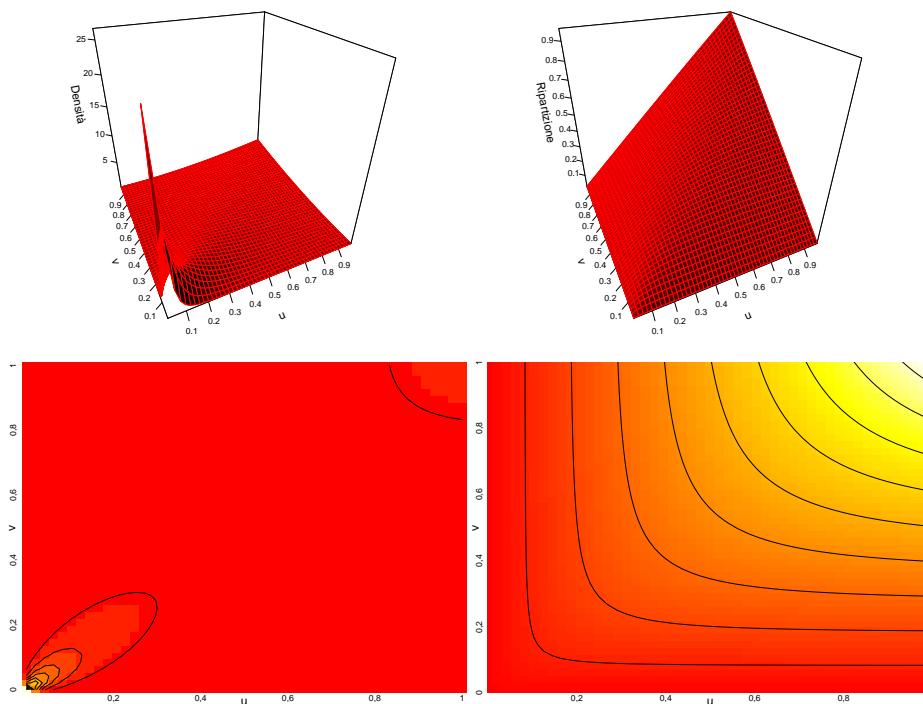


Figura 2.26: Densità e funzione di ripartizione, con rispettive curve di livello, della copula Clayton di parametro $\theta = 2$.

2.4.4 Frank

Il parametro θ da cui dipende tale famiglia di copule è definito per valori appartenenti all'insieme $(-\infty, 0) \cup (0, \infty)$. La forma funzionale della copula è la seguente:

$$C(u, v) = -\frac{1}{\theta} \ln \left[1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{(e^{-\theta} - 1)} \right]$$

dotata di generatore,

$$\varphi(t) = -\ln \left(\frac{e^{-\theta t} - 1}{e^{-\theta} - 1} \right)$$

Il τ di Kendall è definito come,

$$\tau = 1 - \frac{4}{\theta} [1 - D_1(\theta)]$$

dove $D_k(x)$ è la funzione di Debye, pari a $D_k(x) = \frac{k}{x^k} \int_0^x \frac{t^k}{e^t - 1} dt$. La copula di Frank non possiede dipendenza di coda ed è una famiglia comprensiva:

- se $\theta \rightarrow -\infty$, $C = W$;
- se $\theta \rightarrow 0$, $C = \Pi$;
- se $\theta \rightarrow \infty$, $C = M$.

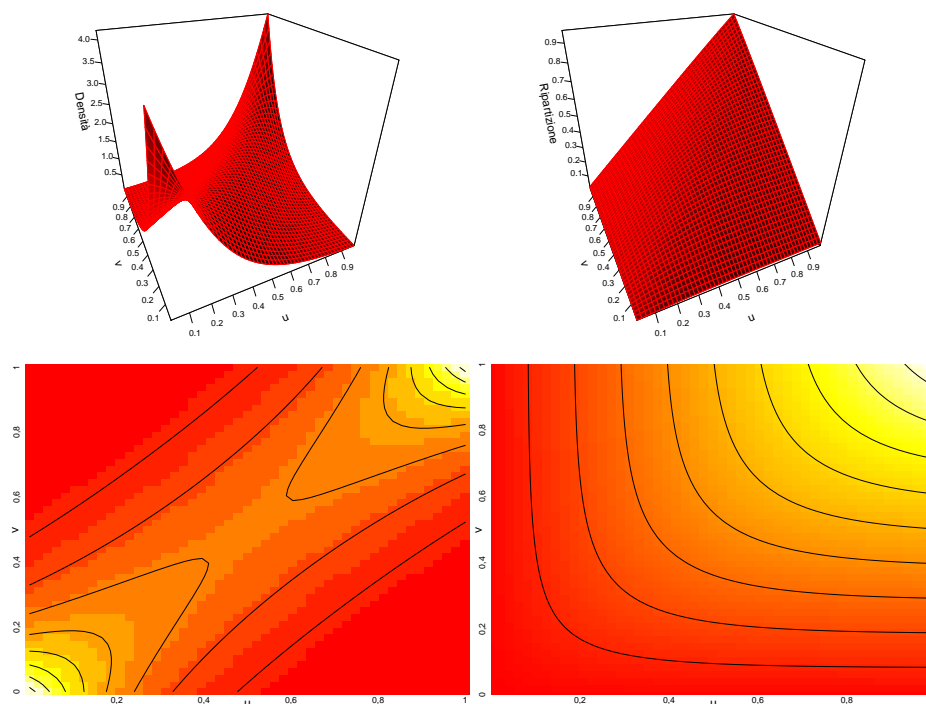


Figura 2.27: Densità e funzione di ripartizione, con rispettive curve di livello, della copula Frank di parametro $\theta = 5$.

2.4.5 Joe

Il parametro caratteristico della copula Joe è definito per valori maggiori o uguali dell'unità, $\theta \in [1, \infty)$. La funzione di ripartizione assume la seguente forma,

$$C(u, v) = 1 - (\bar{u}^\theta + \bar{v}^\theta - \bar{u}^\theta \bar{v}^\theta)^{\frac{1}{\theta}}$$

dove $\bar{u} = 1 - u$ e $\bar{v} = 1 - v$, mentre il generatore della copula risulta essere $\varphi(t) = -\ln[1 - (1 - t)^\theta]$. La misura di dipendenza di concordanza definita dal τ di Kendall risulta essere,

$$\tau = 1 + \frac{4}{\theta^2} \int_0^1 t \ln(t) (1 - t)^{2(1-\theta)/\theta} dt$$

mentre la dipendenza di coda superiore è pari a $\lambda_U = 2 - 2^{\frac{1}{\theta}}$. La famiglia di copule di Joe gode delle seguenti proprietà:

- se $\theta = 1$, $C = \Pi$;
- se $\theta \rightarrow \infty$, $C = M$.

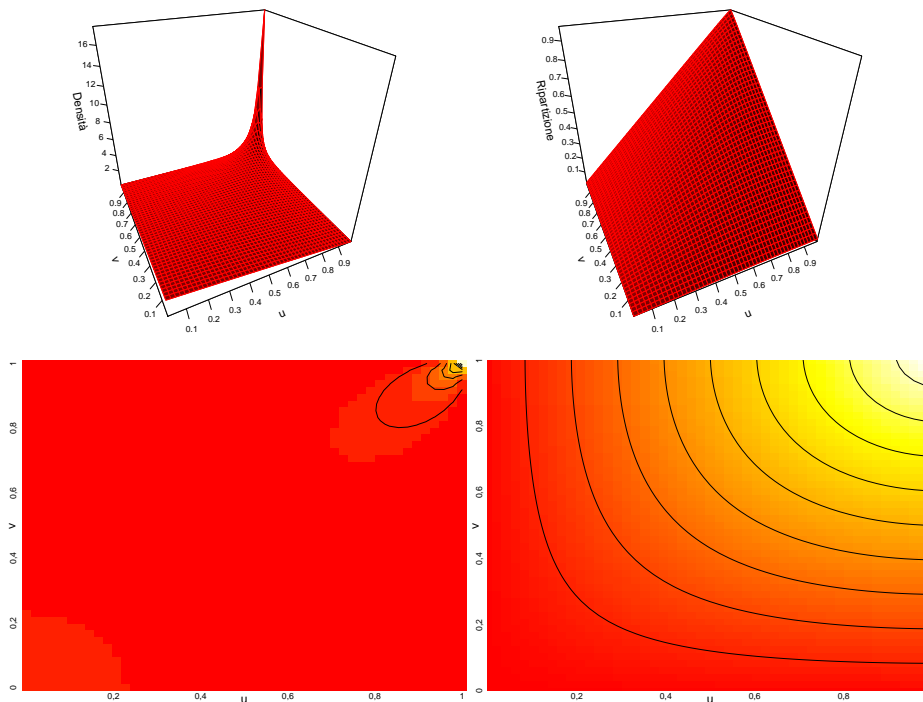


Figura 2.28: Densità e funzione di ripartizione, con rispettive curve di livello, della copula Joe di parametro $\theta = 2$.

2.4.6 BB1

Tale famiglia di copule archimedee, denominata anche come *Clayton generalizzata*, è dotata di due parametri $\theta \in (0, \infty)$ e $\delta \in [1, \infty)$. La funzione di ripartizione è pari a,

$$C(u, v) = \left\{ [(u^{-\theta} - 1)^\delta + (v^{-\theta} - 1)^\delta]^{\frac{1}{\delta}} + 1 \right\}^{-\frac{1}{\theta}}$$

mentre la funzione generatrice della copula risulta essere $\varphi(t) = (t^\theta - 1)^\delta$. La misura di concordanza definita dal τ di Kendall vale,

$$\tau = \frac{(2 + \theta)\delta - 2}{(2 + \theta)\delta}$$

La famiglia di copule BB1 è caratterizzata da dipendenza di coda sia superiore che inferiore, in particolare: $\lambda_L = 2^{-\frac{1}{\theta\delta}}$ e $\lambda_U = 2 - 2^{\frac{1}{\delta}}$. Le proprietà di cui gode, infine, sono:

- se $\theta \rightarrow 0$ e $\delta \rightarrow 1$, $C = \Pi$;
- se $\theta \rightarrow \infty$ o $\delta \rightarrow \infty$, $C = M$.

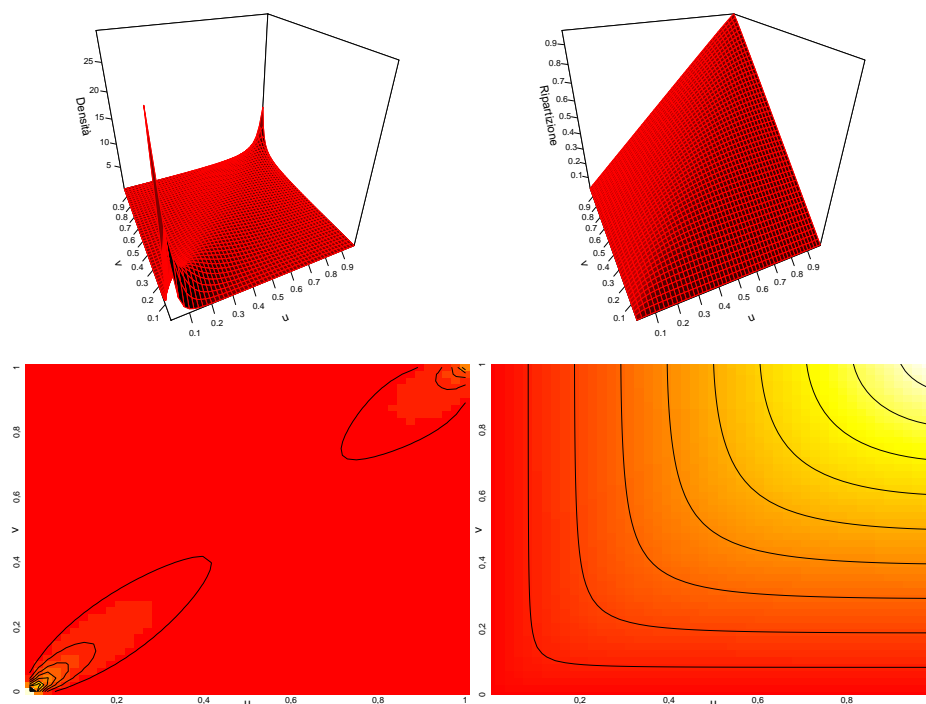


Figura 2.29: Densità e funzione di ripartizione, con rispettive curve di livello, della copula BB1 di parametri $\theta = \delta = 1, 5$.

2.4.7 BB6

La copula BB6 è una famiglia biparametrica in cui $\theta, \delta \in [1, \infty)$. La funzione di ripartizione è la seguente:

$$C(u, v) = 1 - \left(1 - \exp \left\{ - \left[(-\ln(1 - \bar{u}^\theta))^\delta + (-\ln(1 - \bar{v}^\theta))^\delta \right]^{\frac{1}{\delta}} \right\} \right)^{\frac{1}{\theta}}$$

dove $\bar{u} = 1 - u$ e $\bar{v} = 1 - v$ e il generatore della copula è $\varphi(t) = 1 - [1 - \exp(-t^{\frac{1}{\delta}})]^{\frac{1}{\theta}}$. La misura di concordanza definita dal τ di Kendall risulta essere,

$$\tau = 1 + 4 \int_0^1 \frac{-\ln(-(1-t)^\theta + 1)(1-t - (1-t)^{-\theta} + t(1-t)^{-\theta})}{\theta\delta} dt$$

La copula BB6 presenta, infine, solo dipendenza di coda superiore, pari a $\lambda_U = 2 - \frac{1}{\theta\delta}$. Tale famiglia gode delle seguenti proprietà:

- se $\theta = 1$, C coincide con la copula di Gumbel-Hougaard;
- se $\theta \rightarrow \infty$ o $\delta \rightarrow \infty$, $C = M$.

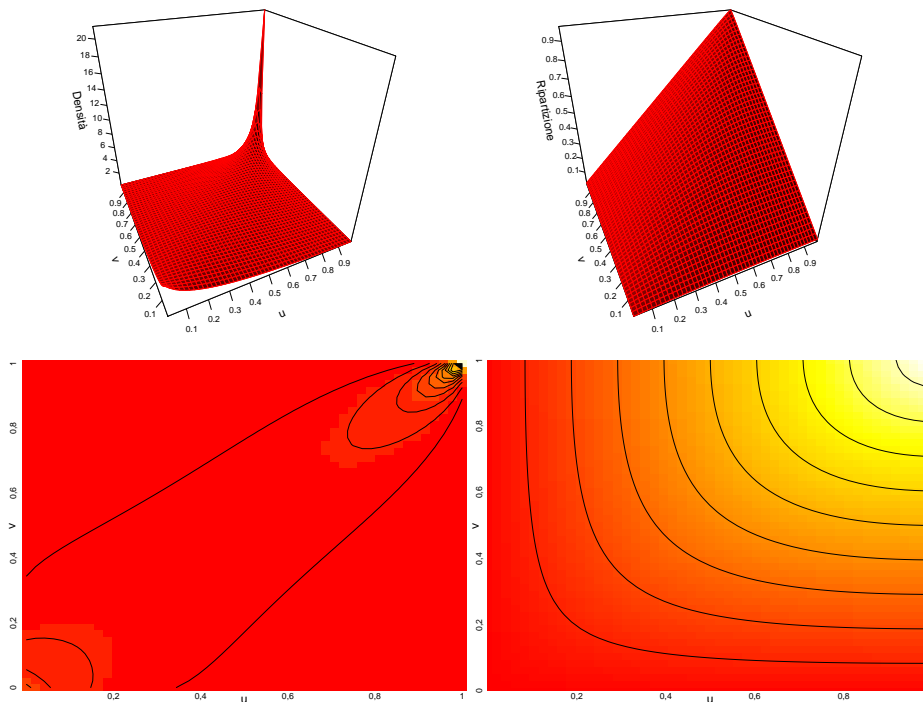


Figura 2.30: Densità e funzione di ripartizione, con rispettive curve di livello, della copula BB6 di parametri $\theta = \delta = 1,5$.

2.4.8 BB7

La famiglia di copule BB7 è caratterizzata dai parametri $\theta \in [1, \infty)$ e $\delta \in [0, \infty)$. La funzione di ripartizione è pari a,

$$C(u, v) = 1 - \left(1 - [(1 - \bar{u}^\theta)^{-\delta} + (1 - \bar{v}^\theta)^{-\delta} - 1]^{-\frac{1}{\delta}} \right)^{\frac{1}{\theta}}$$

in cui $\bar{u} = 1 - u$ e $\bar{v} = 1 - v$ e la funzione generatrice della copula è $\varphi(t) = 1 - [1 - (1 + t)^{-\frac{1}{\theta}}]^{-\frac{1}{\delta}}$. Il τ di Kendall assume la seguente forma funzionale,

$$\tau = 1 + 4 \int_0^1 \frac{[(1 - (1 - t)^\theta)^{-\delta}] - 1}{[-\theta\delta(1 - t)^{\theta-1}(1 - (1 - t)^\theta)^{-\delta-1}] dt}$$

I coefficienti di dipendenza di coda inferiore e superiore, invece, risultano essere rispettivamente $\lambda_L = 2^{-\frac{1}{\delta}}$ e $\lambda_U = 2 - 2^{-\frac{1}{\theta}}$. La copula, inoltre, coincide con il limite superiore di Fréchet se uno dei due parametri tende ad infinito:

- se $\theta \rightarrow \infty$ o $\delta \rightarrow \infty$, $C = M$.

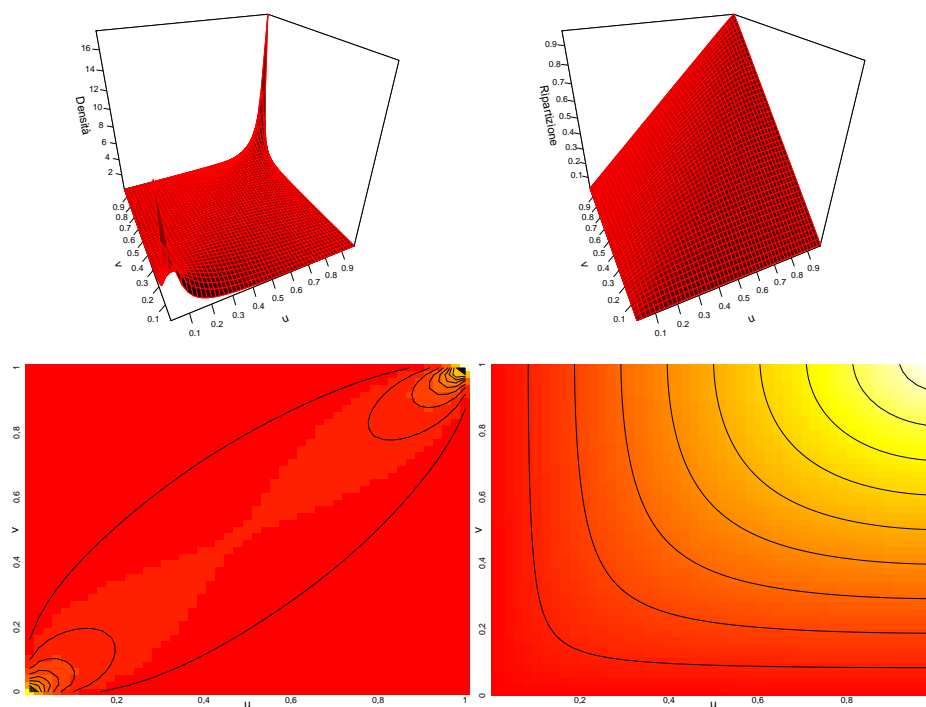


Figura 2.31: Densità e funzione di ripartizione, con rispettive curve di livello, della copula BB7 di parametri $\theta = 2$ e $\delta = 1$.

2.4.9 BB8

Tale famiglia di copule, dotata di parametri $\theta \in [1, \infty)$ e $\delta \in [0, 1]$, ha funzione di ripartizione della seguente forma:

$$C(u, v) = \delta^{-1} \left[1 - \left\{ 1 - [1 - (1 - \delta)^\theta]^{-1} [1 - (1 - \delta u)^\theta] [1 - (1 - \delta v)^\theta] \right\}^{\frac{1}{\theta}} \right]$$

in cui il generatore è dato dalla funzione $\varphi(t) = \delta^{-1} \left[1 - \{1 - [1(1 - \delta)^\theta]e^{-t}\}^{\frac{1}{\theta}} \right]$. La misura di concordanza definita dal τ di Kendall è,

$$\tau = 1 + 4 \int_0^1 \frac{-\ln((1 - t\delta)^\theta - 1)(1 - t\delta - (1 - t\delta)^{-\theta} + (1 - t\delta)^{-\theta}t\delta)}{[(1 - \delta)^\theta - 1]\theta\delta} dt$$

La copula presenta dipendenza di coda superiore solo nel caso in cui $\delta = 1$; in tale circostanza vale che $\lambda_U = 2 - 2^{\frac{1}{\theta}}$. Inoltre, la copula gode della seguente proprietà:

- se $\theta \rightarrow 1$ o $\delta \rightarrow 0$, $C = \Pi$.

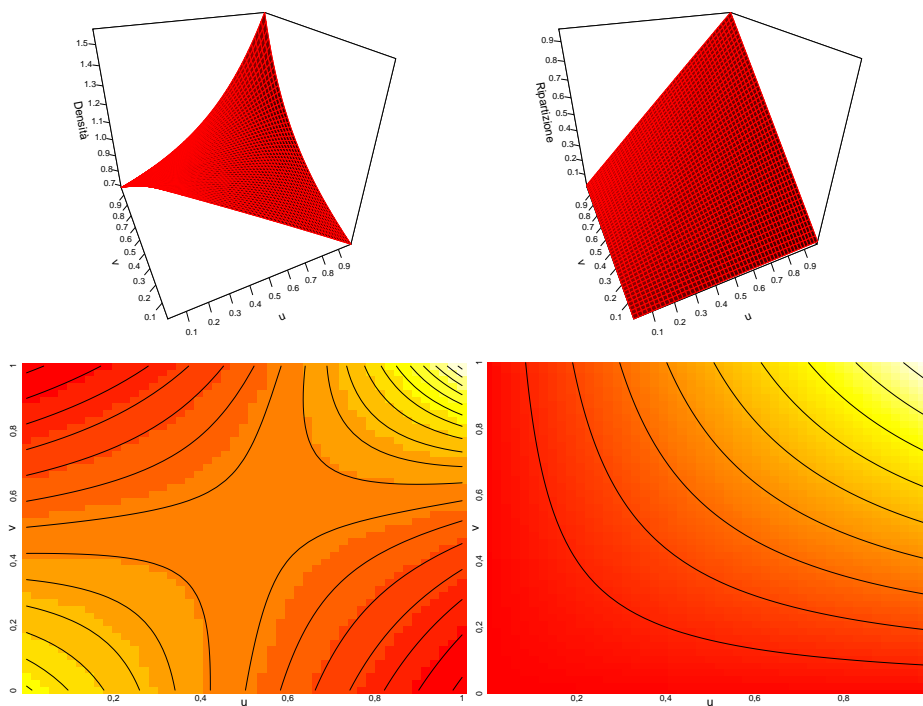


Figura 2.32: Densità e funzione di ripartizione, con rispettive curve di livello, della copula BB8 di parametri $\theta = 2$ e $\delta = 0,5$.

2.5 Altre famiglie di copule

2.5.1 Fréchet

La famiglia di copule di Fréchet definisce una classe di copule biparametriche in cui i parametri caratteristici sono $\theta, \delta \in [0, 1]$, tali per cui $\theta + \delta \leq 1$. La forma funzionale della copula risulta essere la seguente,

$$C(u, v) = \theta M(u, v) + (1 - \theta - \delta)\Pi(u, v) + \delta W(u, v)$$

dove $W(u, v) = \max(u + v - 1; 0)$ e $M(u, v) = \min(u, v)$ rappresentano rispettivamente il limite inferiore e superiore di Fréchet, mentre $\Pi(u, v) = uv$ individua la copula prodotto. La famiglia di copule di Fréchet è comprensiva in quanto permette di ottenere le copule W , Π e M che corrispondono ai casi di perfetta dipendenza negativa (contromonotonicità), indipendenza e perfetta dipendenza positiva (comonotonicità) tra due variabili aleatorie. Valgono, infatti, le seguenti proprietà:

- se $\theta = 0$ e $\delta = 1$, $C = W$;
- se $\theta = \delta = 0$, $C = \Pi$;
- se $\theta = 1$ e $\delta = 0$, $C = M$.

Le copule W e M sono singolari, in quanto $\frac{\partial^2 W}{\partial u \partial v} = 0$ e $\frac{\partial^2 M}{\partial u \partial v} = 0$, e hanno come supporto rispettivamente la diagonale secondaria, data da $u = 1 - v$, e la diagonale principale, pari a $u = v$, del quadrato di lato unitario definito da \mathbb{I}^2 . La copula Π , di contro, è assolutamente continua e definita su tutto il piano \mathbb{I}^2 .

Le copule $C = W$ e $C = \Pi$ sono archimedee e dotate rispettivamente di generatore $\varphi(t) = 1 - t$ e $\varphi(t) = -\ln(t)$. Il τ di Kendall della famiglia di copule di Fréchet è pari a,

$$\tau = \frac{(\theta - \delta)(\theta + \delta + 2)}{3}$$

mentre i coefficienti di dipendenza di coda inferiore e superiore coincidono e sono pari a $\lambda_L = \lambda_U = \theta$.

Una famiglia di copule legata a quella di Fréchet risulta essere quella di Mardia; tale classe di copule dipende dal solo parametro $\theta \in [-1, 1]$,

$$C(u, v) = \frac{\theta^2(1 + \theta)}{2}M(u, v) + (1 - \theta^2)\Pi(u, v) + \frac{\theta^2(1 - \theta)}{2}W(u, v)$$

Anche tale famiglia è comprensiva in quanto per valori del parametro pari a -1, 0 e 1 si ottengono rispettivamente le copule W , Π e M .

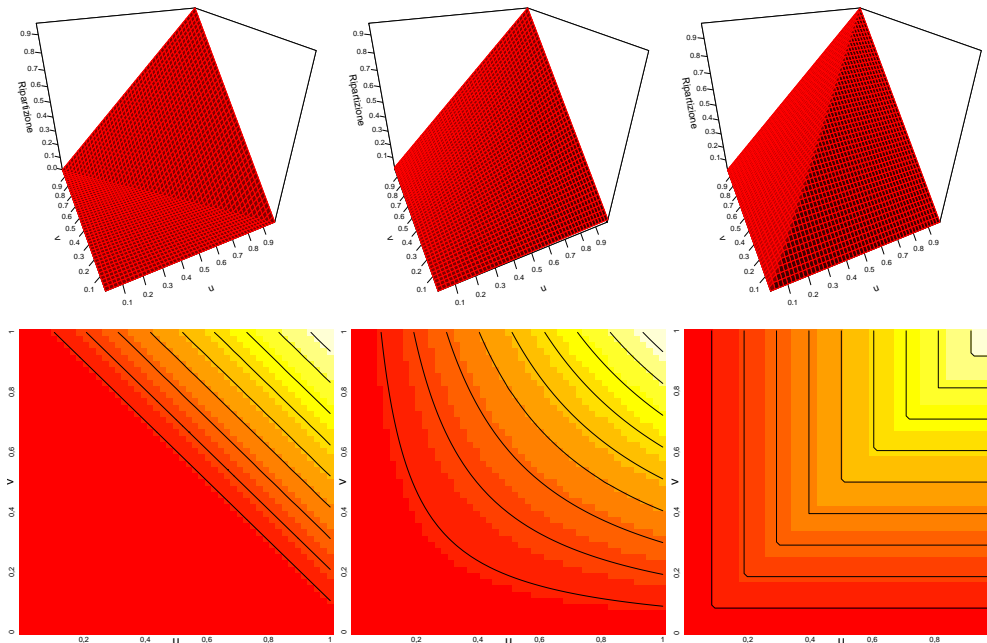


Figura 2.33: Densità e funzione di ripartizione, con rispettive curve di livello, della copula di Fréchet di parametri $\theta = 0$ e $\delta = 1$ (grafico a sinistra, $C = W$), $\theta = \delta = 0$ (grafico centrale, $C = \Pi$) e $\theta = 1$ e $\delta = 0$ (grafico a destra, $C = M$).

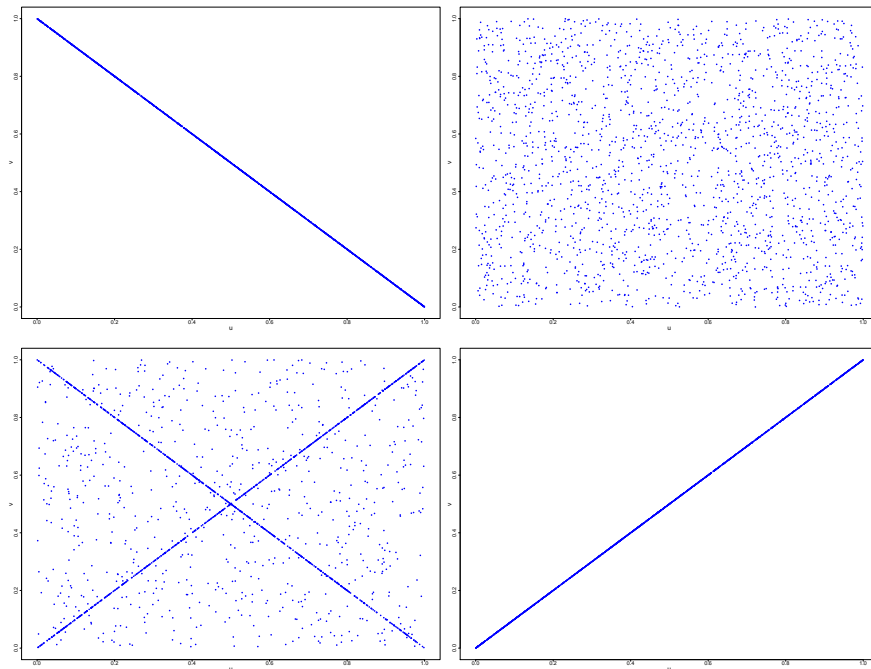


Figura 2.34: 2.000 realizzazioni casuali dalla copula di Fréchet per diverse combinazioni di parametri. In particolare, $\theta = 0$ e $\delta = 1$ (grafico in alto a sinistra), $\theta = \delta = 0$ (grafico in alto a destra), $\theta = \delta = 0,3$ (grafico in basso a sinistra), $\theta = 1$ e $\delta = 0$ (grafico in basso a destra).

2.5.2 Marshall-Olkin

Tale famiglia è caratterizzata da due parametri, θ e $\delta \in [0, 1]$, e possiede la seguente funzione di ripartizione:

$$C(u, v) = \min(u^{1-\theta}v, uv^{1-\delta}) = \begin{cases} u^{1-\theta}v, & u^\theta \geq v^\delta \\ uv^{1-\delta}, & u^\theta \leq v^\delta \end{cases}$$

La copula è dotata sia di una componente assolutamente continua che di una componente singolare, individuata dalla curva $u^\theta = v^\delta$. Il τ di Kendall è definito come,

$$\tau = \frac{\theta\delta}{\theta - \theta\delta + \delta}$$

La copula mostra dipendenza di coda superiore pari a $\lambda_U = \min(\theta, \delta)$. Per tale famiglia di copule valgono i seguenti risultati:

- se $\theta = \delta = 0$, $C = \Pi$;
- se $\theta = \delta = 1$, $C = M$.

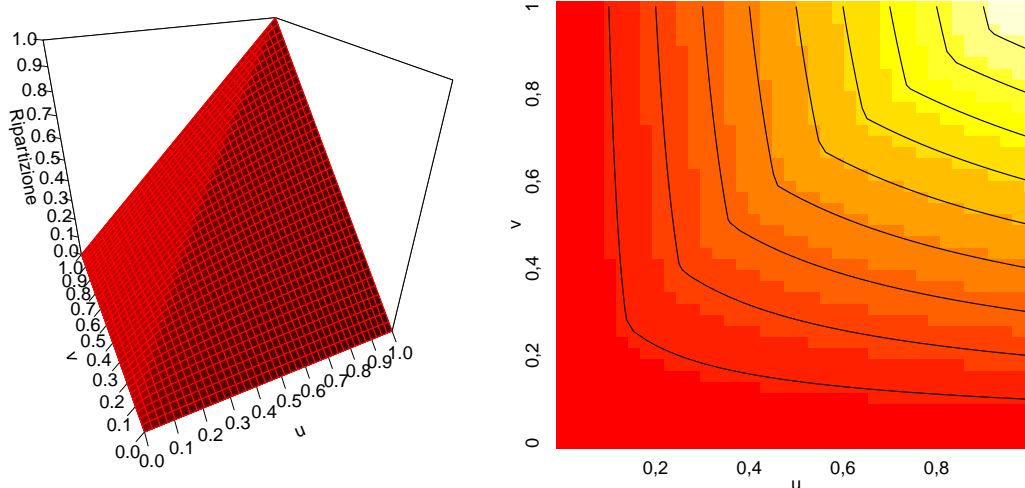


Figura 2.35: Funzione di ripartizione e curve di livello della copula di Marshall-Olkin di parametri $\theta = 0,5$ e $\delta = 0,75$.

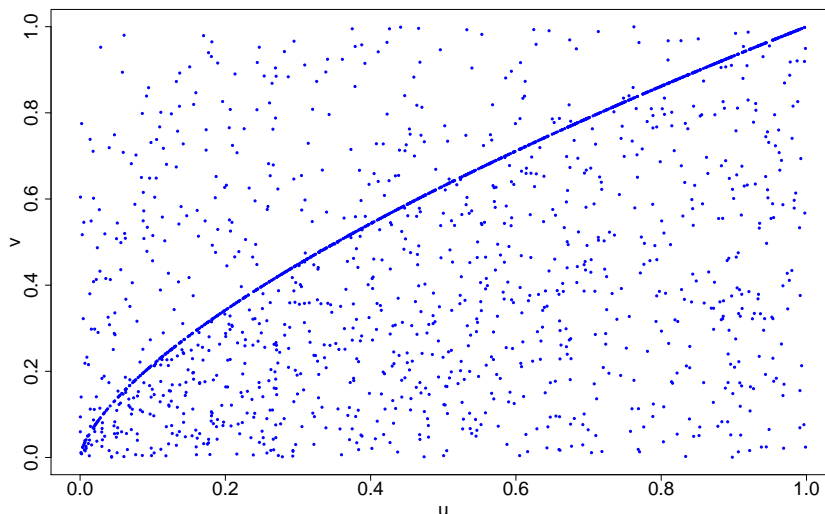


Figura 2.36: 2.000 realizzazioni casuali dalla copula di Marshall-Olkin di parametri $\theta = 0,5$ e $\delta = 0,75$; risulta evidente la singolarità della copula definita dall'insieme dei punti in \mathbb{I}^2 che soddisfano la relazione $u^{0,5} = v^{0,75}$.

2.5.3 Farlie-Gumbel-Morgenstern

Questa famiglia definisce una classe di copule uniparametriche, dotate di parametro $\theta \in [-1, 1]$ e forma funzionale,

$$C(u, v) = uv + \theta uv(1 - u)(1 - v)$$

Tale copula presenta la caratteristica di potere modellizzare solo dipendenze contenute; il τ di Kendall, infatti, definito come,

$$\tau = \frac{2\theta}{9}$$

è lineare rispetto al parametro θ ed è compreso nell'intervallo $[-\frac{2}{9}, \frac{2}{9}]$. Quando $\theta = 0$, inoltre, la copula coincide con la copula prodotto, $C = \Pi$, determinando quindi l'indipendenza tra le variabili marginali. Tale famiglia di copule, infine, non possiede dipendenza di coda.

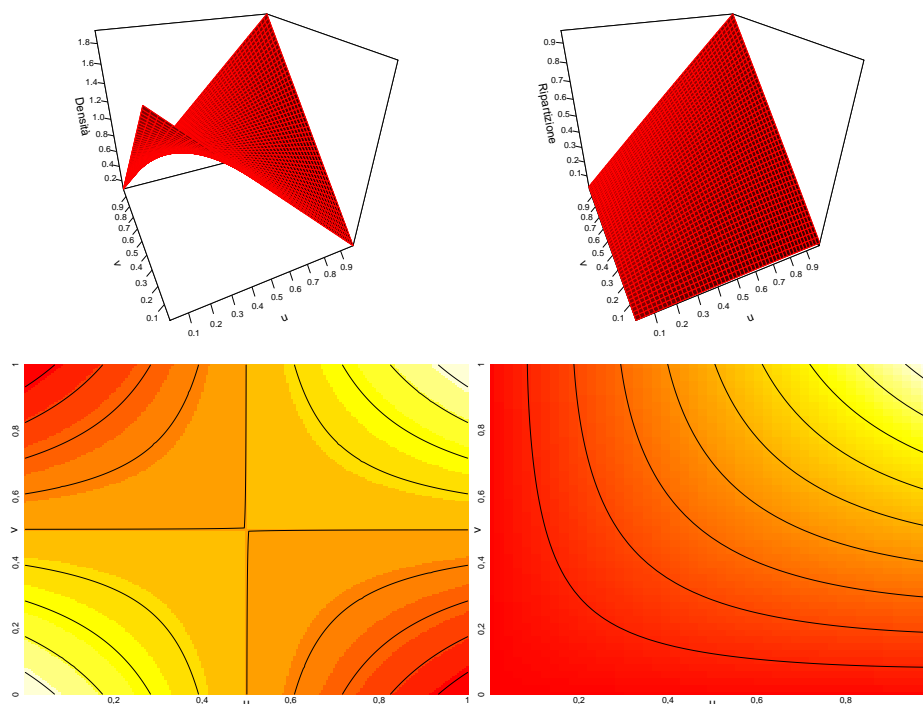


Figura 2.37: Densità e funzione di ripartizione, con rispettive curve di livello, della copula Farlie-Gumbel-Morgenstern di parametro $\theta = 1$.

2.5.4 Plackett

La classe di copule di Plackett è definita dal solo parametro $\theta \in (0, 1) \cup (1, \infty)$. La funzione di ripartizione è la seguente:

$$C(u, v) = \frac{[1 + (\theta - 1)(u + v)] - \sqrt{[1 + (\theta - 1)(u + v)]^2 - 4uv\theta(\theta - 1)}}{2(\theta - 1)}$$

La copula non presenta dipendenza di coda e per il τ di Kendall non esiste una forma analitica in grado di esprimerlo in funzione del parametro θ . Tale famiglia è comprensiva in quanto valgono i seguenti risultati:

- se $\theta \rightarrow 0^+$, $C = W$;
- se $\theta \rightarrow 1$, $C = \Pi$;
- se $\theta \rightarrow \infty$, $C = M$.

I grafici seguenti mostrano la funzione di ripartizione e di densità della copula.

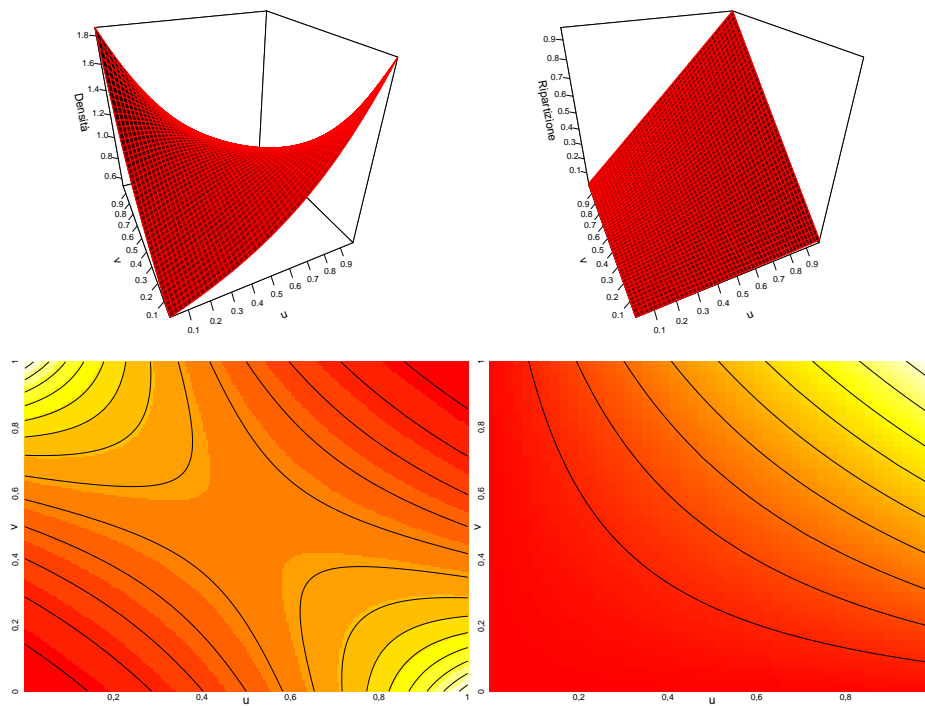


Figura 2.38: Densità e funzione di ripartizione, con rispettive curve di livello, della copula di Plackett di parametro $\theta = 0,5$.

Capitolo 3

La pair-copula construction e le vine copule

Negli ultimi decenni l'avvento di nuovi contesti normativi, quali ad esempio Solvency II nel settore assicurativo, nonché, in ambito finanziario e di risk management, l'esigenza di una adeguata misurazione dei rischi hanno contribuito al notevole sviluppo di modelli stocastici per analizzare la dipendenza esistente tra molteplici fonti di rischio. Le copule costituiscono lo strumento metodologico fondamentale alla base di tali modelli.

Le principali famiglie di copule esistenti in letteratura, quali quelle ellittiche e archimedee, risultano, in genere, inadeguate a descrivere in modo flessibile la dipendenza in contesti multivariati (Berg e Aas (2009), Kurowicka e Joe (2011)). Di contro, tali tipologie di copule, tipicamente sviluppate in un contesto bidimensionale, sono in grado di modellizzare molteplici forme di dipendenza tra due rischi (Joe (1997) e Nelsen (2006)).

Le *vine copule* sono copule multivariate basate su un costruito metodologico che coniuga lo strumento grafico dei *vine* (Bedford e Cooke (2001, 2002)), rappresentato da una struttura ad albero, con copule bidimensionali ottenute tramite la *pair-copula construction* (PCC) (Joe (1996)). Quest'ultima permette di fattorizzare la densità di una variabile aleatoria multivariata per mezzo di copule bivariate, descrivendo, quindi, la dipendenza esistente tra le n fonti di rischio ricorrendo a sole copule in $n = 2$ dimensioni. Il *vine*, infine, consente di organizzare graficamente le svariate scomposizioni della dipendenza multivariata che è possibile ottenere per mezzo della *pair-copula construction*.

Le *vine copule*, quindi, definiscono una copula multivariata utilizzando sole copule bidimensionali, permettendo di modellizzare in modo estremamente flessibile la struttura di dipendenza esistente tra le n variabili marginali oggetto di analisi.

Il capitolo ha l'obiettivo di introdurre la *pair-copula construction*, nonché la cosiddetta "ipotesi semplificata"; quest'ultima rende trattabile da un punto di vista inferenziale la stima delle copule bidimensionali utilizzate per fattorizzare la densità multivariata. In seguito, si presenta il concetto di *vine copula*, ponendo particolare enfasi ad uno specifico sottoinsieme di *regular vine* definito dai *C-vine*

e D-vine. Infine, si mostrano gli approcci e gli algoritmi esistenti in letteratura volti a selezionare e calibrare il modello vine più adatto a descrivere la dipendenza esistente tra i dati multivariati oggetto di studio.

3.1 La pair-copula construction

L'esigenza di scomporre la funzione di densità di una variabile multivariata tramite distribuzioni marginali bivariate compare per la prima volta nei lavori di Joe (1996, 1997). Nella letteratura seguente la fattorizzazione di una variabile aleatoria multidimensionale tramite sole copule bivariate ha preso il nome di pair-copula construction (Aas et al. (2009), Berg e Aas (2009)).

Prima di intraprendere la trattazione della pair-copula construction, è opportuno esporre i seguenti risultati in merito alla riformulazione del teorema di Sklar e delle sue implicazioni nel contesto di variabili aleatorie bidimensionali condizionate rispetto ad un vettore di variabili casuali (Palaro e Hotta (2006), Mesfioui e Quesy (2008)).

In termini generali, si consideri la seguente variabile n -dimensionale $(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n)$; riformulando la relazione (2.2.5), la funzione di ripartizione della v.a. bidimensionale $(\tilde{X}_1, \tilde{X}_2)$ condizionata al vettore in $n - 2$ dimensioni di v.a. $(\tilde{X}_3, \dots, \tilde{X}_n)$ può essere definita come¹:

$$H_{1,2|3,\dots,n}(x_1, x_2|x_3, \dots, x_n) = C_{1,2|3,\dots,n}(F_{1|3,\dots,n}(x_1|x_3, \dots, x_n), F_{2|3,\dots,n}(x_2|x_3, \dots, x_n)|x_3, \dots, x_n) \quad (3.1.1)$$

Analogamente alla relazione (2.2.6), la copula sottostante alla rispettiva funzione di ripartizione condizionata $H_{1,2|3,\dots,n}(x_1, x_2|x_3, \dots, x_n)$ risulta,

$$C_{1,2|3,\dots,n}(x_1, x_2|x_3, \dots, x_n) = H_{1,2|3,\dots,n}(F_{1|3,\dots,n}^{-1}(x_1|x_3, \dots, x_n), F_{2|3,\dots,n}^{-1}(x_2|x_3, \dots, x_n)|x_3, \dots, x_n) \quad (3.1.2)$$

La densità della distribuzione $H_{1,2|3,\dots,n}(x_1, x_2|x_3, \dots, x_n)$, invece, similmente a quanto visto nella (2.2.7), può essere formula come:

¹In tale capitolo, al fine di uniformare la notazione con la letteratura inerente la pair-copula construction e le vine copule, per le funzioni di ripartizione e densità n -variate si utilizzano le seguenti formulazioni: $F_{\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n}(x_1, x_2, \dots, x_n) = F_{1,2,\dots,n}(x_1, x_2, \dots, x_n)$ e $f_{\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n}(x_1, x_2, \dots, x_n) = f_{1,2,\dots,n}(x_1, x_2, \dots, x_n)$. In modo equivalente, per le distribuzioni condizionate si avrà: $F_{\tilde{X}_1, \tilde{X}_2|\tilde{X}_3, \dots, \tilde{X}_n}(x_1, x_2|x_3, \dots, x_n) = F_{1,2|3,\dots,n}(x_1, x_2|x_3, \dots, x_n)$ e $f_{\tilde{X}_1, \tilde{X}_2|\tilde{X}_3, \dots, \tilde{X}_n}(x_1, x_2|x_3, \dots, x_n) = f_{1,2|3,\dots,n}(x_1, x_2|x_3, \dots, x_n)$. Notazione analoga, dove possibile, verrà anche utilizzata per le rispettive funzioni copula.

$$\begin{aligned}
h_{1,2|3,\dots,n}(x_1, x_2|x_3, \dots, x_n) = \\
c_{1,2|3,\dots,n}(F_{1|3,\dots,n}(x_1|x_3, \dots, x_n), F_{2|3,\dots,n}(x_2|x_3, \dots, x_n)|x_3, \dots, x_n) \quad (3.1.3) \\
\cdot f_{1|3,\dots,n}(x_1|x_3, \dots, x_n) \cdot f_{2|3,\dots,n}(x_2|x_3, \dots, x_n)
\end{aligned}$$

Infine, dalla relazione (2.2.8), la densità bivariata di una copula condizionata rispetto ad un vettore di v.a., $c_{1,2|3,\dots,n}(u_1, u_2|u_3, \dots, u_n)$, con \tilde{U}_i v.a. uniformemente distribuite in $[0, 1]$, è definita come segue:

$$c_{1,2|3,\dots,n}(u_1, u_2|u_3, \dots, u_n) = \frac{\partial^2 C_{1,2|3,\dots,n}(u_1, u_2|u_3, \dots, u_n)}{\partial u_1 \partial u_2} \quad (3.1.4)$$

Si consideri una v.a. n -dimensionale $\tilde{\mathbf{X}} = (\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n)$, avente funzione di ripartizione congiunta $F_{1,2,\dots,n}(x_1, x_2, \dots, x_n)$, con variabili marginali continue e dotate di funzione di ripartizione $F_i(x_i)$, con $i = 1, \dots, n$. La funzione di densità della variabile $\tilde{\mathbf{X}}$, ricorrendo alla relazione (2.2.7), può essere formulata esplicitandone la copula n -dimensionale:

$$f_{1,2,\dots,n}(x_1, x_2, \dots, x_n) = c_{1,2,\dots,n}(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) \cdot f_1(x_1) \cdot f_2(x_2) \cdot \dots \cdot f_n(x_n) \quad (3.1.5)$$

dove $c_{1,2,\dots,n}$ individua la funzione di densità della copula in n dimensioni univocamente determinata, come mostrato dal teorema di Sklar (2.2.5), dalla funzione di ripartizione della v.a. $\tilde{\mathbf{X}}$.

Obiettivo della pair-copula construction è quello di scrivere la relazione precedente scomponendo la funzione di densità della copula n -dimensionale tramite copule bidimensionali. A tal fine, è possibile osservare che la densità multivariata di una v.a. può essere scritta come il prodotto delle funzioni di densità delle variabili unidimensionali condizionate²:

$$f_{1,2,\dots,n}(x_1, x_2, \dots, x_n) = f_1(x_1) \cdot f_{2|1}(x_2|x_1) \cdot \dots \cdot f_{n|1,2,\dots,n-1}(x_n|x_1, x_2, \dots, x_{n-1}) \quad (3.1.6)$$

infatti,

$$f_{1,2,\dots,n}(x_1, x_2, \dots, x_n) = f_1(x_1) \cdot \frac{f_{1,2}(x_1, x_2)}{f_1(x_1)} \cdot \dots \cdot \frac{f_{1,2,\dots,n}(x_1, x_2, \dots, x_n)}{f_{1,2,\dots,n-1}(x_1, x_2, \dots, x_{n-1})}$$

Considerando, a titolo esemplificativo, il caso in cui $n = 3$, si ha:

$$f_{1,2,3}(x_1, x_2, x_3) = f_1(x_1) \cdot f_{2|1}(x_2|x_1) \cdot f_{3|1,2}(x_3|x_1, x_2)$$

²La decomposizione è unica a meno di una permutazione delle variabili.

Ricorrendo alla relazione (2.2.7), le funzioni di densità univariate condizionate possono essere espresse in termini di copule bivariate. Per $f_{2|1}(x_2|x_1)$, infatti, vale:

$$f_{2|1}(x_2|x_1) = \frac{f_{1,2}(x_1, x_2)}{f_1(x_1)} = \frac{c_{1,2}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1) \cdot f_2(x_2)}{f_1(x_1)} = c_{1,2}(F_1(x_1), F_2(x_2)) \cdot f_2(x_2)$$

mentre $f_{3|1,2}(x_3|x_1, x_2)$ può essere riscritta osservando quanto segue:

$$\begin{aligned} f_{2,3|1}(x_2, x_3|x_1) &= \frac{f_{1,2,3}(x_1, x_2, x_3)}{f_1(x_1)} = \frac{f_1(x_1) \cdot f_{2|1}(x_2|x_1) \cdot f_{3|1,2}(x_3|x_1, x_2)}{f_1(x_1)} \\ &= f_{2|1}(x_2|x_1) \cdot f_{3|1,2}(x_3|x_1, x_2) \end{aligned}$$

da cui

$$f_{3|1,2}(x_3|x_1, x_2) = \frac{f_{2,3|1}(x_2, x_3|x_1)}{f_{2|1}(x_2|x_1)}$$

Ricorrendo alla relazione (3.1.3), è possibile scrivere $f_{2,3|1}(x_2, x_3|x_1)$ come segue:

$$f_{2,3|1}(x_2, x_3|x_1) = c_{2,3|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)|x_1) \cdot f_{2|1}(x_2|x_1) \cdot f_{3|1}(x_3|x_1)$$

Si noti che, nella relazione precedente, $c_{2,3|1}$ è la copula che descrive la dipendenza tra le v.a. \tilde{X}_2 e \tilde{X}_3 condizionatamente alla v.a. \tilde{X}_1 ; la variabile condizionante \tilde{X}_1 compare non solo negli argomenti della copula, dati dalle funzioni di ripartizione condizionate $F_{2|1}(x_2|x_1)$ e $F_{3|1}(x_3|x_1)$, ma anche direttamente all'interno della forma funzionale della copula, stante a significare il fatto che i parametri caratteristici della copula $c_{2,3|1}$ sono funzioni della v.a. condizionante \tilde{X}_1 .

La funzione di densità della v.a. \tilde{X}_3 condizionata rispetto alle variabili \tilde{X}_1 e \tilde{X}_2 , quindi, diviene:

$$\begin{aligned} f_{3|1,2}(x_3|x_1, x_2) &= \frac{c_{2,3|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)|x_1) \cdot f_{2|1}(x_2|x_1) \cdot f_{3|1}(x_3|x_1)}{f_{2|1}(x_2|x_1)} \\ &= c_{2,3|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)|x_1) \cdot f_{3|1}(x_3|x_1) \\ &= c_{2,3|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)|x_1) \cdot c_{3,1}(F_3(x_3), F_1(x_1)) \cdot f_3(x_3) \end{aligned}$$

dove nella relazione precedente vale $f_{3|1}(x_3|x_1) = c_{3,1}(F_3(x_3), F_1(x_1)) \cdot f_3(x_3)$. Infine, la densità $f_{1,2,3}(x_1, x_2, x_3)$ può essere scritta come:

$$\begin{aligned} f_{1,2,3}(x_1, x_2, x_3) &= f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot c_{1,2}(F_1(x_1), F_2(x_2)) \cdot c_{3,1}(F_3(x_3), F_1(x_1)) \\ &\quad \cdot c_{2,3|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)|x_1) \end{aligned} \tag{3.1.7}$$

Le copule bivariate presenti nella relazione precedente prendono il nome di pair-copule; tramite la pair-copula construction, quindi, è possibile scomporre in modo

ricorsivo la densità di una variabile aleatoria in n dimensioni. Si noti, inoltre, che la decomposizione precedente è solo una delle tre possibili fattorizzazioni in $n = 3$ dimensioni.

In termini generali, ogni elemento della (3.1.6) può essere espresso come il prodotto tra la relativa pair-copula e la funzione di densità marginale condizionata; seguendo la notazione in Aas et al. (2009) si ha,

$$f_{x_i|\mathbf{v}}(x_i|\mathbf{v}) = c_{x_i, v_j|\mathbf{v}_{-j}}(F_{x_i|\mathbf{v}_{-j}}(x_i|\mathbf{v}_{-j}), F_{v_j|\mathbf{v}_{-j}}(v_j|\mathbf{v}_{-j})|\mathbf{v}_{-j}) \cdot f_{x_i|\mathbf{v}_{-j}}(x_i|\mathbf{v}_{-j}) \quad (3.1.8)$$

dove,

- \mathbf{v} rappresenta il vettore (x_1, \dots, x_{i-1}) ;
- v_j individua un generico elemento del vettore \mathbf{v} ;
- \mathbf{v}_{-j} è il vettore \mathbf{v} al netto dell'elemento v_j .

Tale processo implica, oltre alle copule bivariate, anche le funzioni di ripartizione condizionate delle variabili aleatorie marginali, $F_{x_i|\mathbf{v}}(x_i|\mathbf{v})$. Queste ultime possono essere determinate come segue (Joe (1996)):

$$F_{x_i|\mathbf{v}}(x_i|\mathbf{v}) = \frac{\partial C_{x_i, v_j|\mathbf{v}_{-j}}(F_{x_i|\mathbf{v}_{-j}}(x_i|\mathbf{v}_{-j}), F_{v_j|\mathbf{v}_{-j}}(v_j|\mathbf{v}_{-j}))}{\partial F_{v_j|\mathbf{v}_{-j}}(v_j|\mathbf{v}_{-j})} \quad (3.1.9)$$

La fattorizzazione della densità multivariata tramite pair-copula construction è, quindi, per sua natura un processo ricorsivo. Le funzioni di densità condizionate che compaiono nella (3.1.6) possono essere scomposte tramite l'opportuna pair-copula moltiplicata per una funzione di densità condizionata; quest'ultima, a sua volta, può essere scomposta tramite pair-copula. Iterando in modo ricorsivo tale procedimento, è possibile scrivere la densità della variabile multidimensionale di partenza come il prodotto tra le funzioni di densità marginali e le rispettive copule bivariate; data una variabile in n dimensioni, le copule bivariate coinvolte sono pari a $\binom{n}{2}$. La scomposizione della variabile n -dimensionale, inoltre, può essere ottenuta tramite molteplici fattorizzazioni tra loro equivalenti da un punto di vista teorico.

La relazione (3.1.5) può, quindi, essere scritta come³ (Aas et al. (2009)):

$$f_{1, \dots, n}(x_1, \dots, x_n) = \prod_{k=1}^n f_k(x_k) \cdot \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{i, i+j|i+1, \dots, i+j-1}(F_{i|i+1, \dots, i+j-1}(x_i|x_{i+1}, \dots, x_{i+j-1}), F_{i+j|i+1, \dots, i+j-1}(x_{i+j}|x_{i+1}, \dots, x_{i+j-1})|(x_{i+1}, \dots, x_{i+j-1})), \quad (3.1.10)$$

³Come verrà mostrato nel seguito, la fattorizzazione in esame individua la densità di una particolare vine copula definita D-vine.

Esempio. Si consideri la variabile aleatoria $\widetilde{\mathbf{X}} = (\widetilde{X}_1, \widetilde{X}_2, \widetilde{X}_3, \widetilde{X}_4)$ fattorizzata tramite pair-copula construction come segue:

$$\begin{aligned} f_{1,2,3,4}(x_1, x_2, x_3, x_4) &= f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot f_4(x_4) \\ &\quad \cdot c_{1,2}(F_1(x_1), F_2(x_2)) \cdot c_{2,3}(F_2(x_2), F_3(x_3)) \cdot c_{3,4}(F_3(x_3), F_4(x_4)) \\ &\quad \cdot c_{1,3|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)|x_2) \cdot c_{2,4|3}(F_{2|3}(x_2|x_3), F_{4|3}(x_4|x_3)|x_3) \\ &\quad \cdot c_{1,4|2,3}(F_{1|2,3}(x_1|x_2, x_3), F_{4|2,3}(x_4|x_2, x_3)|x_2, x_3) \end{aligned}$$

Ricorrendo alla relazione (3.1.9), le funzioni di ripartizione condizionate delle v.a. marginali sono pari a:

$$\begin{aligned} F_{1|2}(x_1|x_2) &= \frac{\partial C_{1,2}(F_1(x_1), F_2(x_2))}{\partial F_2(x_2)} \\ F_{3|2}(x_3|x_2) &= \frac{\partial C_{3,2}(F_3(x_3), F_2(x_2))}{\partial F_2(x_2)} \\ F_{2|3}(x_2|x_3) &= \frac{\partial C_{2,3}(F_2(x_2), F_3(x_3))}{\partial F_3(x_3)} \\ F_{4|3}(x_4|x_3) &= \frac{\partial C_{4,3}(F_4(x_4), F_3(x_3))}{\partial F_3(x_3)} \\ F_{1|2,3}(x_1|x_2, x_3) &= \frac{\partial C_{1,3|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2))}{\partial F_{3|2}(x_3|x_2)} \\ F_{4|2,3}(x_4|x_2, x_3) &= \frac{\partial C_{4,3|2}(F_{4|3}(x_4|x_3), F_{2|3}(x_2|x_3))}{\partial F_{2|3}(x_2|x_3)} \end{aligned}$$

Si noti che $C_{2,3} = C_{3,2}$ e $C_{3,4} = C_{4,3}$ se le copule godono della proprietà dell'interscambiabilità. ■

La pair-copula construction scompone la densità della variabile n -dimensionale tramite copule bivariate; queste ultime, tuttavia, come già osservato in precedenza, comprendono anche copule condizionate. La variabile condizionante (o le variabili condizionanti nel caso in cui $n \geq 4$) compare non solo negli argomenti della copula, ma influenza anche la forma funzionale dei parametri da cui dipende. A titolo esemplificativo, riprendendo la scomposizione (3.1.7) della densità $f_{1,2,3}$, l'unica copula condizionata coinvolta risulta essere $c_{2,3|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)|x_1)$, dove la variabile condizionante è x_1 , da cui, in generale, dipenderà il parametro (o i parametri) della copula $c_{2,3|1}$. Al fine di rendere trattabile il problema inferenziale legato alla stima dei parametri delle copule condizionate, soprattutto al crescere della dimensione del problema, nella pratica si è soliti formulare la cosiddetta "ipotesi semplificata" della pair-copula construction. Tale assunzione prevede che le copule condizionate dipendano dalle variabili condizionanti *solo*

per mezzo degli argomenti stessi delle copule, definiti dalle funzioni di ripartizione condizionate. In tale circostanza la copula $c_{2,3|1}$ viene approssimata come segue:

$$c_{2,3|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)|x_1) \approx c_{2,3|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1))$$

e la scomposizione della densità $f_{1,2,3}$ risulta:

$$\begin{aligned} f_{1,2,3}(x_1, x_2, x_3) \approx & f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot c_{1,2}(F_1(x_1), F_2(x_2)) \cdot c_{3,1}(F_3(x_3), F_1(x_1)) \\ & \cdot c_{2,3|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)) \end{aligned} \quad (3.1.11)$$

In letteratura recenti lavori si sono focalizzati sull'ipotesi semplificata sottostante alla pair-copula construction. In Stöber, Joe e Czado (2013) si mostra che le uniche copule archimedee multivariate ($n > 2$) che soddisfano l'ipotesi semplificata sono quelle appartenenti alla famiglia Clayton. L'indagine è anche estesa alle copule ellittiche; in tale circostanza, le uniche copule multivariate che verificano l'ipotesi semplificata sono quelle associate alla distribuzione Normale e t-student.

In Mesfioui e Quessy (2008) si presentano dei risultati notevoli relativi alle copule archimedee condizionate. In particolare, considerando una generica copula archimedeica in tre dimensioni, $C(u, v, w)$, si dimostra che la copula condizionata $C(u, v|w)$ appartiene anch'essa alla famiglia di copule archimedee. Gli autori, inoltre, mostrano come la variabile condizionante possa essere interpretata come un parametro di dipendenza. In particolare, la struttura di maggiore dipendenza modellizzabile dalla copula condizionata viene raggiunta quando $w \rightarrow 1$; di contro, il caso opposto si ha quando $w \rightarrow 0$. In quest'ultima circostanza, inoltre, la copula condizionata appartiene alla famiglia Clayton di parametro $\alpha/(\alpha + 1)$, dove $\alpha > 0$ individua l'indice di variazione regolare della funzione inversa del generatore della copula in esame. Gli autori mostrano che le copule archimedee condizionate sono superiormente limitate non dal limite superiore di Fréchet, individuato dalla copula $M = \min(u, v)$, bensì dalla copula Ali-Mikhail-Haq di parametro $\theta = 1$:

$$M^*(u, v) = \frac{uv}{u + v - uv}$$

Come noto, il valore del τ di Kendall in tale circostanza è pari a $1/3$. Lo spettro di dipendenza modellizzabile dalle copule archimedee condizionate risulta essere, quindi, particolarmente contenuto. Anche il coefficiente di dipendenza di coda inferiore risulta essere più contenuto: il valore di λ_L per la copula condizionata, infatti, è pari alla metà del coefficiente di dipendenza di coda inferiore della rispettiva copula non condizionata. Il coefficiente di coda superiore, invece, per le copule archimedee condizionate è sempre pari a zero.

Esempio. In $n > 2$ dimensione, la copula Clayton può essere generalizzata come segue (Nelsen (2006, p. 152)):

$$C(u_1, u_2, \dots, u_n) = (u_1^{-\theta} + u_2^{-\theta} + \dots + u_n^{-\theta} + n - 1)^{-\frac{1}{\theta}} \quad \text{con} \quad \theta > 0$$

In Kraus e Czado (2016, Appendice B) si ricava analiticamente la forma funzionale della copula condizionata nel caso della famiglia Clayton in 3 dimensioni, mostrando come tale copula soddisfi l'ipotesi semplificata della pair-copula construction.

In termini generali, i passaggi da eseguire per ottenere la copula condizionata $C(u, v|w)$ e valutarne il rispetto dell'ipotesi semplificata sono i seguenti:

- calcolare le marginali bivariate $C(u, w)$ e $C(v, w)$;
- quantificare le funzioni di ripartizione delle variabili u e v condizionatamente a w , $F(u|w)$ e $F(v|w)$;
- calcolare le funzioni inverse $F^{-1}(u|w)$ e $F^{-1}(v|w)$;
- calcolare la funzione di ripartizione della distribuzione bivariata condizionata rispetto a w , $H(u, v|w)$;
- estrarre dalla distribuzione $H(u, v|w)$ la copula associata, $C(u, v|w)$: se il parametro (i parametri) di tale copula non dipendono dalla v.a. w segue che la copula $C(u, v|w)$ è nella forma semplificata.

In particolare, considerando $n = 3$, la copula Clayton diviene,

$$C(u, v, w) = (u^{-\theta} + v^{-\theta} + w^{-\theta} + 2)^{-\frac{1}{\theta}}$$

E' immediato osservare come le variabili marginali bidimensionali appartengano anch'esse alla famiglia Clayton, infatti:

$$C(u, w) = C(u, 1, w) = (u^{-\theta} + w^{-\theta} + 2)^{-\frac{1}{\theta}}$$

$$C(v, w) = C(1, v, w) = (v^{-\theta} + w^{-\theta} + 2)^{-\frac{1}{\theta}}$$

Le funzioni di ripartizione unidimensionali delle variabili u e v condizionate w sono rispettivamente:

$$F(u|w) = \frac{\partial C(u, w)}{\partial w} = (u^{-\theta} + w^{-\theta} + 2)^{-\frac{1}{\theta}-1} w^{-\theta-1}$$

$$F(v|w) = \frac{\partial C(v, w)}{\partial w} = (v^{-\theta} + w^{-\theta} + 2)^{-\frac{1}{\theta}-1} w^{-\theta-1}$$

Le funzioni inverse $F^{-1}(u|w)$ e $F^{-1}(v|w)$, inoltre, sono pari a:

$$F^{-1}(u|w) = \left[w^{-\theta} (u^{-\frac{\theta}{1+\theta}} - 1) + 1 \right]^{-\frac{1}{\theta}}$$

$$F^{-1}(v|w) = \left[w^{-\theta} (v^{-\frac{\theta}{1+\theta}} - 1) + 1 \right]^{-\frac{1}{\theta}}$$

Dopo avere quantificato le distribuzioni marginali è necessario valutare dalla copula trivariata $C(u, v, w)$ la distribuzione bivariata $H(u, v|w)$. Quest'ultima è ottenibile derivando rispetto alla variabile condizionante la copula C :

$$H(u, v|w) = \frac{\partial C(u, v, w)}{\partial w} = (u^{-\theta} + v^{-\theta} + w^{-\theta} + 2)^{-\frac{1}{\theta}-1} w^{-\theta-1}$$

Infine, la copula condizionata $C(u, v|w)$ associata alla distribuzione $H(u, v|w)$ può essere ottenuta ricorrendo alla relazione (3.1.2):

$$\begin{aligned} C(u, v|w) &= H(F^{-1}(u|w), F^{-1}(v|w)|w) \\ &= \left[w^{-\theta} (u^{-\frac{\theta}{1+\theta}} - 1) + 1 + w^{-\theta} (v^{-\frac{\theta}{1+\theta}} - 1) + 1 + w^{-\theta} - 2 \right]^{-\frac{1}{\theta}-1} w^{-\theta-1} \\ &= (u^{-\frac{\theta}{1+\theta}} + v^{-\frac{\theta}{1+\theta}} - 1)^{-\frac{1+\theta}{\theta}} \end{aligned}$$

La forma funzionale di $C(u, v|w)$, quindi, coincide con quella della copula bidimensionale Clayton di parametro $\eta = \theta/(\theta + 1)$; il parametro della copula condizionata $C(u, v|w)$ non dipende dalla variabile condizionante e, dunque, l'ipotesi semplificata risulta rispettata.

In Stöber, Joe e Czado (2013, Appendice C) si mostra come la copula Farlie-Gumbel-Morgenster non soddisfi l'ipotesi semplificata. In particolare, in 3 dimensioni tale famiglia di copule può essere estesa come segue,

$$C(u, v, w) = uvw[1 + \theta(1 - u)(1 - v)(1 - w)] \quad \text{con} \quad \theta \in [-1; 1]$$

Le marginali bidimensionali coincidono con la copula prodotto, infatti:

$$C(u, w) = C(u, 1, w) = uw$$

$$C(v, w) = C(1, v, w) = vw$$

Le funzioni di ripartizione unidimensionali delle variabili u e v condizionate a w sono,

$$F(u|w) = \frac{\partial C(u, w)}{\partial w} = u$$

$$F(v|w) = \frac{\partial C(v, w)}{\partial w} = v$$

le cui funzioni inverse coincidono rispettivamente a u e v , $F^{-1}(u|w) = u$ e $F^{-1}(v|w) = v$. La distribuzione bivariata di u e v condizionata alla variabile w risulta essere:

$$H(u, v|w) = \frac{\partial C(u, v, w)}{\partial w} = uv[1 + \theta(1 - u)(1 - v) - 2\theta w(1 - u)(1 - v)]$$

La copula associata alla distribuzione $H(u, v|w)$, infine, è pari a:

$$\begin{aligned} C(u, v|w) &= H(F^{-1}(u|w), F^{-1}(v|w)|w) \\ &= uv[1 + \theta(1 - u)(1 - v) - 2\theta w(1 - u)(1 - v)] \\ &= uv[1 + \theta(1 - 2w)(1 - u)(1 - v)] \end{aligned}$$

che risulta essere la copula bivariata Farlie-Gumbel-Morgenster di parametro $\eta = \theta(1 - 2w)$. L'ipotesi semplificata, quindi, non risulta soddisfatta in quanto il parametro della copula condizionata è funzione della variabile condizionante. ■

L'adeguatezza dell'ipotesi semplificata è stata investigata da ulteriori studi in letteratura. Nel lavoro di Haff, Aas e Frigessi (2010) sono presentati diversi esempi volti a mostrare casi particolari di distribuzioni multivariate in cui l'ipotesi semplificata è valida o meno. Anche in quest'ultima circostanza, gli autori concludono che adottare l'ipotesi semplificata risulta essere una buona approssimazione del vero modello. Nel lavoro di Acar, Genest e Nešlehová (2012) si mostra come l'analisi grafica delle marginali bivariate sottostanti ad una copula (distribuzione) multivariata non sia uno strumento adeguato per comprendere la validità o meno dell'ipotesi semplificata. Su tale tematica, al fine di indagare graficamente l'adeguatezza dell'ipotesi semplificata, Killiches, Kraus e Czado (2016) suggeriscono di considerare, nel contesto trivariato ($n = 3$), non le tre coppie di distribuzioni bivariate marginali, bensì le curve di livello (che in tale contesto sono delle superfici) della distribuzione in 3 dimensioni oggetto di studio. L'analisi sviluppata dagli autori permette quindi di comprendere qualitativamente da un punto di vista grafico quanto il data set di riferimento si discosti dall'ipotesi semplificata.

L'adeguatezza dell'ipotesi semplificata, inoltre, è indagata anche dei lavori di Killiches, Kraus e Czado (2018), Kurz e Spanhel (2017) e Spanhel e Kurz (2019). Infine, in Vatter e Chavez-Demoulin (2015) e Vatter e Nagler (2017) si presentano degli approcci volti a stimare le copule condizionate considerando l'effetto delle variabili condizionanti, interpretate come covariate del fenomeno e modellizzate tramite strutture additive.

3.2 Le vine copule

I *vine* sono uno strumento grafico che permette di descrivere la dipendenza tra n variabili aleatorie; l'introduzione in letteratura di tale metodologia si deve ai

lavori di Cooke (1997) e Bedford e Cooke (2001, 2002). I vine, sviluppati al fine di estendere lo strumento degli alberi Markoviani, basati sulla stringente ipotesi di indipendenza condizionata tra le variabili aleatorie, consentono di organizzare in modo grafico la dipendenza esistente tra le variabili marginali oggetto di studio. L'utilizzo congiunto delle copule bivariate e dei vine, infine, definisce le cosiddette vine copule, le quali permettono di definire le molteplici scomposizioni ottenibili tramite pair-copula construction.

I vine sono particolari strutture grafiche costituite da un insieme annidato di *alberi* connessi tra loro, dove un *albero* è definito come segue.

Definizione (Albero). Un albero $T = \{N, E\}$ è un grafo aciclico dove N rappresenta l'insieme dei nodi ed E l'insieme degli archi, definiti come coppie di nodi non orientati. \square

Un vine \mathcal{V} di n elementi è definito come un insieme di alberi contenuti l'uno nell'altro e collegati tra loro, $\mathcal{V} = \{T_1, \dots, T_{n-1}\}$, dove gli archi dell'albero j sono i nodi dell'albero $j + 1$, con $j = 1, \dots, n - 2$. Un *regular vine* di n variabili, invece, prevede che due archi nell'albero j siano connessi tra loro nell'albero $j + 1$ se e solo se tali archi possiedono un nodo in comune. Da un punto di vista formale i vine e i regular vine possono essere così definiti (Kurowicka e Cooke (2006)).

Definizione (Vine, Regular vine). \mathcal{V} è un vine di n elementi se:

1. $\mathcal{V} = \{T_1, \dots, T_{n-1}\}$.
2. T_1 è un albero con $N_1 = \{1, \dots, n\}$ nodi e un insieme di archi E_1 ; per $i = 2, \dots, n - 1$, T_i è un albero con $N_i = E_{i-1}$ nodi.

\mathcal{V} è un regular vine (R-vine) di n elementi se, oltre alle condizioni (1) e (2), vale la seguente condizione, chiamata di *prossimità*:

3. per $i = 2, \dots, n - 1$, se $\{a, b\} \in E_i$, dove $a = \{a_1, a_2\}$ e $b = \{b_1, b_2\}$, allora $\#a \cap b = 1$, dove $\#$ individua il simbolo di cardinalità di un insieme.

\square

Un R-vine di n elementi, quindi, è un insieme ordinato di $n - 1$ alberi, dove gli archi dell'albero j diventano i nodi dell'albero $j + 1$. La condizione di prossimità, inoltre, garantisce che due archi nell'albero $j + 1$ siano connessi tra loro solo se nell'albero j hanno un nodo in comune. In un regular vine \mathcal{V} di n elementi il numero totale di archi è pari a $\binom{n}{2}$.

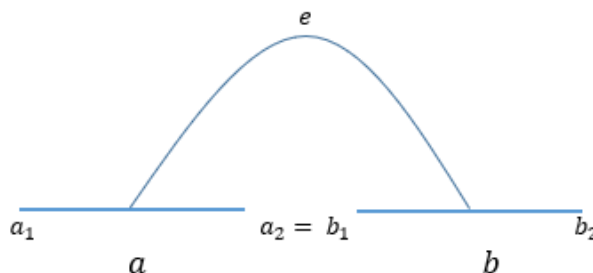


Figura 3.1: Condizione di prossimità. Gli archi a e b condividono il nodo $a_2 = b_1$; risultano, quindi, collegati tra loro dall'arco e che sarà un nodo nell'albero successivo.

Per ogni arco dell'albero è possibile definire l'insieme dei *vincoli*, l'insieme *condizionante* e l'insieme *condizionato*. Gli elementi ottenibili da un dato arco sono definiti come l'insieme dei vincoli dell'arco. Inoltre, quando due archi sono uniti tra loro da un arco nell'albero successivo, l'intersezione tra gli insiemi dei vincoli dei due archi definisce l'insieme condizionante; infine, la differenza simmetrica degli insiemi dei vincoli definisce l'insieme condizionato.

Definizione (Insiemi dei vincoli, condizionante, condizionato). Per ogni generico arco e del vine l'insieme dei vincoli, quello condizionante e quello condizionato sono definiti, adottando la notazione utilizzata in letteratura, come segue.

1. L'insieme dei vincoli associato a $e \in E_i$ è U_e^* .
2. Per $i = 1, \dots, n - 1$, $e \in E_i$ con $e = \{j, k\}$, l'insieme condizionante associato ad e è

$$D_e = U_j^* \cap U_k^*$$

3. Per $i = 1, \dots, n - 1$, $e \in E_i$ con $e = \{j, k\}$, l'insieme condizionato associato ad e è

$$\{C_{e,j}, C_{e,k}\} = U_j^* \Delta U_k^*$$

□

Si noti che per $e \in E_1$, l'insieme condizionante è vuoto.

La condizione di prossimità, che definisce i regular vine, garantisce che l'insieme condizionato sia sempre composto da due elementi. Inoltre, ogni coppia di elementi costituenti l'insieme condizionato compare, all'interno dei nodi, una sola volta nella struttura vine.

Esempio. Si consideri il seguente regular vine (D-vine) costituito da 5 elementi e , quindi, definito da 4 alberi, $\mathcal{V} = \{T_1, T_2, T_3, T_4\}$:

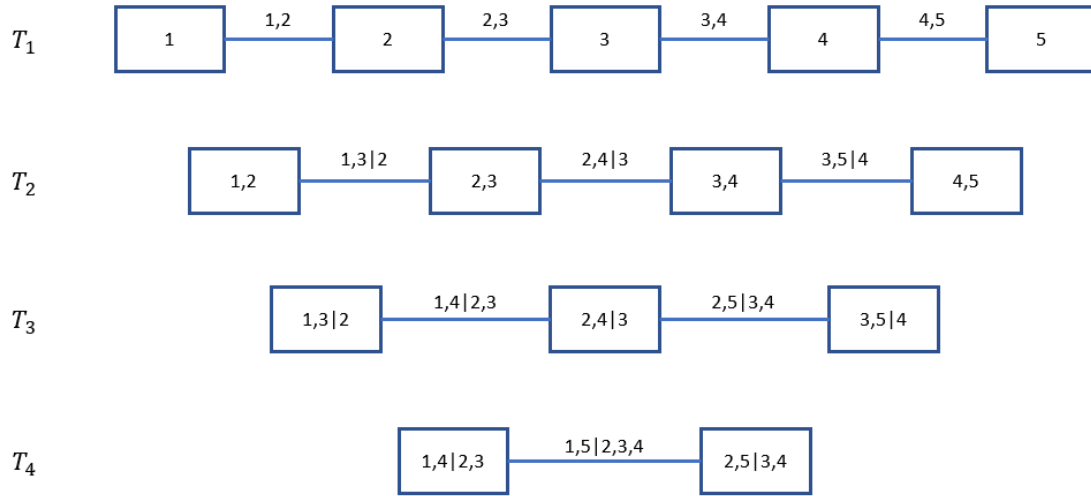


Figura 3.2: Regual vine nella forma di D-vine composto da 5 elementi.

Il primo albero, T_1 , è costituito dai nodi N_1 e dagli archi E_1 seguenti:

- $N_1 = \{1, 2, 3, 4, 5\}$
- $E_1 = \{ \{1, 2\}; \{2, 3\}; \{3, 4\}; \{4, 5\} \}$.

L'insieme dei vincoli per i 4 archi è rispettivamente $\{1, 2\}$, $\{2, 3\}$, $\{3, 4\}$ e $\{4, 5\}$. Nel primo albero gli insiemi condizionanti sono vuoti, mentre gli insiemi condizionati coincidono con quelli dei vincoli.

Il secondo albero, T_2 , possiede nodi, N_2 , coincidenti con gli archi dell'albero precedente e archi E_2 rispettivamente pari a:

- $N_2 = E_1$
- $E_2 = \{ \{ \{1, 2\}, \{2, 3\} \}; \{ \{2, 3\}, \{3, 4\} \}; \{ \{3, 4\}, \{4, 5\} \} \}$.

L'insieme dei vincoli per i tre archi è dato rispettivamente da $\{ \{1, 2\}, \{2, 3\} \}$, $\{ \{2, 3\}, \{3, 4\} \}$ e $\{ \{3, 4\}, \{4, 5\} \}$.

Gli insiemi condizionanti sono definiti dall'intersezione degli insiemi dei vincoli degli archi collegati tra loro; per l'arco $\{ \{1, 2\}, \{2, 3\} \}$ è pari a $\{1, 2\} \cap \{2, 3\} = \{2\}$, per l'arco $\{ \{2, 3\}, \{3, 4\} \}$ è pari a $\{2, 3\} \cap \{3, 4\} = \{3\}$, mentre per l'arco $\{ \{3, 4\}, \{4, 5\} \}$ è $\{3, 4\} \cap \{4, 5\} = \{4\}$.

Gli insiemi condizionati, invece, sono definiti come la differenza simmetrica tra gli insiemi dei vincoli; per ogni arco, quindi, gli insiemi condizionati sono $\{1, 2\} \Delta \{2, 3\} = \{1, 3\}$, $\{2, 3\} \Delta \{3, 4\} = \{2, 4\}$ e $\{3, 4\} \Delta \{4, 5\} = \{3, 5\}$.

Il terzo albero, T_3 , è caratterizzato dai seguenti nodi e archi:

- $N_3 = E_2$
- $E_3 = \{ \{ \{1, 3|2\}, \{2, 4|3\} \}; \{ \{2, 4|3\}, \{3, 5|4\} \} \}$.

Gli insiemi dei vincoli dei due archi sono rispettivamente $\{\{1, 3, 2\}, \{2, 4, 3\}\}$ e $\{\{2, 4, 3\}, \{3, 5, 4\}\}$.

L'insieme condizionante, invece, è per l'arco $\{\{1, 3|2\}, \{2, 4|3\}\}$ pari a $\{\{1, 3, 2\} \cap \{2, 4, 3\}\} = \{2, 3\}$, mentre per l'arco $\{\{2, 4|3\}, \{3, 5|4\}\}$ è pari a $\{\{2, 4, 3\} \cap \{3, 5, 4\}\} = \{3, 4\}$.

Gli insiemi condizionati, invece, sono dati da $\{1, 3, 2\} \triangle \{2, 4, 3\} = \{1, 4\}$ e $\{2, 4, 3\} \triangle \{3, 5, 4\} = \{2, 5\}$.

Il quarto albero, T_4 , infine, è caratterizzato dai seguenti nodi e archi:

- $N_4 = E_3$
- $E_4 = \{\{1, 4|2, 3\}, \{2, 5|3, 4\}\}$.

L'insieme dei vincoli dell'unico arco è pari a $\{\{1, 4, 2, 3\}, \{2, 5, 3, 4\}\}$.

L'insieme condizionate, invece, è definito come $\{1, 4, 2, 3\} \cap \{2, 5, 3, 4\} = \{2, 3, 4\}$, mentre l'insieme condizionato è pari a $\{1, 4, 2, 3\} \triangle \{2, 5, 3, 4\} = \{1, 5\}$. ■

In Bedford e Cooke (2001) si mostra come la densità di una variabile aleatoria multivariata in n dimensioni possa essere rappresentata tramite vine. In particolare, le vine copule sono definite associando ad ogni arco del vine una copula condizionata in cui le v.a. sono definite dall'insieme condizionato, mentre le variabili condizionanti sono definite dall'insieme condizionante dell'arco. Nel primo albero sono coinvolte solo copule non condizionate in numero pari a $n - 1$. In particolare, la struttura dei regular vine garantisce che tutte le copule coinvolte siano bivariate; in tale fattispecie, è possibile fattorizzare la variabile aleatoria multivariata tramite pair-copula construction, definendo una *vine copula bivariata*, dove l'aggettivo qualificativo *bivariato* indica che la copula multivariata è fattorizzata ricorrendo a sole copule bidimensionali. In tale situazione, quindi, le copule bivariate coinvolte sono pari a $\binom{n}{2}$, di cui $n - 1$ non condizionate e le restanti $(n - 1)(n - 2)/2$ condizionate. In letteratura l'attenzione è stata posta quasi esclusivamente sulle vine copule bivariate, tanto da riferirsi a queste ultime semplicemente con il termine di vine copule (omettendo l'aggettivo "bivariato"). Nel seguito, se non diversamente specificato, con il termine vine copula si farà riferimento alle sole vine copule bivariate, basate, quindi, sulla PCC.

Le vine copule, quindi, sono copule multivariate costruite tramite la fattorizzazione determinata dalla pair-copula construction, le cui molteplici scomposizioni sono organizzate e definite ricorrendo alla struttura grafica definita dai regular vine. In Bedford e Cooke (2001) si dà evidenza di come scrivere la densità di una v.a. n -dimensionale tramite regular vine. Si consideri la v.a. $\tilde{\mathbf{X}} = (\tilde{X}_1, \dots, \tilde{X}_n)$ n -dimensionale dotata di variabili marginali con funzioni di ripartizione e densità rispettivamente $F_i(x_i)$ e $f_i(x_i)$, con $i = 1, \dots, n$. Dato un regular vine di n elementi (variabili aleatorie), $\mathcal{V} = \{T_1, \dots, T_{n-1}\}$, per ogni arco $e = \{j, k\} \in T_i$, per $i = 1, \dots, n - 1$, con insieme condizionato $\{j, k\}$ e insieme condizionante D_e , siano le funzioni di ripartizione e densità della copula condizionata rispettivamente $C_{i,j|D_e}$ e $c_{i,j|D_e}$. La funzione di densità della v.a. $\tilde{\mathbf{X}}$ può essere scritta come:

$$f_{1,\dots,n}(x_1, \dots, x_n) = \prod_{i=1}^n f_i(x_i) \cdot \prod_{e_{j,k} \in E_i} c_{j,k|D_e}(F_{j|D_e}(x_j|D_e), F_{k|D_e}(x_k|D_e)|D_e) \quad (3.2.1)$$

Risulta evidente l'equivalenza tra la relazione precedente e la relazione (3.1.10); le molteplici scomposizioni della densità multivariata ottenibili tramite PCC, quindi, possono essere organizzate tramite regular vine.

Si noti che la formula (3.2.1) definisce la densità di una di una v.a. multidimensionale dotata di generiche v.a. marginali $f_i(x_i)$; in quanto tale, quindi, dovrebbe essere denominata più propriamente *vine distribution* e non *vine copula*. L'utilizzo in letteratura di quest'ultimo termine, tuttavia, è risultato preponderante in quanto l'obiettivo di tale costrutto metodologico risulta quello di modellizzare la struttura di dipendenza tra le v.a. marginali; coerentemente con la teoria delle copule, infatti, l'analisi del modello sottostante alle distribuzioni marginali può essere implementata disgiuntamente rispetto allo studio del fenomeno della dipendenza. Nel proseguo, senza ulteriori distinzioni, si farà uso del termine *vine copula*.

In letteratura le vine copule maggiormente studiate e utilizzate sono quelle associate alle strutture dei regular vine definite C-vine (*canonical vine*) e D-vine (*drawable vine*).

Definizione (C-vine, D-vine). Un regular vine $\mathcal{V} = \{T_1, \dots, T_{n-1}\}$ è chiamato,

1. C-vine se ogni albero T_i ha un unico nodo di grado $n - i$;
2. D-vine se ogni nodo in T_1 ha al più grado pari a 2, dove il grado è definito come il numero di archi connessi al nodo.

□

Il numero di regular vine cresce in modo notevole con la dimensione del problema; in Morales-Nápoles, Cooke e Kurowicka (2008) si fornisce una formula analitica al fine di enumerare tutti i regular vine possibili in n dimensioni. Considerando n v.a., gli autori dimostrano che le configurazioni differenti di regular vine sono in numero pari a:

$$n!/2 \times 2^{\binom{n-2}{2}}$$

Anche circoscrivendo l'analisi ai soli C-vine e D-vine e considerando solo copule che godono della proprietà dell'interscambiabilità il numero di strutture vine risulta essere considerevole, pari a $n!/2$.

Si noti che per $n = 3$ tutti i regular vine appartengono alla stessa classe, per $n = 4$, invece, possono dare luogo sia a C-vine che a D-vine, infine, per $n \geq 5$ compaiono regular vine che non sono né C-vine né D-vine. Questi ultimi, quindi, costituiscono delle strutture limite all'interno della classe dei regular vine, permettendo di semplificare il problema inferenziale relativo alla scelta della tipologia

di struttura vine da utilizzare.

La tabella seguente riporta il numero di regular vine, C-vine e D-vine che è possibile ottenere al variare della dimensione del problema. I grafici riportati nel seguito, invece, propongono degli esempi di strutture R-vine, C-vine e D-vine composte da 5 elementi.

n	R-vine	C-vine, D-vine
3	3	3
4	24	12
5	48	60
6	23.040	360
7	2.580.480	2.520
8	6,606 e+08	20.160
9	3,805 e+11	181.440
10	4,871 e+14	1.814.400

Tabella 3.1: Numero di R-vine, C-vine e D-vine ottenibili in n dimensioni.

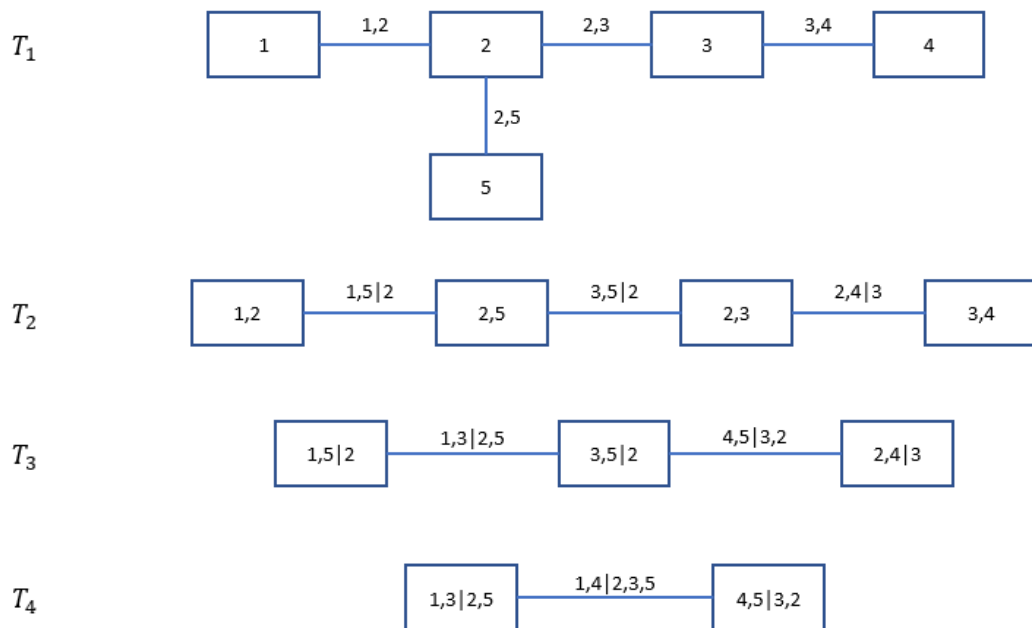


Figura 3.3: Esempio di regular vine composto da 5 elementi.

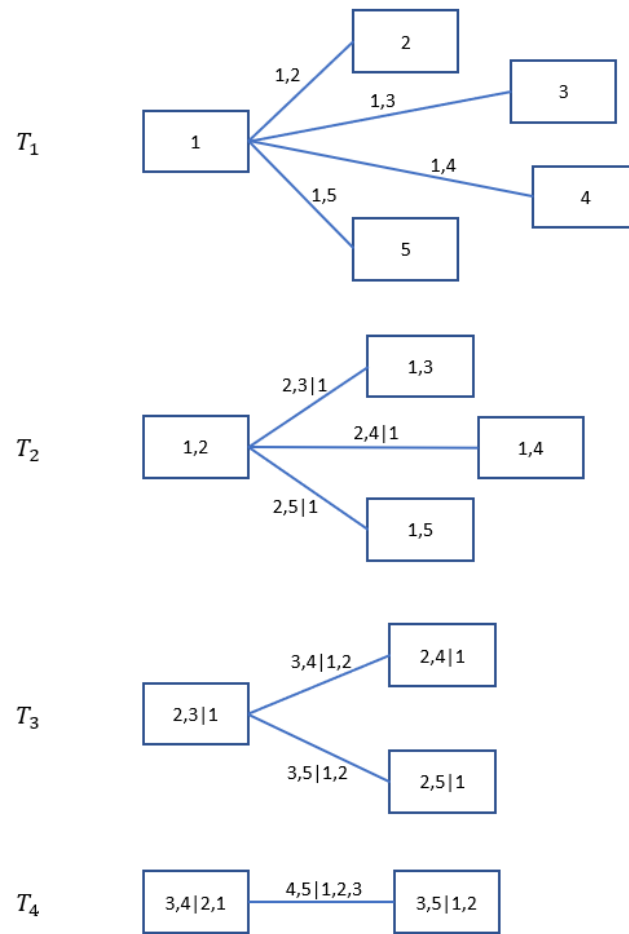


Figura 3.4: Esempio di C-vine composto da 5 elementi.

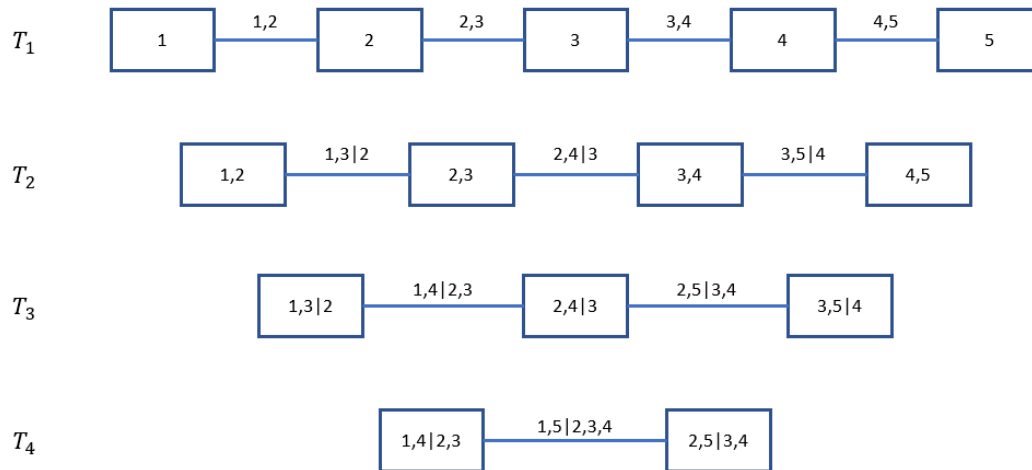


Figura 3.5: Esempio di D-vine composto da 5 elementi.

La funzione di densità di una v.a. n -dimensionale rappresentata tramite D-vine e C-vine risulta la seguente (Aas et al. (2009)):

$$f_{1,\dots,n}(x_1, \dots, x_n) = \prod_{k=1}^n f_k(x_k) \cdot \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{i,i+j|i+1,\dots,i+j-1}(F_{i|i+1,\dots,i+j-1}(x_i|x_{i+1}, \dots, x_{i+j-1}), F_{i+j|i+1,\dots,i+j-1}(x_{i+j}|x_{i+1}, \dots, x_{i+j-1})|(x_{i+1}, \dots, x_{i+j-1})), \quad (3.2.2)$$

$$f_{1,\dots,n}(x_1, \dots, x_n) = \prod_{k=1}^n f_k(x_k) \cdot \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{j,i+j|1,\dots,j-1}(F_{j|1,\dots,j-1}(x_j|x_1, \dots, x_{j-1}), F_{i+j|1,\dots,j-1}(x_{i+j}|x_1, \dots, x_{j-1})|(x_1, \dots, x_{j-1})), \quad (3.2.3)$$

Si noti che in un D-vine, l'ordine delle variabili nel primo albero determina univocamente tutti gli alberi successivi. Considerando un C-vine, invece, in ogni albero è necessario scegliere il nodo di grado massimo.

Per implementare da un punto di vista pratico modelli stocastici risulta necessario potere generare delle realizzazioni casuali dal modello di C-vine o D-vine oggetto di analisi. Un algoritmo di carattere generale, analogo a quanto visto per le copule bivariate con il *metodo della distribuzione condizionata*, risulta essere quello descritto nel seguito.

Considerando il caso n -dimensionale e assumendo che le variabili marginali siano uniformi, i passaggi da eseguire sono (Aas et al. (2009)):

- 1) generare w_i , con $i = 1, \dots, n$, v.a. indipendenti e distribuite uniformemente in $[0, 1]$;
- 2) porre:

$$\begin{aligned} x_1 &= w_1 \\ x_2 &= F_{2|1}^{-1}(w_2|x_1) \\ x_3 &= F_{3|1,2}^{-1}(w_3|x_1, x_2) \\ &\dots = \dots \\ x_n &= F_{n|1,\dots,n-1}^{-1}(w_n|x_1, \dots, x_{n-1}) \end{aligned}$$

dove le funzioni $F_{j|1,\dots,j-1}(x_j|x_1, \dots, x_{j-1})$, per ogni j , sono determinate ricorsivamente tramite la relazione (3.1.9).

3.2.1 Il processo di stima delle vine copule

Dato un campione di osservazioni generate da una v.a. n -dimensionale, nel contesto in esame il focus della stima risulta essere la parte del modello vine associato

alle sole copule, lasciando ad altra sede il problema della modellizzazione delle v.a. marginali. Nelle relazioni (3.2.1), (3.2.2) e (3.2.3), quindi, si assume implicitamente che le variabili marginali siano delle variabili uniformi in $[0, 1]$; la parte del modello associata alle funzioni di densità delle marginali, quindi, svanisce in quanto pari all'unità, $\prod_{k=1}^n f_k(x_k) = 1$. In tale circostanza, la vine distribution coincide effettivamente con una vine copula.

Da un punto di vista teorico, le molteplici configurazioni ottenibili tramite PCC con cui è possibile fattorizzare la funzione di densità di una variabile multidimensionale sono tra loro equivalenti; scomporre la v.a. $f_{1,\dots,n}(x_1, \dots, x_n)$ tramite C-vine, D-vine, o altro regular vine, quindi, risulta essere indifferente. Da un punto di vista inferenziale, di contro, l'utilizzo dell'ipotesi semplificata associata alla PCC, l'uso solamente di alcune famiglie bivariate di copule per stimare la struttura di dipendenza tra le variabili coinvolte, nonché la procedura di stima dei parametri associati alle copule creano l'esigenza di selezionare la vine copula più adeguata a descrivere la struttura di dipendenza insita nei dati oggetto di studio. In termini pratici, quindi, l'utilizzo di differenti strutture di regular vine porta a risultati diversi in termini di adattamento della vine copula selezionata rispetto ai dati osservati.

Il problema inferenziale legato alle vine copule, basato sull'ipotesi semplificata della PCC, si compone di tre parti (Dißmann et al. (2013)):

1. scelta del modello regular vine (C-vine o D-vine) e dell'ordine delle variabili;
2. selezione delle famiglie di copule bivariate da adottare per ogni arco del vine;
3. stima dei parametri caratteristici per ogni famiglia di copula selezionata.

In termini inferenziali, il primo passaggio necessario da valutare al fine di stimare la vine copula più adeguata ai dati è quello della scelta della struttura di albero, nonché dell'ordinamento delle variabili da utilizzare. Le diverse fattorizzazioni della PCC, infatti, non sono equivalenti tra loro in quanto la scelta di diverse tipologie di copule, la stima dei parametri associati, nonché l'aderenza o meno all'ipotesi semplificata portano, come detto, i diversi modelli vine a differenti livelli di adattamento rispetto ai dati. In linea teorica, al fine di selezionare la struttura vine più adeguata è necessario stimare tutte le possibili strutture esistenti in n dimensioni. Come mostrato in tabella 3.1, tuttavia, al crescere della dimensione del problema il numero di regular vine cresce enormemente. L'esigenza pratica di selezionare la struttura vine più adeguata senza enumerarne tutte le possibili configurazioni esistenti ha trovato risposta in letteratura nello sviluppo di specifici algoritmi.

In Dißmann et al. (2013) si propone un algoritmo euristico che consiste nella selezione degli alberi di un regular vine massimizzando la dipendenza tra le coppie di variabili aleatorie. L'obiettivo dell'algoritmo consiste, nello specifico, nel cogliere il maggior grado di dipendenza nei primi alberi i quali hanno la maggiore

influenza sul livello di adattamento complessivo del vine. In particolare, il primo albero viene definito selezionando la configurazione di v.a. che massimizza il valore assoluto del τ di Kendall empirico. Ad ogni arco congiungente due v.a., quindi, viene associato un peso pari al valore assoluto del τ di Kendall osservato tra le variabili stesse; il criterio di scelta prevede di selezionare l'albero che massimizza la somma dei pesi. In modo analogo tutti gli alberi successivi sono definiti selezionando l'ordine delle variabili aleatorie che massimizzano la dipendenza tra le coppie di variabili coinvolte. Si noti che, considerando un D-vine, l'algoritmo di *Dißmann* consiste nel scegliere l'ordine delle variabili aleatorie del solo primo albero; tutti quelli successivi, infatti, sono univocamente determinati dalla configurazione dell'albero T_1 . Di contro, in un C-vine l'algoritmo permette di scegliere la configurazioni delle coppie di variabili casuali per ogni albero del vine.

In Czado, Jeske e Hofmann (2013) vengono proposti degli algoritmi di scelta del vine alternativi a quello precedente; lo scopo degli autori è quello di selezionare gli alberi della struttura vine ricorrendo a grandezze statistiche di adattamento del modello. In particolare, gli algoritmi prevedono di selezionare gli alberi che massimizzano per ogni arco dei pesi, definiti dalle grandezze AIC o dai valori del p-value di opportuni test statistici di adattamento relativi alle copule bivariate stimate. Si noti che tali algoritmi non velocizzano il problema di selezione del vine in quanto richiedono di stimare in ogni albero tutte le pair-copule coinvolte al fine di valutare l'albero che massimizza il criterio di scelta previsto dal metodo.

In Kraus e Czado (2017), infine, è proposto un approccio di stima della struttura del regular vine che si prefigge un duplice obiettivo; costruire una vine copula che, analogamente all'algoritmo di *Dißmann*, colga il più possibile la dipendenza tra le v.a. negli alberi iniziali e, inoltre, che sia aderente il più possibile all'ipotesi semplificata sottostante alla PCC. Quest'ultima è valutata ricorrendo al test statistico, richiamato nel seguito, definito *constant conditional correlations test* (CCC), sviluppato in Kurz e Spanhel (2017).

Tale test permette di valutare il grado di adeguatezza dell'ipotesi semplificata per ogni pair-copula coinvolta nel vine. Considerando una generica pair-copula della relazione (3.2.1), l'ipotesi semplificata prevede che le variabili aleatorie $\tilde{U}_{j|D_e} = F_{j|D_e}(x_j|D_e)$ e $\tilde{U}_{k|D_e} = F_{k|D_e}(x_k|D_e)$ siano indipendenti dalle variabili condizionanti definite dall'insieme D_e , $(\tilde{U}_{j|D_e}, \tilde{U}_{k|D_e}) \perp \tilde{U}_{D_e}$. Il test è costruito considerando il coefficiente di correlazione lineare condizionato, $\rho_{j,k|D_e}$; in particolare, si verifica l'ipotesi che la grandezza $\rho_{j,k|D_e}$ associata alla pair-copula $C_{j,k|D_e}$ sia costante rispetto alle variabili condizionanti \tilde{U}_{D_e} . Il test prevede di dividere il supporto delle v.a. condizionanti Ω_0 secondo la seguente partizione $\Gamma = \{\Omega_1, \dots, \Omega_L\}$, con $L \in \mathbb{N}$, $\Omega_i \cap \Omega_j = \emptyset$ per $i \neq j$ e $P(\tilde{U}_{D_e} \in \Omega_i) > 0$ per $i = 1, \dots, L$. L'ipotesi nulla assume che la correlazione condizionata tra le variabili non dipenda dalla partizione delle variabili condizionanti, $\rho_{\omega_1} = \dots = \rho_{\omega_L}$, dove $\rho_{\omega_i} = \rho(\tilde{U}_{j|D_e}, \tilde{U}_{k|D_e} | \tilde{U}_{D_e} \in \Omega_i)$. Gli autori costruiscono una statistica test che tiene in considerazione sia la dimensione delle v.a. condizionanti che le diverse configurazioni con cui è possibile partizionarne il supporto.

L'algoritmo proposto da Kraus e Czado (2017) seleziona la struttura del regular

vine, quindi, come segue. Il primo albero T_1 , in cui non sono presenti copule condizionate, è scelto ricorrendo all'algoritmo di *Dißmann*. Nei livelli successivi, invece, vengono scelti gli alberi che massimizzano i pesi associati agli archi congiungenti le v.a.; i pesi, in tale algoritmo, tuttavia sono funzione sia del τ di Kendall tra le variabili che del valore di p-value associato al test di ipotesi CCC. Si noti che nel caso specifico di un D-vine, l'algoritmo in questione coincide con quello di *Dißmann* in quanto la configurazione del primo albero determina univocamente tutti gli alberi successivi. Gli autori, ricorrendo a un vasto studio simulativo mostrano come l'algoritmo, in genere, porti a selezionare delle strutture di vine copule con un adattamento migliore rispetto all'algoritmo proposto da Dißmann et al. (2013); tale risultato, inoltre, è accentuato al crescere della dimensione del problema. In dimensioni contenute, infatti, l'ordinamento delle variabili nel primo albero, che è determinato sulla base del solo criterio di massimizzazione della dipendenza, tende ad avere un impatto preponderante sul livello complessivo di adattamento del modello vine.

Per scegliere tra due o più modelli stimati di vine copule è possibile ricorrere ai criteri di informazione di Akaike (AIC) e Bayesiano (BIC). Tali grandezze permettono di individuare il modello più parsimonioso e che meglio si adatta ai dati. I valori AIC e BIC di un modello sono definiti come:

$$AIC = 2k - 2 \ln(\hat{L})$$

e

$$BIC = \ln(H)k - 2 \ln(\hat{L})$$

dove k rappresenta il numero di parametri stimati nel modello, \hat{L} è il valore di massima verosimiglianza del modello, H , infine, è la dimensione campionaria.

Anche i test statistici sviluppati da Vuong (1989) e Clarke (2007) permettono di scegliere tra due modelli concorrenti tra loro quello più adeguato.

La statistica test di Vuong è determinata a partire dal rapporto delle verosimiglianze dei due modelli oggetto di confronto. In particolare, si consideri il campione di osservazioni $\mathbf{x} = (x_{1,h}, \dots, x_{n,h})$, con $h = 1, \dots, H$, generate dalla v.a. $\tilde{\mathbf{X}}$ n -dimensionale e si considerino due modelli vine n -dimensionali dotati rispettivamente di densità $f_1(\hat{\boldsymbol{\theta}}_1)$ e $f_2(\hat{\boldsymbol{\theta}}_2)$, in cui $\hat{\boldsymbol{\theta}}_1$ e $\hat{\boldsymbol{\theta}}_2$ rappresentano i parametri stimati. La statistica test è definita come:

$$v = \frac{\frac{1}{H} \sum_{h=1}^H m_h}{\sqrt{\sum_{h=1}^H (m_h - \bar{m})^2}}$$

dove

$$m_h = \ln \left[\frac{f_1(u_h; \hat{\boldsymbol{\theta}}_1)}{f_2(u_h; \hat{\boldsymbol{\theta}}_2)} \right] \quad \text{con} \quad h = 1, \dots, H$$

e \bar{m} rappresenta il valore atteso dei termini m_h . La statistica v , quindi, è la somma standardizzata delle differenze logaritmiche dei termini di verosimiglianza;

tale grandezza, inoltre, è asintoticamente normale, $v \sim N(0, 1)$.

L'ipotesi nulla del test prevede che i due modelli siano equivalenti tra loro, ossia:

$$H_0 = E(m_h) = 0 \quad \forall h = 1, \dots, H$$

Il modello $f_1(\hat{\theta}_1)$ sarà preferito rispetto al secondo modello di vine copula, con un livello di confidenza α , nel caso in cui il valore assunto dalla statistica test sia superiore a $\Phi^{-1}(1 - \alpha/2)$, dove Φ^{-1} rappresenta l'inverso della funzione di ripartizione della v.a. Normale standard, o analogamente per valori del p-value minori di $\alpha/2$. Di contro, il modello $f_2(\hat{\theta}_2)$ risulta statisticamente più adeguato se $v < -\Phi^{-1}(1 - \alpha/2)$, dunque per valori del p-value maggiori di $1 - \alpha/2$. Infine, per valori della statistica test compresi tra $[-\Phi^{-1}(1 - \alpha/2), \Phi^{-1}(1 - \alpha/2)]$ non è possibile discriminare tra i due modelli. I valori soglia di tale test, inoltre, possono essere corretti per tenere conto della parsimonia, intesa come numero di parametri, dei modelli considerati; in particolare, tale correzione si basa sulle grandezze AIC e BIC.

Un test in parte analogo è stato proposto da Clarke in cui l'ipotesi nulla oggetto di valutazione risulta essere:

$$H_0 = P(m_h > 0) = 0,5 \quad \forall h = 1, \dots, H$$

Sotto l'ipotesi nulla di equivalenza dei due modelli i valori del rapporto di verosimiglianza tra i due modelli sono distribuiti in modo uniforme intorno a zero; statisticamente, quindi, metà dei valori m_h saranno maggiori di zero. Ponendo,

$$d_h = \ln(f_1(u_h; \hat{\theta}_1)) - \ln(f_2(u_h; \hat{\theta}_2))$$

la statistica test definita dall'autore risulta essere:

$$B = \sum_{h=1}^H I_{(0, \infty)}(d_h)$$

dove $I(\cdot)$ è la funzione indicatrice, definita come:

$$I(A) = \begin{cases} 1, & \text{se } A \text{ è vera} \\ 0, & \text{altrimenti} \end{cases}$$

La statistica B è distribuita come una variabile aleatoria Binomiale di parametri H e $p = 1/2$, $Bin(H, 1/2)$. Il modello $f_1(\hat{\theta}_1)$ è statisticamente migliore rispetto al secondo modello, al livello di confidenza α , se la statistica test assume un valore maggiore della soglia definita dalla funzione quantilica della distribuzione $Bin(H, 1/2)$ calcolata nel punto $1 - \alpha/2$, ossia per valori del p-value inferiori a $\alpha/2$. Di contro, il modello $f_2(\hat{\theta}_2)$ sarà preferito nel caso in cui la statistica test assuma un valore inferiore alla soglia definita dalla funzione quantilica della distribuzione $Bin(H, 1/2)$ calcolata nel punto $\alpha/2$, ossia per valori del p-value maggiori di $1 - \alpha/2$. Infine, per valori del p-value compresi nell'intervallo $[\alpha/2, 1 - \alpha/2]$ i

due modelli risultano statisticamente equivalenti tra loro. Anche per tale test, infine, è possibile correggere, ricorrendo alle grandezze AIC o BIC, i valori del p-value per tenere in debita considerazione la parsimonia dei modelli oggetto di confronto.

Dopo avere scelto la struttura vine più idonea, è necessario selezionare un'adeguata famiglia di copule bivariate per ogni arco degli alberi associati al vine. Nel contesto delle copule, la selezione del modello viene perseguita ricorrendo sia a strumenti grafici che a test statistici.

Dato un campione bivariato di H osservazioni, è possibile calcolare la funzione di ripartizione empirica di ogni variabile marginale i -esima come segue:

$$u_{i,h} = F_{i,h}(x_{i,h}) = \frac{1}{H+1} \sum_{s=1}^H I(X_{i,s} \leq x_{i,h})$$

dove la divisione per $H+1$ al posto che H viene adottata per evitare problemi al limite. Nel contesto delle copule, i valori assunti dalla funzione di ripartizione empirica di una v.a. marginale sono anche denominati *pseudo-osservazioni* (si veda ad esempio Genest e Favre (2007) e Haff (2012)). Tali grandezze sono per costruzione invarianti rispetto a trasformazioni monotone crescenti dei dati in quanto dipendono esclusivamente dal rango delle osservazioni, infatti: $u_{i,h} = F_{i,h}(x_{i,h}) = \frac{1}{H+1} \sum_{s=1}^H I(X_{i,s} \leq x_{i,h}) = \frac{1}{H+1} \sum_{s=1}^H I(r_{i,s} \leq r_{i,h})$, dove $r_{i,s}$ individua il rango dell'osservazione $x_{i,s}$.

Considerando, quindi, le realizzazioni casuali di una variabile bidimensionale, la rappresentazione grafica delle pseudo-osservazioni, definite per costruzione nel piano $[0, 1] \times [0, 1]$ ed invarianti rispetto a trasformazioni monotone crescenti delle variabili, fornisce indicazioni di carattere qualitativo sulla struttura di dipendenza sottostante ai dati.

Un ulteriore strumento grafico utile per indagare la dipendenza sottostante ai dati risulta essere il cosiddetto grafico *Chi*, originariamente introdotto da Fisher e Switzer (1985). Date le pseudo-osservazioni $u_{1,h}$ e $u_{2,h}$, con $h = 1, \dots, H$, il *Chi-plot* prevede di rappresentare i punti (χ_h, λ_h) definiti come segue:

$$\chi_h = \frac{\check{F}_{1,2}(u_{1,h}, u_{2,h}) - \check{F}_1(u_{1,h})\check{F}_2(u_{2,h})}{\sqrt{\check{F}_1(u_{1,h})(1 - \check{F}_1(u_{1,h}))\check{F}_2(u_{2,h})(1 - \check{F}_2(u_{2,h}))}}$$

$$\lambda_h = 4 \text{sign} [\bar{F}_1(u_{1,h})\bar{F}_2(u_{2,h})] \max [(\bar{F}_1(u_{1,h}))^2, (\bar{F}_2(u_{2,h}))^2]$$

dove \check{F}_1 , \check{F}_2 e $\check{F}_{1,2}$ rappresentano rispettivamente le funzioni di ripartizione empiriche delle v.a. uniformi \tilde{U}_1 , \tilde{U}_2 e della v.a. $(\tilde{U}_1, \tilde{U}_2)$, mentre $\bar{F}_1 = \check{F}_1 - 0,5$ e $\bar{F}_2 = \check{F}_2 - 0,5$. Le grandezze χ_h e λ_h dipendono solamente dai ranghi delle osservazioni e, per costruzione, sono definite nell'intervallo $[0, 1]$. In particolare,

λ_h misura la distanza del punto $(u_{1,h}, u_{2,h})$ rispetto al centro del data set osservato; sotto la situazione di indipendenza tra le due variabili marginali, vale che $\lambda_h \sim U[-1, 1]$, mentre $\chi_h \sim N(0, 1/H)$. In situazioni di dipendenza positiva tra le osservazioni i punti (χ_h, λ_h) si distribuiscono sopra lo zero, viceversa per variabili marginali dipendenti in modo negativo tra loro.

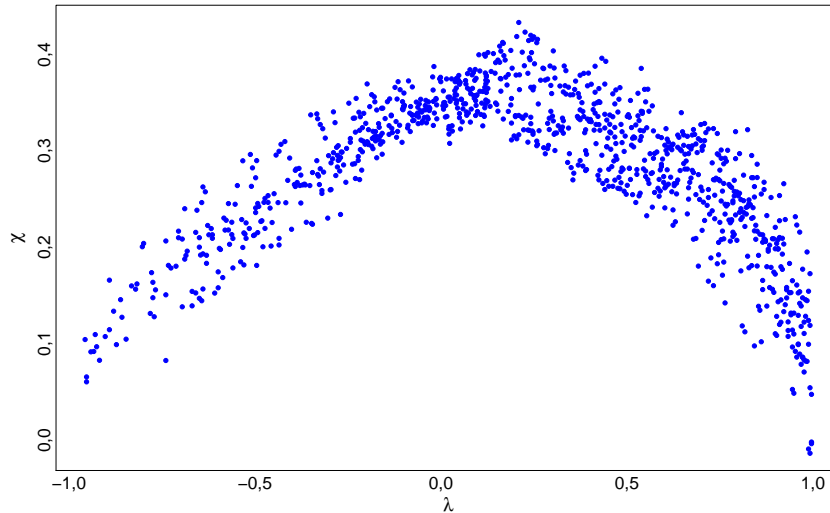


Figura 3.6: *Chi-plot* ottenuto con 1.000 realizzazioni casuali da una copula Normale di parametro $\rho = 0,5$.

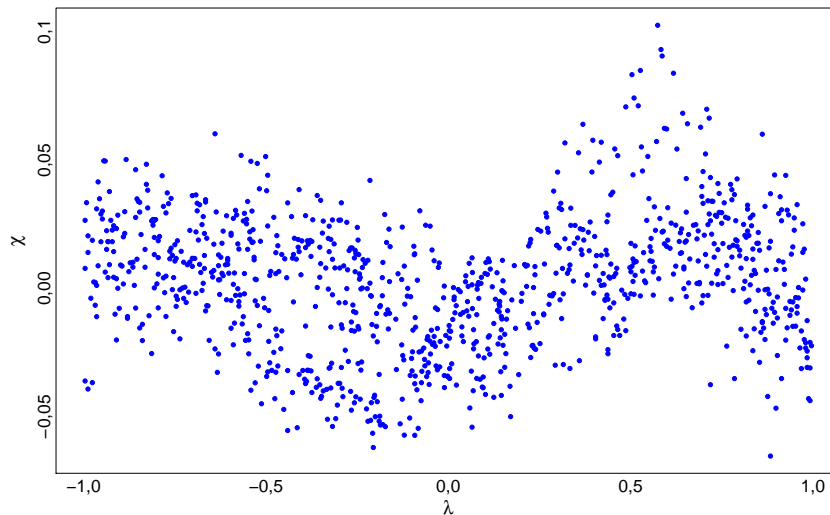


Figura 3.7: *Chi-plot* ottenuto con 1.000 realizzazioni casuali da una copula Normale di parametro $\rho = 0$.

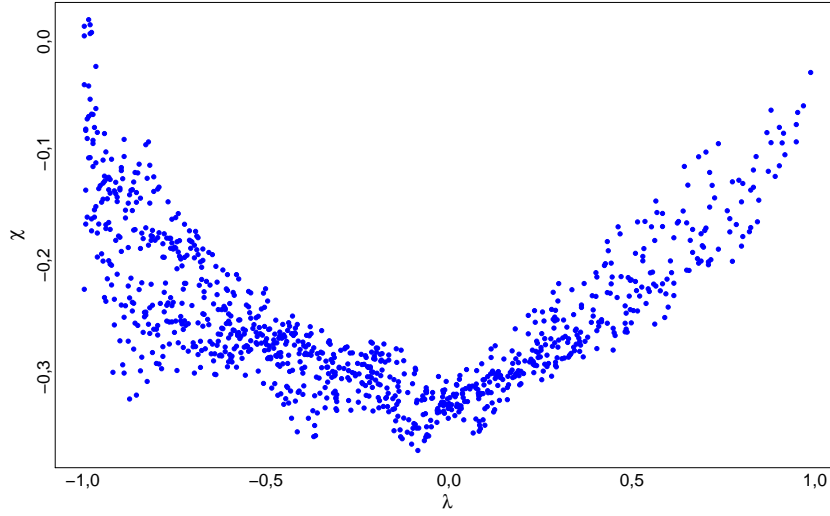


Figura 3.8: *Chi-plot* ottenuto con 1.000 realizzazioni casuali da una copula Normale di parametro $\rho = -0,5$.

Altro strumento grafico basato sui ranghi delle osservazioni è costituito dal cosiddetto *K - plot* (Genest e Favre (2007)). Tale approccio prevede di rappresentare nel piano $[0, 1] \times [0, 1]$ i punti $(\check{F}_{1,2}^{(h)}(u_{1,h}, u_{2,h}), W_{h:H})$, dove:

$$\check{F}_{1,2}^{(1)} < \dots < \check{F}_{1,2}^{(h)} < \dots < \check{F}_{1,2}^{(H)}$$

rappresentano le statistiche d'ordine associate ai valori assunti dalla funzione di ripartizione empirica bivariata. La grandezza $W_{h:H}$, invece, individua il valore atteso della statistica d'ordine h -esima del campione casuale di dimensione H della v.a. $\tilde{W} = C(u, v)$ sotto l'ipotesi di indipendenza delle v.a. \tilde{U} e \tilde{V} . La grandezza $W_{h:H}$ è definita come,

$$W_{h:H} = H \binom{H-1}{h-1} \int_0^1 \omega k_0(\omega) (K_0(\omega))^{h-1} (1 - K_0(\omega))^{H-h} d\omega$$

dove

$$K_0(\omega) = P(\tilde{U}\tilde{V} \leq \omega) = \omega - \omega \ln(\omega)$$

mentre k_0 individua la rispettiva funzione di densità. In tale rappresentazione grafica, la situazione di indipendenza è caratterizzata dai punti $(\check{F}_{1,2}^{(h)}, W_{h:H})$ disposti lungo la diagonale principale del quadrato di lato unitario. Ogni scostamento da tale situazione è sintomo della presenza di dipendenza nei dati. Vi è dipendenza negativa quando i punti si dispongono al disotto della diagonale; in particolare, nel caso di contromonotonicità i punti sono allineati all'asse delle ascisse. Di contro, nel caso di dipendenza positiva i punti risultano al di sopra della diagonale. Nel caso di comonotonicità, infine, i punti coincidono con la curva $K_0(\omega)$. Tale strumento, oltre che per effettuare analisi esplorative dei dati,

può essere utilizzato per confrontare i punti $(\check{F}_{1,2}^{(h)}, W_{h:H})$ determinati dalle osservazioni con quelli associati ad estrazioni casuali del modello prescelto al fine di valutarne graficamente l'adattamento.

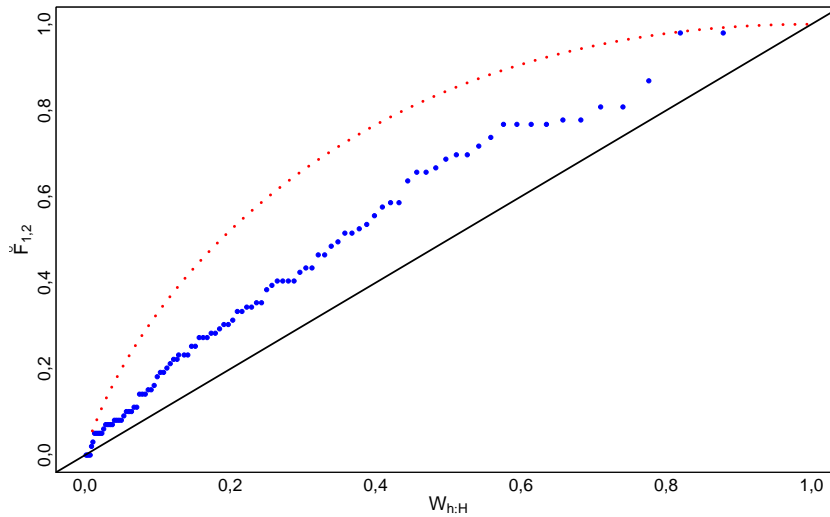


Figura 3.9: K -plot ottenuto con 100 realizzazioni casuali da una copula Normale di parametro $\rho = 0,5$. La curva $K_0(\omega)$ è individuata dalla linea tratteggiata di colore rosso.

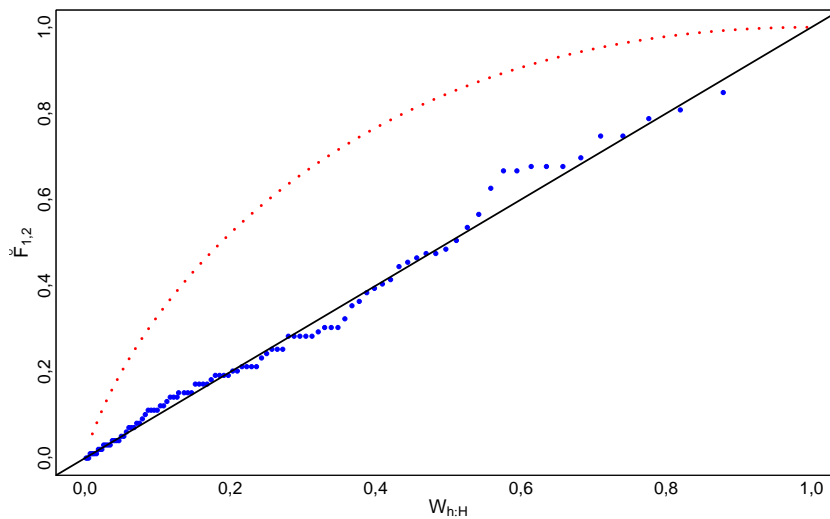


Figura 3.10: K -plot ottenuto con 100 realizzazioni casuali da una copula Normale di parametro $\rho = 0$. La curva $K_0(\omega)$ è individuata dalla linea tratteggiata di colore rosso.

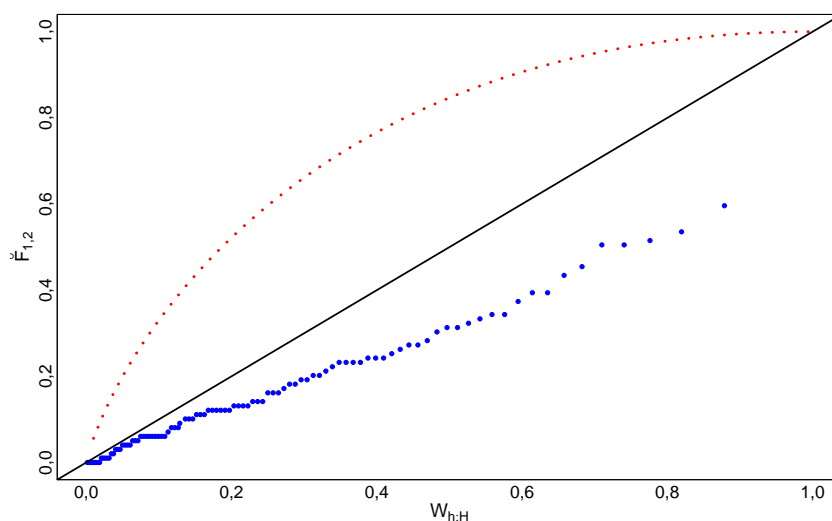


Figura 3.11: K -plot ottenuto con 100 realizzazioni casuali da una copula Normale di parametro $\rho = -0,5$. La curva $K_0(\omega)$ è individuata dalla linea tratteggiata di colore rosso.

La funzione *lambda* rappresenta un ulteriore approccio che permette di individuare tra molteplici modelli quello più appropriato (Wang e Wells (2000)). Tale strumento prevede di confrontare la funzione $\lambda(t)$ empirica con quella teorica associata al modello selezionato. In particolare, la funzione $\lambda(t)$ è definita come segue:

$$\lambda(t) = t - K(t) \quad \text{con} \quad t \in [0, 1]$$

dove

$$K(t) = P(C(u, v) \leq t) \quad \text{con} \quad t \in [0, 1]$$

individua la funzione di ripartizione della copula $C(u, v)$. Si noti che, ricorrendo alla relazione (2.4.2), per le copule archimedee vale:

$$\lambda(t) = \frac{\varphi(t)}{\varphi'(t^+)}$$

dove φ è il generatore della copula archimedeae.

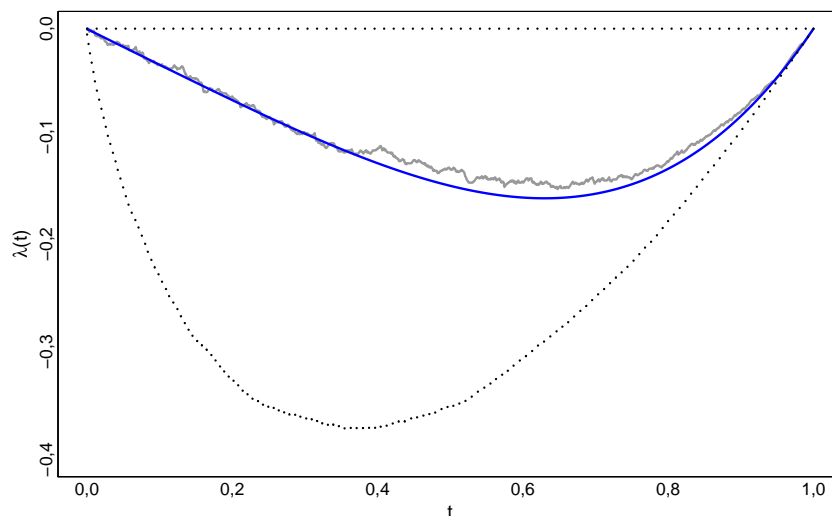


Figura 3.12: *Lambda-plot* ottenuto con 1.000 realizzazioni casuali da una copula Clayton di parametro $\theta = 3$ ($\tau = 0,6$). Le curve in colore grigio e blu rappresentano rispettivamente la funzione $\lambda(t)$ empirica e teorica; la linea nera tratteggiata orizzontale corrisponde alla funzione $\lambda(t)$ nel caso di perfetta dipendenza positiva, $\tau = 1$, mentre la curva inferiore individua il caso in cui $\tau = 0$.

Ricorrendo alla funzione $K(t)$, in Genest, Quessy e Rémillard (2006) è proposto un test statistico per valutare la bontà di adattamento di una prescelta famiglia di copule rispetto ai dati oggetto di analisi. In particolare, il test prevede di confrontare la funzione di ripartizione empirica della copula, definita $K_h(t)$, con la rispettiva funzione di ripartizione teorica, definita $K(\theta_h, t)$, ottenuta sotto l'ipotesi nulla $H_0 : C \in \mathcal{C}$. Le statistiche test costruite dagli autori, basate rispettivamente sui criteri di Cramér-von Mises e di Kolmogorov-Smirnov, sono le seguenti:

$$S_h = \int_0^1 |\mathbb{K}_h(t)|^2 k(\theta_h, t) dt$$

e

$$T_h = \sup_{0 \leq t \leq 1} |\mathbb{K}_h(t)|$$

dove $\mathbb{K}_h(t)$ individua il processo $\mathbb{K}_h(t) = \sqrt{h}\{K_h(t) - K(\theta_h, t)\}$. Il test statistico prevede di rifiutare l'ipotesi H_0 quando i valori osservati della statistica S_h o della statistica T_h risultano maggiori del valore assunto dal percentile al livello di confidenza $100(1 - \alpha)\%$ della distribuzione statistica sotto l'ipotesi nulla. Si noti che gli stimatori S_h e T_h dipendono dall'ignoto parametro θ associato al modello oggetto di analisi; al fine di superare tale inconveniente, gli autori propongono un approccio bootstrap parametrico al fine di quantificare i valori del p-value.

Altro test statistico utile per valutare la correttezza del modello associato alla copula bivariata risulta quello basato sulla cosiddetta matrice di informazione di

White. Considerando il metodo di stima della massima verosimiglianza, infatti, in White (1982) si dimostra che nel caso di corretta specificazione del modello, l'informazione di Fisher coincide con $-\mathbb{H}(\theta) = \mathbb{C}(\theta)$, dove $\mathbb{H}(\theta)$ e $\mathbb{C}(\theta)$ rappresentano rispettivamente il valore atteso della matrice Hessiana e il valore atteso del prodotto della rispettiva funzione score. Il test di ipotesi, quindi, è costruito come segue:

$$H_0 : \mathbb{H}(\theta) + \mathbb{C}(\theta) = 0$$

$$H_1 : \mathbb{H}(\theta) + \mathbb{C}(\theta) \neq 0$$

Dopo avere selezionato il modello vine e le famiglie di copule bivariate associate agli archi di ogni albero, è necessario stimare i parametri delle copule bidimensionali coinvolte. In letteratura esistono differenti approcci⁴ utili per risolvere il problema inferenziale, quali ad esempio il metodo dei momenti basato sull'inversione del τ di Kendall e del ρ di Spearman. Tali metodi, tuttavia, non risultano perseguibili nel caso in cui la famiglia di copule selezionata sia caratterizzata da un vettore di parametri. Di contro, l'approccio della verosimiglianza, sotto la condizione che la copula C sia assolutamente continua, permette di stimare il vettore di parametri caratteristici della copula in esame. Quest'ultimo metodo è quello utilizzato nel contesto delle vine copule; in letteratura, infatti, tale approccio è stato sviluppato al fine di adattarlo alle specificità della pair-copula construction.

Data la matrice di $h = 1, \dots, H$ osservazioni, $\mathbf{x} = (x_{1,h}, \dots, x_{n,h})$, della v.a. n -dimensionale $\widetilde{\mathbf{X}} = (\widetilde{X}_1, \dots, \widetilde{X}_n)$, la funzione di log-verosimiglianza, associata ad esempio ad una struttura D-vine, risulta essere⁵:

$$l(\boldsymbol{\theta}; \mathbf{x}) = \sum_{h=1}^H \sum_{j=1}^{n-1} \sum_{i=1}^{n-j} \ln [c_{i,i+j|i+1,\dots,i+j-1}(F_{i|i+1,\dots,i+j-1}(x_i|x_{i+1}, \dots, x_{i+j-1}), F_{i+j|i+1,\dots,i+j-1}(x_{i+j}|x_{i+1}, \dots, x_{i+j-1}); \boldsymbol{\theta}_{i,\dots,i+j})] \quad (3.2.1.1)$$

dove $\boldsymbol{\theta}$ individua il parametro, o il vettore dei parametri, caratteristico delle singole copule bivariate oggetto di stima.

Data la non conoscenza delle variabili marginali, le quali vengono modellizzate separatamente dal modello di dipendenza, la stima del vettore dei parametri delle copule è ottenuta seguendo un approccio semiparametrico (abbreviato in

⁴Per una panoramica generale sui metodi inferenziali esistenti in letteratura nel contesto delle copule si rimanda a Genest e Favre (2007).

⁵Si ricorda che le variabili marginali non sono oggetto di stima nel modello vine copula e, quindi, sono considerate uniformi, $\prod_{k=1}^n f_k(x_k) = 1$.

letteratura con l'acronimo SP, *semiparametric*), in cui vengono utilizzate le funzioni di ripartizione empiriche, o pseudo-osservazioni, delle variabili marginali. Si noti che le funzioni di ripartizione della v.a. marginali permettono, ricorrendo iterativamente alla (3.1.9), di specificare gli argomenti di tutte le copule condizionate coinvolte nella vine copula. Il metodo della massima verosimiglianza prevede di stimare i parametri massimizzando la relazione (3.2.1.1) rispetto al vettore $\boldsymbol{\theta}$. Si noti, che le copule che compaiono negli alberi successivi al primo, (T_2, \dots, T_{n-1}) , dipendono dai parametri delle copule precedenti; gli addendi della relazione (3.2.1.1), quindi, devono essere considerati congiuntamente al fine di ottenere il vettore di parametri $\hat{\boldsymbol{\theta}}^{SP}$ che massimizza la funzione di verosimiglianza della vine copula in esame. Lo stimatore $\hat{\boldsymbol{\theta}}^{SP}$, sotto opportune condizioni di regolarità, risulta essere consistente e asintoticamente Normale (Haff (2013), Tsukahara (2005)).

La massimizzazione congiunta della funzione di verosimiglianza della vine copula risulta essere, al crescere della dimensione e del numero dei parametri del problema, un'operazione computazionalmente onerosa che necessita di un vettore adeguato di valori di partenza al fine di stimare i parametri delle copule coinvolte nel vine. Per gestire tale aspetto, in letteratura l'approccio SP è stato modificato al fine di adattarlo alla logica della PCC, dando luogo ad un approccio di stima semiparametrico *stepwise* (SSP, *stepwise semiparametric*). La procedura stepwise prevede di stimare i parametri delle copule bivariate della PCC presenti nella (3.2.1.1) separatamente, considerando una copula alla volta, iniziando dall'albero al livello superiore per poi, in seguito, stimare i parametri delle copule site ai livelli inferiori.

A titolo esemplificativo si consideri un D-vine in 4 dimensioni con funzione di densità, ponendo $f_i(x_i) = 1$ con $i = 1, \dots, 4$, fattorizzata come segue:

$$\begin{aligned} f_{1,2,3,4}(x_1, x_2, x_3, x_4) = & c_{1,2}(F_1(x_1), F_2(x_2)) \cdot c_{2,3}(F_2(x_2), F_3(x_3)) \cdot c_{3,4}(F_3(x_3), F_4(x_4)) \\ & \cdot c_{1,3|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \cdot c_{2,4|3}(F_{2|3}(x_2|x_3), F_{4|3}(x_4|x_3)) \\ & \cdot c_{1,4|2,3}(F_{1|2,3}(x_1|x_2, x_3), F_{4|2,3}(x_4|x_2, x_3)) \end{aligned}$$

Considerando $h = 1, \dots, H$ realizzazioni della variabile casuale $\widetilde{\mathbf{X}}$, la funzione di log-verosimiglianza del D-vine, dopo avere sostituito alle funzioni di ripartizione marginali le rispettive funzioni empiriche, diviene:

$$\begin{aligned} l(\boldsymbol{\theta}; \mathbf{x}) = & \sum_{h=1}^H c_{1,2}(u_{(1),h}, u_{(2),h}; \boldsymbol{\theta}_{1,2}) + \sum_{h=1}^H c_{2,3}(u_{(2),h}, u_{(3),h}; \boldsymbol{\theta}_{2,3}) + \sum_{h=1}^H c_{3,4}(u_{(3),h}, u_{(4),h}; \boldsymbol{\theta}_{3,4}) \\ & + \sum_{h=1}^H c_{1,3|2}(u_{(1|2),h}, u_{(3|2),h}; \boldsymbol{\theta}_{1,3|2}) + \sum_{h=1}^H c_{2,4|3}(u_{(2|3),h}, u_{(4|3),h}; \boldsymbol{\theta}_{2,4|3}) \\ & + \sum_{h=1}^H c_{1,4|2,3}(u_{(1|2,3),h}, u_{(4|2,3),h}; \boldsymbol{\theta}_{1,4|2,3}) \end{aligned}$$

L'approccio stepwise prevede di stimare disgiuntamente i parametri delle singole copule coinvolte nel vine; in particolare, poiché le copule che compaiono negli alberi successivi al primo dipendono dalle copule dei livelli precedenti, la procedura di stima prevede di quantificare inizialmente i parametri delle copule del primo albero. I termini da massimizzare disgiuntamente, quindi, risultano:

$$\begin{aligned} & - \sum_{h=1} c_{1,2}(u_{(1),h}, u_{(2),h}; \boldsymbol{\theta}_{1,2}); \\ & - \sum_{h=1} c_{2,3}(u_{(2),h}, u_{(3),h}; \boldsymbol{\theta}_{2,3}); \\ & - \sum_{h=1} c_{3,4}(u_{(3),h}, u_{(4),h}; \boldsymbol{\theta}_{3,4}). \end{aligned}$$

Dopo avere stimato i parametri delle copule del primo albero, ricorrendo alla relazione (3.1.9), è possibile calcolare le funzioni di ripartizione delle variabili condizionate coinvolte nel secondo albero: $F_{1|2}(x_1|x_2)$, $F_{3|2}(x_3|x_2)$, $F_{2|3}(x_2|x_3)$ e $F_{4|3}(x_4|x_3)$. In seguito, la massimizzazione delle funzioni di log-verosimiglianza delle copule associate al secondo livello del D-vine permette di stimarne i parametri. Le funzioni da massimizzare rispetto ai termini $\boldsymbol{\theta}$ sono,

$$\begin{aligned} & - \sum_{h=1} c_{1,3|2}(u_{(1|2),h}, u_{(3|2),h}; \boldsymbol{\theta}_{1,3|2}); \\ & - \sum_{h=1} c_{2,4|3}(u_{(2|3),h}, u_{(4|3),h}; \boldsymbol{\theta}_{2,4|3}). \end{aligned}$$

La quantificazione dei vettori dei parametri associati alle copule $c_{1,3|2}$ e $c_{2,4|3}$ permette, grazie alla (3.1.9), di calcolare gli argomenti della copula associata all'ultimo albero del vine, che risultano essere le pseudo-osservazioni delle variabili $\tilde{X}_{1|2,3}$ e $\tilde{X}_{4|2,3}$. La stima del vettore di parametri $\boldsymbol{\theta}_{1,4|2,3}$, infine, è ottenibile massimizzando la funzione di log-verosimiglianza della copula $c_{1,4|2,3}$:

$$- \sum_{h=1} c_{1,4|2,3}(u_{(1|2,3),h}, u_{(4|2,3),h}; \boldsymbol{\theta}_{1,4|2,3}).$$

Lo stimatore $\hat{\boldsymbol{\theta}}^{SSP}$ risulta anch'esso uno stimatore consistente e asintoticamente Normale; $\hat{\boldsymbol{\theta}}^{SSP}$ tuttavia è asintoticamente meno efficiente rispetto a $\hat{\boldsymbol{\theta}}^{SP}$ in quanto ad ogni livello non tiene conto delle informazioni negli alberi successivi (Haff (2012, 2013)). Lo stimatore stepwise, tuttavia, riduce l'onere computazione del processo di stima, risultando particolarmente adeguato in contesti inferenziali di dimensione elevata in cui il numero di parametri diviene numeroso. L'approccio stepwise, inoltre, può essere utilizzato per determinare i valori di partenza dell'approccio SP; stime molto differenti tra i metodi SP e SSP sono sintomo, inoltre, dell'errata scelta delle famiglie di copule utilizzate nel vine. In generale, come mostrato in Haff (2012), lo stimatore $\hat{\boldsymbol{\theta}}^{SSP}$, sebbene non massimizzi la (3.2.1.1), fornisce stime molto prossime a quelle dello stimatore $\hat{\boldsymbol{\theta}}^{SP}$.

Si noti, infine, che in letteratura è presente un filone di ricerca, ad oggi ancora limitato, che affronta il problema inferenziale delle vine copule non tramite il

classico approccio frequentista bensì ricorrendo al framework statistico bayesiano. In particolare, in Min e Czado (2010) la stima dei parametri delle pair-copule è perseguita tramite algoritmo Markov Chain Monte Carlo (MCMC) al fine di effettuare una stima intervallare dei parametri. In seguito, gli stessi autori (Min e Czado (2011)), considerando la famiglia dei D-vine, ricorrono ad un approccio definito *reversible jump* Markov Chain Monte Carlo (Green (1995)) con lo scopo di selezionare non solo le copule bivariate coinvolte ma anche la struttura del modello D-vine. L'approccio bayesiano declinato tramite la procedura *reversible jump* MCMC è utilizzato nel contesto delle vine copule con il primario scopo di selezionare la struttura di modello di vine copula più adeguata, evitando di enumerarne in modo esplicito tutte le possibili configurazioni come richiesto dagli algoritmi sviluppati sotto l'impostazione frequentista, quali ad esempio l'algoritmo di *Dißmann*. In Gruber e Czado (2015) la procedura *reversible jump* MCMC è estesa al fine di considerare tutte le possibili strutture di regular vine esistenti, permettendo di selezionare sia la configurazione più adeguata, nonché le relative pair-copule e i parametri sottostanti. Gli autori sviluppano l'approccio ricorrendo ad una procedura sequenziale in cui la stima considera tutti i singoli alberi dei vari livelli condizionatamente a quelli dei livelli superiori. Gruber e Czado (2018), infine, estendono il loro precedente lavoro sviluppando una procedura di stima bayesiana che considera simultaneamente tutti gli alberi associati ai livelli della struttura del regular vine.

Capitolo 4

Caso studio: la quantificazione del non-life reserve risk

Nell'ultimo decennio le vine copule sono state notevolmente utilizzate in ambito finanziario e di risk management al fine di costruire modelli di dipendenza multivariati, ne sono un esempio i lavori di Aas (2016), Allen, Ashraf et al. (2013), Allen, McAleer e Singh (2017), Dißmann et al. (2013), Geidosh e Fischer (2016) e Nikoloulopoulos, Joe e Li (2012). Nel contesto assicurativo e attuariale, di contro, l'applicazione di tale metodologia è risultata particolarmente ridotta e la si ritrova nei lavori di Eling e Jung (2018), Mejdoub e Ben Arab (2017) e Schirmacher e Schirmacher (2008). In letteratura, inoltre, da ciò che è noto allo scrivente, non esistono applicazioni di modelli di vine copule al fine di valutare il rischio di riservazione delle assicurazioni danni.

Il presente capitolo ha lo scopo di mostrare un caso studio reale in cui si quantifica il non-life reserve risk considerando congiuntamente 4 linee di attività di una compagnia di assicurazione operante nel mercato italiano. Il data set oggetto di studio è costituito dai triangoli di run-off degli importi e dei numeri dei sinistri pagati dei 4 rami assicurativi. Coerentemente con la teoria delle copule, il rischio di riservazione è valutato modellizzando separatamente la struttura di dipendenza tra i rischi e le variabili marginali.

In particolare, le vine copule sono utilizzate per modellizzare la dipendenza esistente tra le Lob considerate. La dipendenza implicita esistente nei triangoli di run-off è estratta ricorrendo ad un approccio bootstrap dipendente; tale procedura permette di superare il problema, tipico del contesto in esame, della limitata numerosità dei dati, permettendo di calibrare il modello di vine copula su un campione di dati con numerosità arbitraria.

Le variabili marginali rappresentative del reserve risk delle singole linee di business, di contro, sono modellizzate tramite l'approccio Collective Risk Model (CRM) come sviluppato in Ricotta e Luini (2019); in tale contesto, la variabilità afferente ai parametri del modello, nonché la correlazione tra gli stessi sono quantificate ricorrendo ad una metodologia bayesiana. In seguito, i rischi marginali

sono aggregati per mezzo della vine copula stimata in precedenza permettendo di valutare a livello complessivo il requisito patrimoniale di solvibilità (SCR) afferente al reserve risk, come prescritto dalla normativa comunitaria Solvency II.

4.1 Il data set di riferimento

La valutazione del rischio di riservazione presentata nel seguito si basa sui triangoli di run-off della riserva sinistri afferenti a 4 Lob e provenienti da una reale compagnia di assicurazione danni operante nel mercato italiano. I dati, al fine di preservarne la confidenzialità, sono stati parzialmente modificati. Le grandezze prese a riferimento sono gli importi e il numero dei sinistri pagati per gli anni di calendario dal 1995 al 2006 e relativi a 12 anni di sviluppo di pagamenti; i triangoli di run-off, quindi, possiedono righe e colonne in numero pari a 12, $i = 1, \dots, N$ e $j = 1, \dots, N$ con $N = 12$, che individuano rispettivamente gli anni di accadimento dei sinistri (cd. *accident years*, AY) e le antiduarate o anni di sviluppo dei sinistri (cd. *development years*, DY). Le linee di attività considerate sono le seguenti: responsabilità civile derivante dalla circolazione di autoveicoli terrestri (RCA), responsabilità civile generale (RCG), assicurazione incendio ed eventi naturali (IEN) e assicurazione altri danni ai beni (ADB). In Appendice A si riportano i triangoli di run-off osservati degli importi e dei numeri cumulati dei sinistri pagati per le 4 linee di business considerate.

La stima deterministica della riserva sinistri è ottenuta tramite metodo Frequency-Severity¹ basato sulla meccanica Chain-Ladder. Tale metodo si basa sul data set delle osservazioni storiche dei triangoli di run-off incrementali del numero ($n_{i,j}$) e importo ($X_{i,j}$) dei sinistri, individuati rispettivamente con $D^n = \{n_{i,j} \leq N + 1\}$ e $D = \{X_{i,j} \leq N + 1\}$; il metodo, attraverso i triangoli cumulati, proietta il numero dei sinistri e il rispettivo costo medio, permettendo di quantificare gli importi dei sinistri che si stima vengano liquidati nei futuri anni di calendario. Dato il triangolo cumulato storico del numero dei sinistri pagati, il primo step del metodo, coerentemente con la logica Chain-Ladder, consiste nel calcolare il fattore di sviluppo, o link ratio, (λ_j^n) per ogni DY:

$$\lambda_j^n = \frac{\sum_{i=1}^{N-j} n_{i,j+1}^c}{\sum_{i=1}^{N-j} n_{i,j}^c} \quad \text{con} \quad j = 1, \dots, N - 1$$

dove $n_{i,j}^c$ individua il numero cumulato dei sinistri pagati nella generica cella (i, j) del triangolo di run-off. Il numero atteso stimato dei sinistri cumulati è ottenuto come:

$$\hat{n}_{i,j}^c = n_{i,N-i+1}^c \prod_{k=N-i+1}^{j-1} \lambda_k^n \quad \text{con} \quad \begin{cases} i = 1, \dots, N \\ j = N - i + 2, \dots, N \end{cases}$$

¹Per ulteriori dettagli si veda, ad esempio, Friedland (2010).

dove $n_{i,N-i+1}^c$ rappresenta l'ultimo valore osservato per la generazione di accadimento i -esima, situato sull'ultima diagonale del triangolo che individua l'ultimo anno di calendario disponibile. In seguito, la stima del numero incrementale dei sinistri $\hat{n}_{i,j}$ è ottenuta tramite differenza delle rispettive grandezze cumulate. Analogamente, lo stesso procedimento è applicato al triangolo dei costi medi cumulati, definito dal rapporto tra gli ammontari dei pagamenti cumulati ($C_{i,j}$) e i rispettivi numeri cumulati:

$$CM_{i,j}^c = \frac{C_{i,j}}{n_{i,j}^c}$$

Il link ratio, λ_j^{CM} , per ogni anno di sviluppo del triangolo è definito come:

$$\lambda_j^{CM} = \frac{\sum_{i=1}^{N-j} CM_{i,j+1}^c}{\sum_{i=1}^{N-j} CM_{i,j}^c} \quad \text{con} \quad j = 1, \dots, N-1$$

Il triangolo inferiore dei costi medi cumulati, in seguito, è così stimato:

$$\widehat{CM}_{i,j}^c = CM_{i,N-i+1}^c \prod_{k=N-i+1}^{j-1} \lambda_k^{CM} \quad \text{con} \quad \begin{cases} i = 1, \dots, N \\ j = N - i + 2, \dots, N \end{cases}$$

La stima degli importi cumulati dei sinistri pagati che verranno liquidati negli anni di calendario futuri è ottenuta moltiplicando il numero cumulato stimato dei sinistri per la rispettiva stima del costo medio cumulato:

$$\hat{C}_{i,j} = \hat{n}_{i,j}^c \cdot \widehat{CM}_{i,j}^c \quad \text{con} \quad \begin{cases} i = 1, \dots, N \\ j = N - i + 2, \dots, N \end{cases}$$

Infine, la differenza tra le grandezze $\hat{C}_{i,j}$ permette di ottenere l'importo stimato dei sinistri incrementali che verranno liquidati negli anni futuri, individuati con $\hat{X}_{i,j}$. Si noti che il rapporto tra le grandezze $\hat{X}_{i,j}$ e $\hat{n}_{i,j}$ permette di valutare il costo medio futuro dei sinistri incrementali, $\hat{m}_{i,j} = \hat{X}_{i,j} / \hat{n}_{i,j}$.

La tabella seguente riporta i valori di riserva sinistri e i numeri dei sinistri, per le 4 Lob oggetto di studio, stimati tramite il metodo deterministico Frequency-Severity.

Lob	Riserva sinistri	Numero sinistri
RCA	947.803	66.761
RCG	332.007	11.063
IEN	136.717	4.716
ADB	41.509	1.598

Tabella 4.1: Ammontare di riserva sinistri e numero dei sinistri stimati tramite il metodo Frequency-Severity (importi in migliaia di euro).

Nel presente caso studio il metodo Frequency-Severity è alla base dell'approccio bootstrap dipendente utilizzato al fine di ottenere il data set sintetico con cui calibrare il modello di vine copula; tale metodo deterministico di stima della riserva sinistri è adottato, inoltre, per quantificare i parametri della *frequency* ($\hat{n}_{i,j}$) e della *severity* ($\hat{m}_{i,j}$) del modello CRM utilizzato per modellizzare le variabili aleatorie marginali.

4.2 La modellizzazione della dipendenza tramite vine copule

La valutazione della dipendenza del rischio di riservazione tra più linee di business nel contesto delle assicurazioni danni è in genere basata sulla serie storica di indicatori sintetici del rischio oggetto di studio; ne è un esempio il cosiddetto run-off ratio, definito come il rapporto tra lo smontamento della riserva sinistri (cd. run-off) osservato su un arco temporale di riferimento e il valore iniziale stimato di riserva sinistri. La profondità storica di tali indicatori, tuttavia, è limitata in quanto sono costruiti considerando, in genere, orizzonti temporali di osservazione annui.

L'approccio presentato nel seguito oltrepassa l'inconveniente della limitata numerosità dei dati e permette di quantificare la dipendenza tra il reserve risk di più Lob partendo direttamente dai triangoli di run-off osservati. Alla base dell'approccio sviluppato vi è la metodologia ampiamente utilizzata in letteratura e nella prassi attuariale definita dal metodo bootstrap (Efron e Tibshirani (1993)) applicato alla meccanica Chain-Ladder (England e Verrall (2002)). La metodologia statistica definita dal bootstrap è nata con lo scopo di effettuare inferenza sui dati oggetto di studio. Considerando un campione di valori indipendenti e identicamente distribuiti tra loro, il metodo prevede di rideterminare il campione stesso tramite estrazioni casuali e con reimmissione dei dati di partenza; in tale contesto, stante l'indipendenza e l'identica distribuzione dei dati, il metodo è qualificato come non parametrico in quanto non si basa su alcuna ipotesi distributiva delle grandezze osservate. La generazione di un numero sufficientemente elevato di pseudo-campioni, infine, permette di valutare la distribuzione degli stimatori di interesse, permettendo di fare inferenza sul campione originario di partenza. Di contro, quando i dati originari non rispettano l'ipotesi di indipendenza e identica distribuzione, il metodo bootstrap viene applicato ai residui di un modello che si ipotizza abbia generato i dati; in tale circostanza il bootstrap è definito parametrico. In particolare, il bootstrap parametrico basato sul metodo Chain-Ladder è la metodologia adottata e sviluppata originariamente da England e Verrall (1999) al fine di effettuare la valutazione stocastica della riserva sinistri; di seguito, tale approccio è adottato per ottenere un data set *sintetico* di valori di riserva sinistri dipendenti tra loro al fine di quantificare e modellizzare la dipendenza esistente tra il reserve risk delle Lob oggetto di studio. L'approccio

adottato per estrarre la dipendenza tra i triangoli di run-off delle varie linee di business è basato su uno schema bootstrap dipendente in cui il ricampionamento dei residui, basati sulla meccanica Chain-Ladder, è effettuato congiuntamente per ogni Lob considerata; in particolare, ad ogni step di ricampionamento vengono estratte le grandezze che occupano la medesima posizione (cella) nei triangoli di run-off delle 4 linee di business. La procedura, quindi, consiste nel ricampionare gli indici i e j che individuano la posizione dei dati nel triangolo; lo schema di ricampionamento congiunto, quindi, si sostanzia nell'estrarre, ad ogni step simulativo, per tutti i triangoli di run-off delle Lob considerate gli stessi indici. Tale approccio bootstrap è applicato ai triangoli del numero e degli importi dei sinistri, coerentemente con il metodo Frequency-Severity di stima deterministica della riserva sinistri, permettendo di ottenere per ogni Lob un vettore di valori stimati di riserva sinistri; il ricampionamento congiunto determina dei valori di riserva per i 4 rami dipendenti tra loro. Lo schema congiunto di ricampionamento applicato alle Lob, dunque, permette di estrarre la dipendenza contenuta nei dati osservati nei triangoli di run-off ricorrendo al metodo deterministico di stima della riserva sinistri Frequency-Severity; la dipendenza presente nel data set sintetico così ottenuto, quindi, è funzione non solo della dipendenza insita nei dati e colta attraverso il ricampionamento congiunto, ma anche del metodo deterministico adottato per stimare la riserva sinistri sottostante all'approccio bootstrap parametrico. Si noti che tale approccio può essere applicato, *mutatis mutandis*, considerando qualsiasi metodo deterministico di stima della riserva sinistri basato sui triangoli di run-off.

Nel dettaglio, i passaggi della procedura bootstrap, coerentemente con il metodo Frequency-Severity, per ogni step di ricampionamento, considerando i triangoli osservati del numero e degli importi cumulati dei sinistri di una generica Lob, sono i seguenti.

- 1) Determinare i fattori di sviluppo del metodo Chain-Ladder del triangolo delle grandezze cumulate osservate.
- 2) Dai link ratio e dai dati osservati sull'ultima diagonale del triangolo determinare ricorsivamente le grandezze cumulate del triangolo superiore; tramite differenza, in seguito, calcolare le grandezze incrementalì.
- 3) Calcolare i residui del modello tra le grandezze incrementalì del passo precedente e i dati osservati dei triangoli di run-off.
- 4) Ricampionare con reimmissione i residui creando il triangolo di run-off dei residui e, in seguito, da questi calcolare le grandezze cumulate.

Tale procedura permette di ottenere, per ogni step simulativo, il triangolo ricampionato del numero e degli importi dei sinistri pagati; in seguito, applicando il metodo Frequency-Severity si giunge alla stima del valore di riserva sinistri. Lo schema di campionamento congiunto interviene al punto (4) della procedura bootstrap; in particolare, il ricampionamento dei residui è effettuato considerando per

tutte e 4 le Lob gli stessi indici di posizione i e j . Le grandezze ricampionate ad ogni step simulativo, quindi, occupano la stessa posizione, ossia la stessa cella, nei triangoli di run-off delle Lob considerate. Il ricampionamento congiunto, quindi, permette di estrarre implicitamente la dipendenza insita nei dati dei triangoli di run-off, ottenendo dei vettori di riserva sinistri, per costruzione, dipendenti tra loro. Si noti che la dipendenza è ricavata facendo uso del metodo deterministico prescelto di stima della riserva sinistri, che in tale circostanza risulta essere il Frequency-Severity.

Lo schema di ricampionamento congiunto di cui sopra permette di ottenere un data set *sintetico* di valori di riserva sinistri dipendenti tra loro su cui stimare il modello di vine copula. In particolare, nel presente caso studio il numero di iterazioni prescelto è pari a $H = 10.000$, ottenendo quindi un data set corrispondente a una matrice di dati di 4 colonne, pari al numero di Lob considerate, e 10.000 righe. Quest'ultima rappresenta le realizzazioni casuali della funzione di densità multivariata empirica oggetto di analisi, $f_{1,2,3,4}(x_1, x_2, x_3, x_4)$, dove per brevità di notazione, le v.a. marginali RCA, RCG, IEN e ADB sono state indicate con la numerazione progressiva 1, 2, 3 e 4, da cui $f_{1,2,3,4}(x_1, x_2, x_3, x_4) = f_{RCA,RCG,IEN,ADB}(x_{RCA}, x_{RCG}, x_{IEN}, x_{ADB})$.

$$\begin{array}{c}
 \text{RCA} \quad \text{RCG} \quad \text{IEN} \quad \text{ADB} \\
 \begin{array}{c} 1 \\ \cdot \\ \cdot \\ \cdot \\ H \end{array} \left[\begin{array}{cccc} x_{RCA,1} & x_{RCG,1} & x_{IEN,1} & x_{ADB,1} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ x_{RCA,H} & x_{RCG,H} & x_{IEN,H} & x_{ADB,H} \end{array} \right]
 \end{array}$$

La tabella seguente riporta la differenza tra il valore di riserva deterministica ottenuta con il Frequency-Severity e il valore atteso di riserva calcolato sul vettore ricostruito tramite metodo bootstrap dipendente:

Lob	Riserva sinistri	Valore atteso (bootstrap)	Δ %
RCA	947.803	947.884	0,01%
RCG	332.007	332.481	0,14%
IEN	136.717	136.627	-0,07%
ADB	41.509	41.594	0,20%

Tabella 4.2: Confronto tra il valore deterministico e il valore atteso della riserva sinistri ricampionata con metodo bootstrap (importi in migliaia di euro).

Il metodo di campionamento bootstrap congiunto determina un data set sintetico di valori di riserva sinistri che mostra una dipendenza espressa in termini di τ di Kendall, ρ di Spearman e ρ di Pearson come riportato nelle tabelle seguenti. Come noto, il coefficiente di correlazione lineare è una misura di dipendenza che è funzione anche dalle caratteristiche distributive delle variabili marginali; nel

presente caso studio, tuttavia, le distribuzioni dei rischi marginali non sono considerate nel processo di modellizzazione della struttura di dipendenza sottostante alle Lob considerate. Al fine di studiare le relazioni di dipendenza esistenti tra i rischi marginali, quindi, è preferibile considerare le misure di concordanza del τ di Kendall e del ρ di Spearman.

τ	RCA	RCG	IEN	ADB
RCA	1			
RCG	0,13	1		
IEN	0,11	0,14	1	
ADB	0,07	0,11	0,11	1

Tabella 4.3: Valori del τ di Kendall calcolati sulla matrice di valori di riserva sinistri determinati tramite approccio bootstrap dipendente.

ρ_S	RCA	RCG	IEN	ADB
RCA	1			
RCG	0,20	1		
IEN	0,17	0,22	1	
ADB	0,11	0,18	0,17	1

Tabella 4.4: Valori del ρ di Spearman calcolati sulla matrice di valori di riserva sinistri determinati tramite approccio bootstrap dipendente.

ρ	RCA	RCG	IEN	ADB
RCA	1			
RCG	0,21	1		
IEN	0,16	0,22	1	
ADB	0,13	0,12	0,11	1

Tabella 4.5: Valori del ρ di Pearson calcolati sulla matrice di valori di riserva sinistri determinati tramite approccio bootstrap dipendente.

I grafici seguenti mostrano gli scatterplot delle coppie di vettori di riserva sinistri del data set sintetico, nonché le rispettive pseudo-osservazioni.

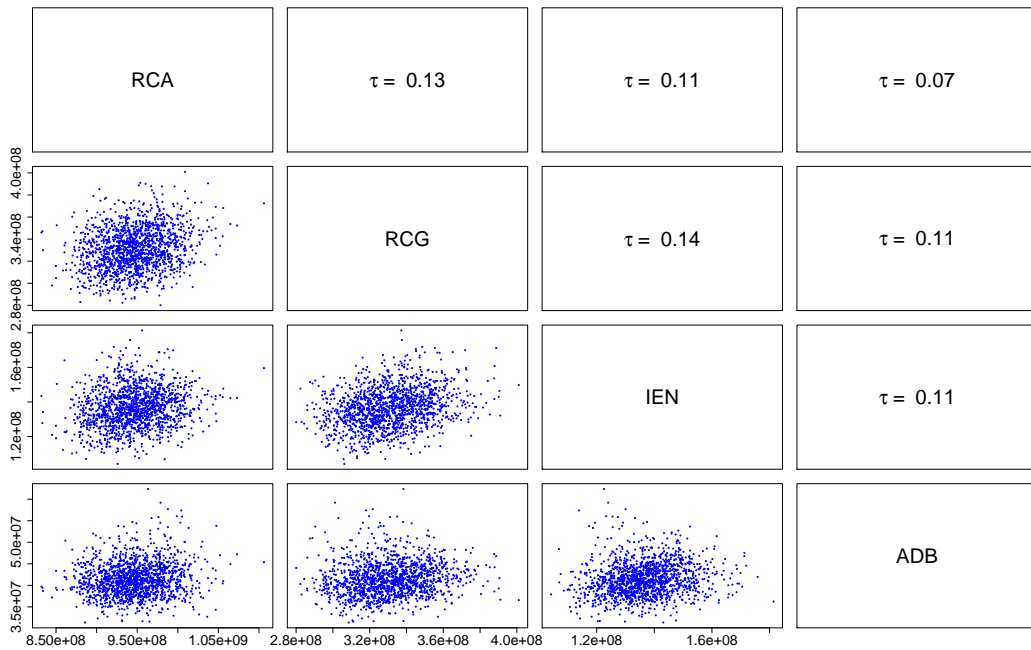


Figura 4.1: Scatterplot per le 4 Lob considerate dei valori di riserva ottenuti tramite l'approccio di ricampionamento bootstrap dipendente.

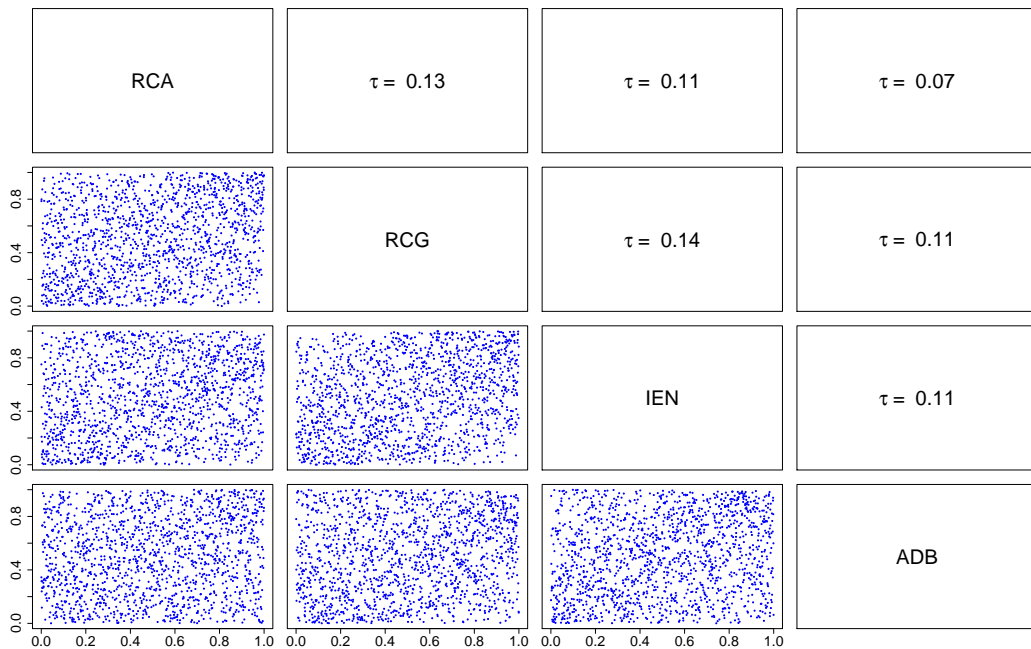


Figura 4.2: Scatterplot per le 4 Lob considerate delle pseudo-osservazioni ottenute tramite l'approccio di ricampionamento bootstrap dipendente.

Dopo avere ottenuto, come mostrato in precedenza, il data set sintetico dei valori dipendenti di riserva sinistri per le 4 Lob considerate, è possibile su tali dati stimare il modello di vine copula più adeguato. L'intero processo di stima della vine copula è stato implementato non effettuando alcuna analisi distributiva sulle variabili marginali; queste sono state considerate uniformi in $[0, 1]$, $f(x) = 1$, e rimappate all'interno del modello di vine copula con le rispettive pseudo-osservazioni, da cui $\tilde{X}_1 = \tilde{U}_1$, $\tilde{X}_2 = \tilde{U}_2$, $\tilde{X}_3 = \tilde{U}_3$ e $\tilde{X}_4 = \tilde{U}_4$. Coerentemente con la teoria delle copule la struttura di dipendenza è modellizzata separatamente dal modello sottostante alle variabili marginali. Inoltre, tutte le analisi riportate nel seguito, sviluppate al fine di individuare il modello vine più adeguato ai dati, sono state svolte sotto l'assunzione di ipotesi semplificata della pair-copula construction.

Come mostrato nel paragrafo 3.2.1, il processo di stima di una vine copula consta di tre passaggi: la scelta del modello vine, l'individuazione delle famiglie di copule bivariate e, infine, la rispettiva stima dei parametri. Dato il numero di variabili aleatorie marginali pari a quattro, tutte le possibili strutture di vine copule ricadono nella classe dei D-vine o dei C-vine, caratterizzate da tre livelli di alberi e complessivamente da sei archi, che individuano le copule bivariate coinvolte nella fattorizzazione della variabile aleatoria in 4 dimensioni. Stante il numero contenuto di variabili marginali, al fine di selezione la struttura vine più adeguata sono state stimate tutte le possibili strutture di regular vine, pari a $n!/2 = 12$ D-vine e C-vine rispettivamente. All'interno della classe dei D-vine e C-vine, in seguito, sono stati individuati i modelli migliori in termini di AIC. Infine, tramite i test statistici di Vuong e Clarke è stato individuato il modello di vine copula più adeguato.

L'individuazione delle famiglie di copule bivariate coinvolte in ogni albero della struttura vine è stata perseguita ricorrendo sia ad approcci grafici che statistici. In particolare, gli strumenti grafici definiti dal *Chi-plot*, *K-plot* e *Lambda-plot* hanno permesso di indagare la struttura di dipendenza sottostante ai dati. In seguito, per ogni copula bivariata coinvolta, partendo da un set di famiglie di copule considerate, è stata prescelta quella che mostra il migliore valore di AIC. Le famiglie di copule bivariate utilizzate nel processo di stima sono le copule ellittiche Normale e t-student, nonché le copule archimedee Clayton, Gumbel, Frank, Joe, BB1, BB6, BB7 e BB8; per le copule archimedee, inoltre, si sono considerate anche le rispettive copule ruotate di 90° , 180° e 270° .

I parametri (θ) delle copule coinvolte, infine, sono stimati tramite metodo della verosimiglianza basato sulle pseudo-osservazioni delle variabili marginali (approccio semiparametrico). In particolare, coerentemente con la teoria e prassi delle vine copule, i parametri sono determinati ricorsivamente per ogni albero tramite approccio semiparametrico stepwise, stimando le grandezze $\hat{\theta}^{SSP}$; tali stime, in seguito, sono utilizzate come valori di partenza per la procedura di massimizzazione della verosimiglianza della funzione di densità complessiva della vine copula, giungendo alla stima delle grandezze $\hat{\theta}^{SP}$ (Haff (2012, 2013)).

4.2.1 Stima del D-vine

Considerando tutte le possibili configurazioni di D-vine, pari a 12 date 4 variabili casuali, la tabella seguente riassume i valori di AIC, BIC e di log-verosimiglianza delle strutture D-vine stimate come riassunto precedentemente. La prima colonna contiene l'ordine delle variabili nel primo albero, il quale, come noto, per un D-vine individua univocamente la configurazione degli alberi successivi.

#	Ordine variabili 1° albero	AIC	BIC	Log-veros.
1	{RCA, RCG, ADB, IEN}	-1783	-1726	899
2	{RCA, IEN, ADB, RCG}	-1779	-1721	897
3	{RCG, IEN, ADB, RCA}	-1767	-1702	892
4	{IEN, RCG, RCA, ADB}	-1730	-1672	873
5	{IEN, RCG, ADB, RCA}	-1722	-1665	869
6	{IEN, RCA, ADB, RCG}	-1686	-1628	851
7	{RCA, RCG, IEN, ADB}	-1668	-1610	842
8	{RCA, IEN, RCG, ADB}	-1657	-1592	837
9	{RCG, RCA, ADB, IEN}	-1651	-1586	834
10	{RCG, RCA, FIRE, ADB}	-1587	-1529	801
11	{IEN, RCA, RCG, ADB}	-1586	-1528	801
12	{RCG, IEN, RCA, ADB}	-1542	-1477	780

Tabella 4.6: Strutture D-vine ordinate per valori crescenti della grandezza AIC.

Il D-vine che mostra il migliore adattamento ai dati in termini di AIC, che considera anche la parsimonia del modello, è quello che possiede nel primo albero l'ordinamento di v.a. {RCA, RCG, ADB, IEN}. Al fine di scegliere il modello D-vine più adeguato, tuttavia, sono stati considerati come criteri valutativi, non solo i valori di AIC, ma anche i risultati dei test statistici di Vuong e Clarke che permettono di confrontare due modelli concorrenti tra loro. I modelli presi in considerazione, in particolare, sono i primi tre modelli D-vine in termini di AIC, individuati nel primo albero dagli ordinamenti {RCA, RCG, ADB, IEN}, {RCA, IEN, ADB, RCG} e {RCG, IEN, ADB, RCA}, le cui funzioni di densità sono fattorizzate come segue²:

$$\begin{aligned}
 f_{1,2,3,4}^I(u_1, u_2, u_3, u_4) = & c_{1,2}(F_1(u_1), F_2(u_2)) \cdot c_{2,4}(F_2(u_2), F_4(u_4)) \cdot c_{4,3}(F_4(u_4), F_3(u_3)) \\
 & \cdot c_{1,4|2}(F_{1|2}(u_1|u_2), F_{4|2}(u_4|u_2)) \cdot c_{2,3|4}(F_{2|4}(u_2|u_4), F_{3|4}(u_3|u_4)) \\
 & \cdot c_{1,3|2,4}(F_{1|2,4}(u_1|u_2, u_4), F_{3|2,4}(u_3|u_2, u_4))
 \end{aligned}$$

²Si ricorda che le variabili marginali sono considerate uniformi, $f_1(x_1) = f_2(x_2) = f_3(x_3) = f_4(x_4) = 1$.

$$\begin{aligned}
 f_{1,2,3,4}^{\text{II}}(u_1, u_2, u_3, u_4) = & c_{1,3}(F_1(u_1), F_3(u_3)) \cdot c_{3,4}(F_3(u_3), F_4(u_4)) \cdot c_{4,3}(F_4(u_4), F_3(u_3)) \\
 & \cdot c_{1,4|3}(F_{1|3}(u_1|u_3), F_{4|3}(u_4|u_3)) \cdot c_{3,2|4}(F_{3|4}(u_3|u_4), F_{2|4}(u_2|u_4)) \\
 & \cdot c_{1,2|3,4}(F_{1|3,4}(u_1|u_3, u_4), F_{2|3,4}(u_2|u_3, u_4))
 \end{aligned}$$

$$\begin{aligned}
 f_{1,2,3,4}^{\text{III}}(u_1, u_2, u_3, u_4) = & c_{2,3}(F_2(u_2), F_3(u_3)) \cdot c_{3,4}(F_3(u_3), F_4(u_4)) \cdot c_{4,1}(F_4(u_4), F_1(u_1)) \\
 & \cdot c_{2,4|3}(F_{2|3}(u_2|u_3), F_{4|3}(u_4|u_3)) \cdot c_{3,1|4}(F_{3|4}(u_3|u_4), F_{1|4}(u_1|u_4)) \\
 & \cdot c_{2,1|3,4}(F_{2|3,4}(u_2|u_3, u_4), F_{1|3,4}(u_1|u_3, u_4))
 \end{aligned}$$

Nelle tabelle seguenti per i tre D-vine considerati si riportano le copule bivariate selezionate e i relativi parametri stimati, individuati con $\hat{\theta}^{SP}$ e $\hat{\delta}^{SP}$ nel caso di copule caratterizzate da due parametri.

Ordine variabili 1° albero: {RCA, RCG, ADB, IEN}

Arco	Copula	$\hat{\theta}^{SP}$	$\hat{\delta}^{SP}$
$c_{1,2}$	Normale	0,220	/
$c_{2,4}$	BB8 ruotata 180°	2,148	0,557
$c_{4,3}$	BB8 ruotata 180°	1,943	0,618
$c_{1,4 2}$	Clayton ruotata 180°	0,128	/
$c_{2,3 4}$	Normale	0,198	/
$c_{1,3 2,4}$	Clayton	0,157	/
Log-verosimiglianza		899	
AIC		-1783	
BIC		-1726	

Tabella 4.7: Caratteristiche stimate del D-vine con ordine delle variabili nel primo albero pari a {RCA, RCG, ADB, IEN}.

Ordine variabili 1° albero: {RCA, IEN, ADB, RCG}

Arco	Copula	$\hat{\theta}^{SP}$	$\hat{\delta}^{SP}$
$c_{1,3}$	Clayton	0,220	/
$c_{3,4}$	BB8 ruotata 180°	1,979	0,598
$c_{4,3}$	BB8 ruotata 180°	2,154	0,561
$c_{1,4 3}$	Clayton ruotata 180°	0,137	/
$c_{3,2 4}$	Normale	0,194	/
$c_{1,2 3,4}$	Normale	0,181	/
Log-verosimiglianza		897	
AIC		-1779	
BIC		-1721	

Tabella 4.8: Caratteristiche stimate del D-vine con ordine delle variabili nel primo albero pari a {RCA, IEN, ADB, RCG}.

Ordine variabili 1° albero: {RCG, IEN, ADB, RCA}

Arco	Copula	$\hat{\theta}^{SP}$	$\hat{\delta}^{SP}$
$c_{2,3}$	Normale	0,220	/
$c_{3,4}$	BB8 ruotata 180°	1,788	0,677
$c_{4,1}$	Clayton ruotata 180°	0,148	/
$c_{2,4 3}$	BB8 ruotata 180°	2,302	0,467
$c_{3,1 4}$	BB8 ruotata 180°	1,545	0,771
$c_{2,1 3,4}$	Normale	0,181	/
Log-verosimiglianza		892	
AIC		-1767	
BIC		-1702	

Tabella 4.9: Caratteristiche stimate del D-vine con ordine delle variabili nel primo albero pari a {RCG, IEN, ADB, RCA}.

La selezione del modello più adeguato tra i tre individuati tramite criterio dell'AIC si basa, come detto, anche sui test statistici di Vuong e Clarke. Al fine di tenere in considerazione la differente dimensionalità dei modelli in esame, come mostrato in Vuong (1989) e Clarke (2007), sono state calcolate anche le statistiche test aggiustate coerentemente con le grandezze AIC e BIC, in cui i fattori di correzione risultano rispettivamente pari a $[p - q]$ e $[(p/2) \ln(H) - (q/2) \ln(H)]$, dove p e q rappresentano il numero dei parametri del primo e del secondo modello oggetto di test. La tabella seguente riporta gli esiti dei due test statistici valutati a un livello di confidenza pari ad $\alpha = 5\%$; l'ipotesi nulla H_0 prevede l'equivalenza tra i due modelli oggetto di analisi.

H0	Vuong			Clarke		
	Statistica test (AIC adj.) (BIC adj.)	P-value	Esito	Statistica test (AIC adj.) (BIC adj.)	P-value	Esito
$f_{1,2,3,4}^I$	0,4254	0,3353	non rifiuto H0	5054	0,1379	non rifiuto H0
\equiv	"	"	"	"	"	"
$f_{1,2,3,4}^{II}$	"	"	"	"	"	"
$f_{1,2,3,4}^I$	0,5722	0,2836	non rifiuto H0	5149	0,0014	$f_{1,2,3,4}^I$
\equiv	0,6537	0,2567	"	5156	0,0009	"
$f_{1,2,3,4}^{III}$	0,9474	0,1717	"	5184	0,0001	"
$f_{1,2,3,4}^{II}$	0,3527	0,3622	non rifiuto H0	4947	0,1469	non rifiuto H0
\equiv	0,4252	0,3353	"	4956	0,1922	"
$f_{1,2,3,4}^{III}$	0,6867	0,2461	"	4974	0,3050	"

Tabella 4.10: Test di Vuong e Clarke per i D-vine $f_{1,2,3,4}^I$, $f_{1,2,3,4}^{II}$ e $f_{1,2,3,4}^{III}$.

Gli esiti del test di Vuong per tutti e tre i confronti dei D-vine portano a non rifiutare l'ipotesi nulla di equivalenza tra i modelli statistici. Anche il test di

Clarke conferma l'equivalenza dei modelli a meno del D-vine $f_{1,2,3,4}^I$ che risulta superiore rispetto al modello $f_{1,2,3,4}^{III}$.

Stante gli esiti dei test di ipotesi di cui sopra e ricorrendo al criterio di Akaike, il modello selezionato è il D-vine $f_{1,2,3,4}^I$, dotato di ordinamento iniziale delle variabili nel primo albero pari a $\{RCA, RCG, ADB, IEN\}$.

Si noti che l'algoritmo di *Dißmann* per selezionare la struttura di D-vine porterebbe a scegliere la configurazione $\{RCA, RCG, IEN, ADB\}$, che risulta essere, tra le 12 strutture D-vine, la settima in termini di AIC.

Di seguito si riportano i dettagli delle analisi del processo di stima del D-vine selezionato. Come già anticipato, il processo di inferenza del modello di vine copula non prende in considerazione alcuna ipotesi modellistica sulle variabili marginali; il processo di stima dei parametri del modello si basa, quindi, su un approccio semiparametrico, dove le variabili di partenza del data set sintetico oggetto di studio, individuate con $\tilde{X}_1, \tilde{X}_2, \tilde{X}_3$ e \tilde{X}_4 , sono sostituite con le rispettive pseudo-osservazioni $\tilde{U}_1, \tilde{U}_2, \tilde{U}_3$ e \tilde{U}_4 , che coincidono con i valori delle funzioni di ripartizione empiriche delle singole variabili marginali. I parametri della vine copula sono inizialmente stimati, tramite verosimiglianza, seguendo l'approccio stepwise semiparametrico (SSP) in modo ricorsivo per tutti gli alberi dei livelli della struttura vine. In seguito, tali stime ($\hat{\theta}^{SSP}$) sono adottate come valori iniziali nel processo di massimizzazione della funzione di verosimiglianza complessiva del D-vine, ottenendo le stime dei parametri del modello ($\hat{\theta}^{SP}$).

Le quattro variabili marginali del data set oggetto di studio determinano, per il modello selezionato $f_{1,2,3,4}^I$, una struttura D-vine dotata di tre alberi, $\mathcal{V} = \{T_1, T_2, T_3\}$.

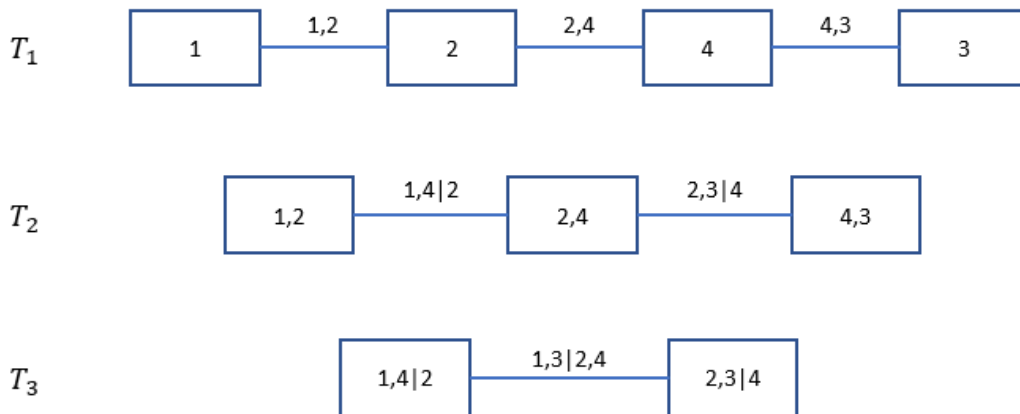


Figura 4.3: Struttura grafica del D-vine selezionato.

Il primo albero T_1 è caratterizzato dai nodi $N_1 = \{1, 2, 4, 3\}$ e dagli archi $E_1 = \{\{1, 2\}; \{2, 4\}; \{4, 3\}\}$. Prima di stimare la copule bivariate associate agli archi dell'albero, è stato implementato un test statistico volto a valutare l'indipendenza

dei dati bivariati coinvolti. La statistica test, costruita sotto l'ipotesi nulla di indipendenza dei dati, è la seguente (Genest e Favre (2007)):

$$\sqrt{\frac{9H(H-1)}{2(2H+5)}} \cdot |\hat{\tau}| \sim N(0, 1)$$

dove H è la numerosità campionaria e $\hat{\tau}$ il valore del τ di Kendall empirico calcolato sui dati bivariati oggetto di analisi. Considerando il primo arco dell'albero, $\{1, 2\}$, corrispondente alle variabili RCA e RCG, il τ di Kendall è pari a 0,132 e il test di indipendenza, ad un livello di confidenza $\alpha = 5\%$, porta a rifiutare l'ipotesi H_0 :

Test di indipendenza

valori soglia:

$$\Phi^{-1}(\alpha/2) = -1,960,$$

$$\Phi^{-1}(1 - \alpha/2) = 1,960$$

H0	Statistica test	P-value	Esito
$\tilde{U}_1 \perp \tilde{U}_2$	19,6859	0,0000	rifiuto H0

Tabella 4.11: Test di indipendenza tra le variabili \tilde{U}_1 e \tilde{U}_2 .

I grafici mostrati nel seguito, relativi al *K-plot* e *Chi-plot*, permettono di indagare la dipendenza presente nei dati oggetto di analisi e palesano anch'essi una relazione positiva tra le variabili \tilde{U}_1 e \tilde{U}_2 ; in tale circostanza, i punti del *K-plot* sono disposti al di sopra della diagonale principale, mentre nel *Chi-plot* i valori della grandezza χ risultano maggiori di zero.

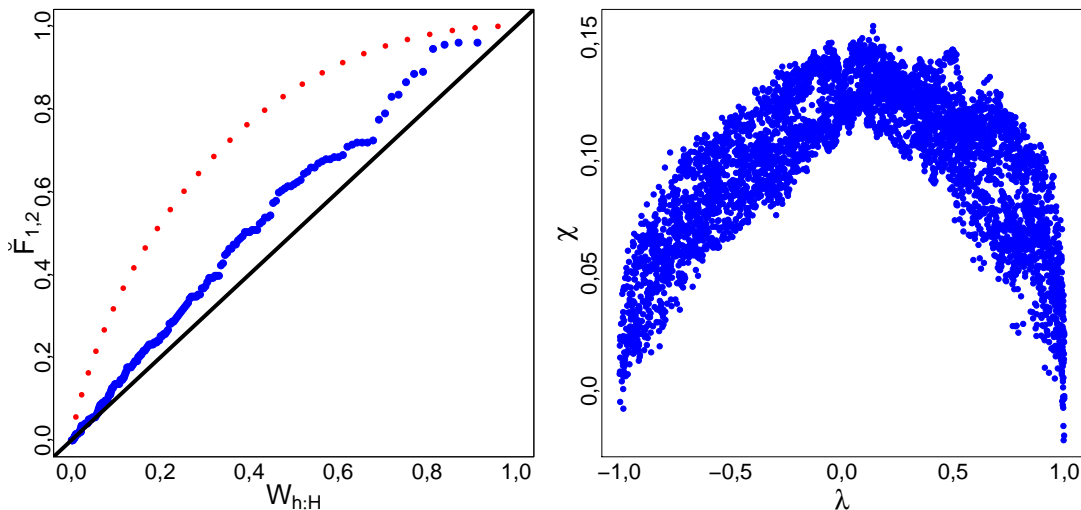


Figura 4.4: *K-plot* e *Chi-plot* per le variabili \tilde{U}_1 e \tilde{U}_2 .

La copula selezionata tra il set di famiglie considerate è scelta in base al criterio AIC; i parametri della copula, inoltre, sono stimati tramite metodo della

massima verosimiglianza. Per l'arco $\{1, 2\}$, la procedura di stima porta a identificare la copula Normale, di parametro, stimato tramite approccio stepwise, pari a $\hat{\theta}^{SSP} = \hat{\rho}^{SSP} = 0,211$, a cui corrisponde un τ teorico di 0,136. Il grafico denominato *Lambda-plot*, confrontando la funzione $\lambda(t)$ empirica con quella teorica del modello selezionato, permette di valutare graficamente l'adattamento dei dati alla copula stimata.

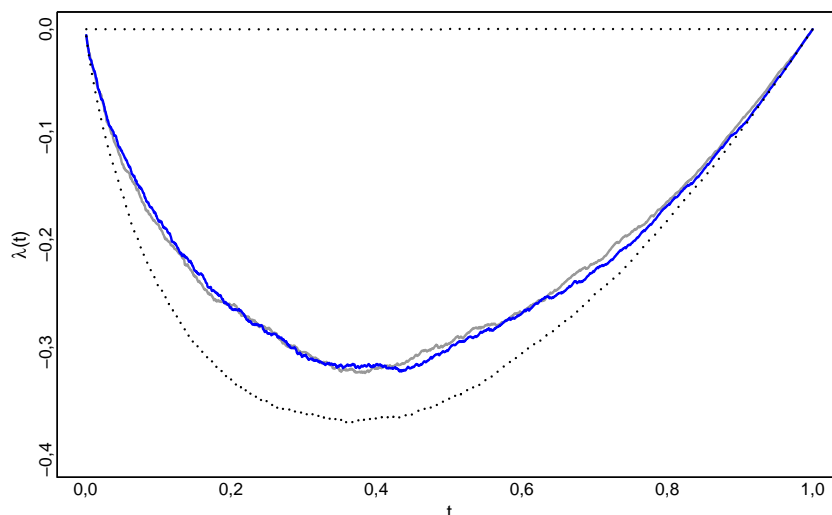


Figura 4.5: *Lambda-plot* delle variabili \tilde{U}_1 e \tilde{U}_2 (curva di colore grigio) e della copula stimata Normale di parametro $\hat{\theta}^{SSP} = \hat{\rho}^{SSP} = 0,211$ (curva di colore blu).

Il secondo arco dell'albero T_1 , dato dalle variabili $\{2,4\}$, corrispondenti a RCG e ADB, mostra una dipendenza empirica misurata tramite τ di Kendall pari a 0,113. Il test di indipendenza porta a concludere che le due v.a. sono tra loro dipendenti:

Test di indipendenza

valori soglia:

$$\Phi^{-1}(\alpha/2) = -1,960,$$

$$\Phi^{-1}(1 - \alpha/2) = 1,960$$

H0	Statistica test	P-value	Esito
$\tilde{U}_2 \perp \tilde{U}_4$	16,9924	0,0000	rifuto H0

Tabella 4.12: Test di indipendenza tra le variabili \tilde{U}_2 e \tilde{U}_4 .

Gli strumenti grafici del *K-plot* e *Chi-plot* mostrano la presenza di dipendenza positiva tra le variabili in esame.

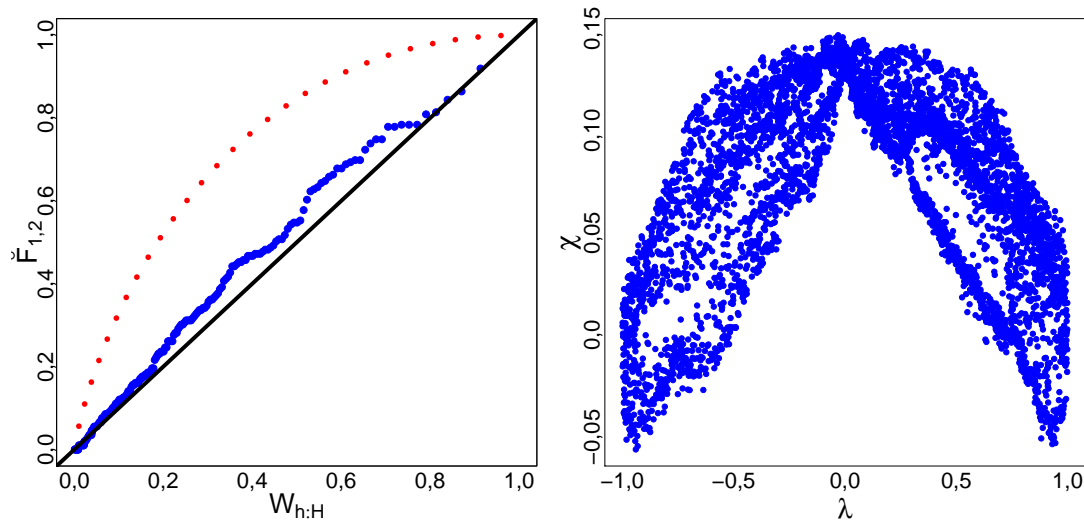


Figura 4.6: *K-plot* e *Chi-plot* per le variabili \tilde{U}_2 e \tilde{U}_4 .

La copula selezionata associata alle variabili \tilde{U}_2 e \tilde{U}_4 è la copula BB8 ruotata di 180° , con parametri stimati pari a $\hat{\theta}^{SSP} = 1,996$ e $\hat{\delta}^{SSP} = 0,597$, a cui corrisponde un τ teorico di 0,114. Tale copula non possiede dipendenza di coda né superiore né inferiore. Di seguito si riporta il *Lambda-plot* della copula stimata.

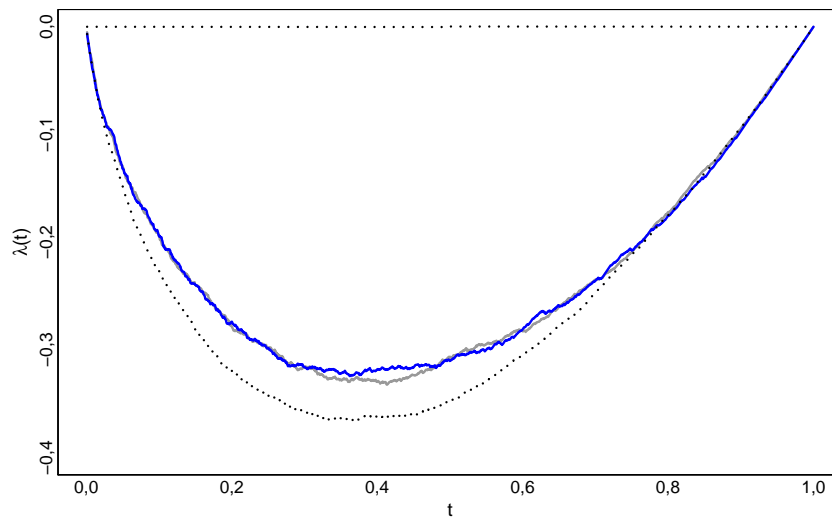


Figura 4.7: *Lambda-plot* delle variabili \tilde{U}_2 e \tilde{U}_4 (curva di colore grigio) e della copula stimata BB8 ruotata di 180° di parametri $\hat{\theta}^{SSP} = 1,996$ e $\hat{\delta}^{SSP} = 0,597$ (curva di colore blu).

Il grafico precedente, a riprova dell'adeguata selezione della copula in esame, mostra un buon accostamento tra la funzione $\lambda(t)$ empirica, calcolata sul vettore bidimensionale (u_2, u_4) e la funzione $\lambda(t)$ teorica della copula selezionata.

Il terzo nodo del primo albero, $\{4,3\}$, associato alle variabili ADB e IEN, possiede un valore di τ pari a 0,112. Il test di indipendenza, costruito sul valore del τ di Kendall empirico, porta a rifiutare l'ipotesi nulla di indipendenza tra i due vettori marginali \tilde{U}_4 e \tilde{U}_3 .

Test di indipendenza

valori soglia:

$$\Phi^{-1}(\alpha/2) = -1,960,$$

$$\Phi^{-1}(1 - \alpha/2) = 1,960$$

H0	Statistica test	P-value	Esito
$\tilde{U}_4 \perp \tilde{U}_3$	16,8093	0,0000	rifiuto H0

Tabella 4.13: Test di indipendenza tra le variabili \tilde{U}_4 e \tilde{U}_3 .

I grafici del *K-plot* e *Chi-plot*, mostrati nel seguito, evidenziano la situazione di dipendenza positiva tra le variabili marginali \tilde{U}_4 e \tilde{U}_3 ; per il *K-plot*, infatti, i punti risiedono sopra la diagonale principale, mentre nel *Chi-plot* i valori di χ sono concentrati sopra zero.

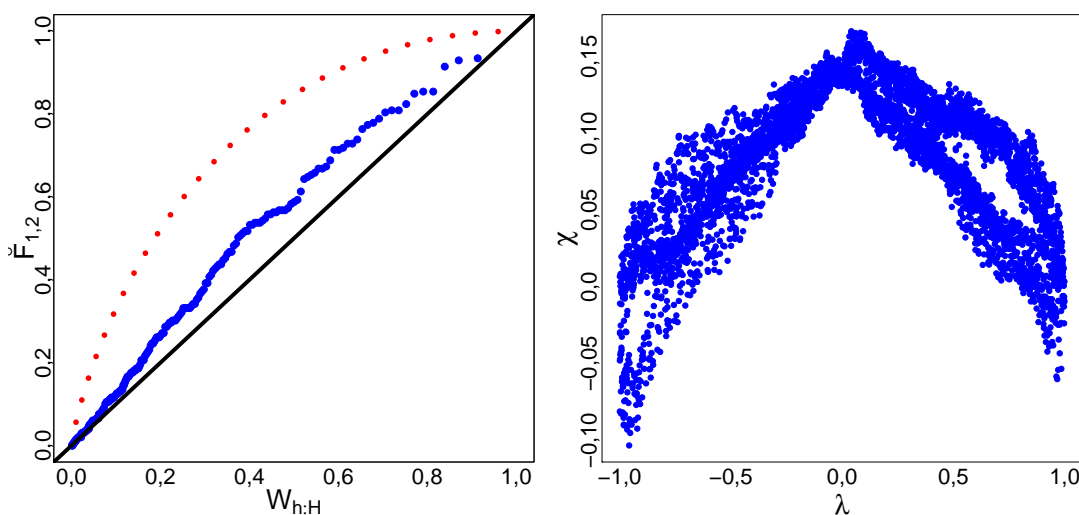


Figura 4.8: *K-plot* e *Chi-plot* per le variabili \tilde{U}_4 e \tilde{U}_3 .

La copula bivariata selezionata per le variabili \tilde{U}_4 e \tilde{U}_3 risulta essere la copula BB8 ruotata di 180° con parametri stimati pari a $\hat{\theta}^{SSP} = 1,793$ e $\hat{\delta}^{SSP} = 0,673$, che possiede un valore teorico di τ pari a 0,113. Di seguito si riporta il rispettivo *Lambda-plot*.

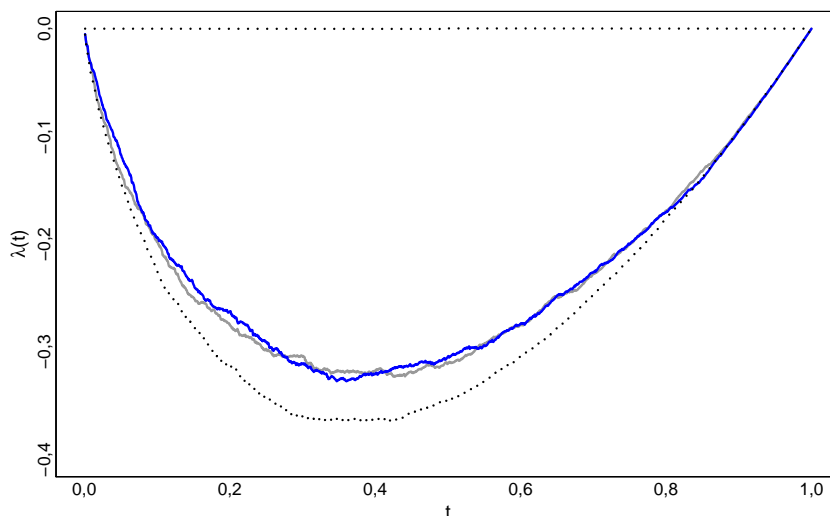


Figura 4.9: *Lambda-plot* delle variabili \tilde{U}_4 e \tilde{U}_3 (curva di colore grigio) e della copula stimata BB8 ruotata di 180° di parametri $\hat{\theta}^{SSP} = 1,793$ e $\hat{\delta}^{SSP} = 0,673$ (curva di colore blu).

Il secondo albero della struttura D-vine possiede nodi, N_2 , coincidenti con gli archi dell'albero precedente e archi, esplicitandone gli insiemi condizionati e condizionanti, pari a $E_2 = \{\{1, 4|2\}; \{2, 3|4\}\}$. Le copule bivariate coinvolte in tale albero sono $c_{1,4|2}(F_{1|2}(u_1|u_2), F_{4|2}(u_4|u_2))$ e $c_{2,3|4}(F_{2|4}(u_2|u_4), F_{3|4}(u_3|u_4))$. Ricorrendo alla relazione (3.1.9) è possibile determinare gli argomenti delle copule bivariate coinvolte, dati dalle funzioni di ripartizione condizionate. Considerando la copula $c_{1,4|2}(F_{1|2}(u_1|u_2), F_{4|2}(u_4|u_2))$, gli argomenti $F_{1|2}$ e $F_{4|2}$ sono determinati come:

$$F_{1|2}(u_1|u_2) = \frac{\partial C_{1,2}(F_1(u_1), F_2(u_2))}{\partial F_2(u_2)}$$

$$F_{4|2}(u_4|u_2) = \frac{\partial C_{4,2}(F_4(u_4), F_2(u_2))}{\partial F_2(u_2)}$$

In merito alla copula $c_{2,3|4}(F_{2|4}(u_2|u_4), F_{3|4}(u_3|u_4))$, le funzioni di ripartizione condizionate $F_{2|4}$ e $F_{3|4}$ sono ottenute come,

$$F_{2|4}(u_2|u_4) = \frac{\partial C_{2,4}(F_2(u_2), F_4(u_4))}{\partial F_4(u_4)}$$

$$F_{3|4}(u_3|u_4) = \frac{\partial C_{3,4}(F_3(u_3), F_4(u_4))}{\partial F_4(u_4)}$$

Si noti che, avendo selezionato solo famiglie di copule che godono della proprietà dell'interscambiabilità, valgono le relazioni $C_{4,2} = C_{2,4}$ e $C_{3,4} = C_{4,3}$.

Considerando il primo arco, dato da $\{1, 4|2\}$, associato alle Lob RCA e ADB condizionate al ramo RCG, la variabile bivariata coinvolta è costituita dalla coppia

($\tilde{U}_{1|2} = F_{1|2}, \tilde{U}_{4|2} = F_{4|2}$). Il τ di Kendall empirico calcolato su tali variabili è pari a 0,051 e il test di indipendenza porta a rifiutare l'ipotesi di indipendenza tra le variabili coinvolte.

Test di indipendenza

valori soglia:

$$\Phi^{-1}(\alpha/2) = -1,960,$$

$$\Phi^{-1}(1 - \alpha/2) = 1,960$$

H0	Statistica test	P-value	Esito
$\tilde{U}_{1 2} \perp \tilde{U}_{4 2}$	7,6804	0,0000	rifiuto H0

Tabella 4.14: Test di indipendenza tra le variabili $\tilde{U}_{1|2}$ e $\tilde{U}_{4|2}$.

I grafici seguenti del *K-plot* e *Chi-plot* evidenziano il basso grado di dipendenza positiva tra le variabili coinvolte.

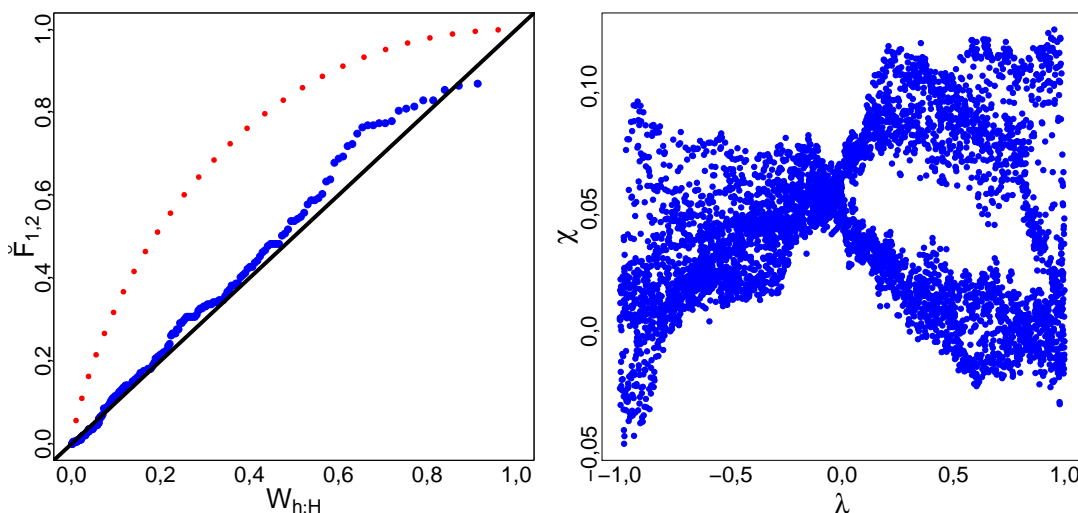


Figura 4.10: *K-plot* e *Chi-plot* per le variabili $\tilde{U}_{1|2}$ e $\tilde{U}_{4|2}$.

La copula bivariata selezionata è la copula Clayton ruotata di 180° di parametro stimato $\hat{\theta}^{SSP} = 0,126$. Il τ teorico della copula, a fronte del parametro stimato, è pari a $\tau = 0,059$. Tale copula possiede dipendenza di coda superiore pari a $\lambda_U = 2^{-\frac{1}{\theta}}$; stante il valore stimato del parametro, il coefficiente di dipendenza di coda superiore assume un valore contenuto, pari a $\lambda_U = 0,004$. Di seguito si riporta il *Lambda-plot*.

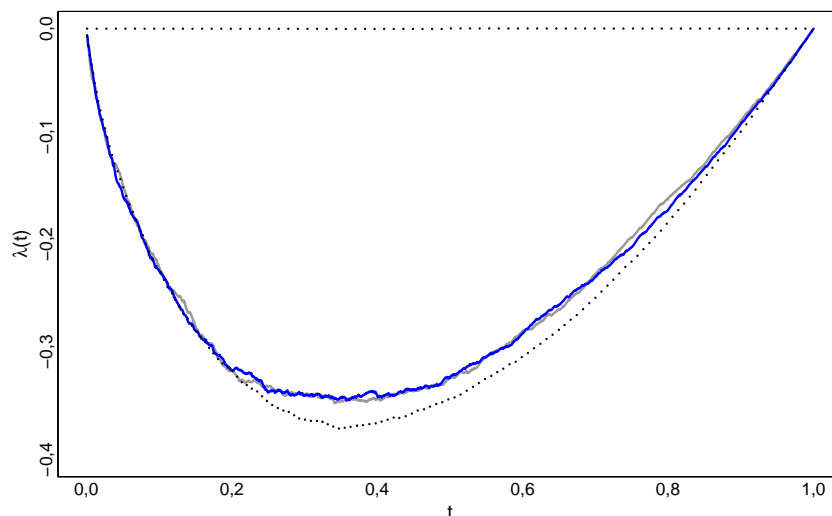


Figura 4.11: *Lambda-plot* delle variabili $\tilde{U}_{1|2}$ e $\tilde{U}_{4|2}$ (curva di colore grigio) e della copula stimata Clayton ruotata di 180° di parametro $\hat{\theta}^{SSP} = 0,126$ (curva di colore blu).

Nel presente caso studio il processo di stima del D-vine è costruito sotto l'assunzione di ipotesi semplificata della pair-copula construction; di seguito, tuttavia, si riporta il risultato del test statistico proposto da Kurz e Spanhel (2017), denominato *constant conditional correlations test* (CCC), sviluppato per valutare la validità dell'ipotesi semplificata. La statistica test costruita sotto l'ipotesi H_0 di indipendenza risulta essere una variabile χ^2 con un grado di libertà. La tabella seguente riporta i risultati del test associati alla copula $c_{1,4|2}$.

Test CCC

valori soglia:

$$F_{\chi^2}^{-1}(\alpha/2) = 0,001,$$

$$F_{\chi^2}^{-1}(1 - \alpha/2) = 5,024$$

H_0	Statistica test	P-value	Esito
$(\tilde{U}_{1 2}, \tilde{U}_{4 2}) \perp \tilde{U}_2$	13,0878	0,0003	rifiuto H_0

Tabella 4.15: Test di indipendenza tra la variabile $(\tilde{U}_{1|2}, \tilde{U}_{4|2})$ e la variabile \tilde{U}_2 .

L'esito del test mostra che il vettore bidimensionale $(\tilde{U}_{1|2}, \tilde{U}_{4|2})$ non è indipendente dalla variabile \tilde{U}_2 . Senza ulteriori informazioni, quindi, l'analisi porterebbe a concludere la violazione dell'ipotesi semplificata. Tuttavia, sfruttando i risultati teorici presenti in Stöber, Joe e Czado (2013) è possibile valutare in modo più approfondito la validità o meno dell'ipotesi semplificata sottostante alla PCC. Il test statistico CCC, infatti, permette solamente di valutare la dipendenza o meno di un vettore aleatorio bidimensionale rispetto a una o più v.a. condizionanti. Nella situazione in cui il test porti a rifiutare l'ipotesi di indipendenza, tuttavia, non è

possibile affermare che sia violata l'ipotesi semplificata; a tale scopo, come mostrato in Stöber, Joe e Czado (2013), è necessario valutare se la copula sottostante alle variabili condizionate e condizionanti appartenga alle famiglie Clayton, Normale o t-student, che sono le uniche tre famiglie di copule multivariate che soddisfano l'ipotesi semplificata. Si noti che le copule bivariate condizionate ottenute da tali famiglie appartengono a loro volta alla famiglia della copula multidimensionale non condizionata³; nel contesto in esame, quindi la selezione per la copula $c_{1,4|2}$ della famiglia Clayton ruotata di 180° implica che la copula trivariata sottostante alle variabili condizionate e alla variabile condizionante non appartiene ad una delle tre famiglie che soddisfano l'ipotesi semplificata. A fronte di tale considerazione, tuttavia, si è proceduto a stimare la copula $c(F_{1|2}(u_1|u_2), F_{4|2}(u_4|u_2), u_2)$; la selezione della copula trivariata ha preso in considerazione le copule in $n = 3$ dimensioni appartenenti alle famiglie Normale, t-student, Clayton, Joe, Frank e Gumbel. La copula che mostra il migliore adattamento risulta essere la t-student. Tale risultato, quindi, palesa la difficoltà di stimare un adeguato modello di dipendenza in contesti multivariati ($n > 2$); come noto, infatti, il modello ignoto di dipendenza dato dalla copula $c(F_{1|2}(u_1|u_2), F_{4|2}(u_4|u_2), u_2)$ non è una t-student, tuttavia, la non adeguata flessibilità delle famiglie di copule multivariate porta a selezionare come modello più adeguato la copula t-student. A fronte di tale analisi, quindi, è possibile concludere che l'ipotesi semplificata sottostante alla PCC è, verosimilmente, non rispettata; la selezione della copula non condizionata trivariata t-student, tuttavia, porta a pensare che la violazione dell'ipotesi semplificata non sia particolarmente marcata. In definitiva, quindi, la copula non condizionata è un'approssimazione della copula condizionata:

$$c_{1,4|2}(F_{1|2}(u_1|u_2), F_{4|2}(u_4|u_2)) \approx c_{1,4|2}(F_{1|2}(u_1|u_2), F_{4|2}(u_4|u_2)|u_2)$$

In merito al secondo nodo, $\{2, 3|4\}$, concernente le linee di business RCG e IEN condizionate alla Lob ADB, le variabili coinvolte sono $\tilde{U}_{2|4} = F_{2|4}$ e $\tilde{U}_{3|4} = F_{3|4}$ che mostrano un grado di dipendenza espresso in termini di τ di Kendall pari a 0,128. Il test di indipendenza conferma che le due v.a. coinvolte sono tra loro statisticamente dipendenti.

Test di indipendenza

valori soglia:

$$\Phi^{-1}(\alpha/2) = -1,960,$$

$$\Phi^{-1}(1 - \alpha/2) = 1,960$$

H0	Statistica test	P-value	Esito
$\tilde{U}_{2 4} \perp \tilde{U}_{3 4}$	19,1557	0,0000	rifuto H0

Tabella 4.16: Test di indipendenza tra le variabili $\tilde{U}_{2|4}$ e $\tilde{U}_{3|4}$.

³Si veda il paragrafo 3.1 per un esempio relativo alla famiglia Clayton e l'Appendice B di Stöber, Joe e Czado (2013) per le copule Normale e t-student.

I grafici del *K-plot* e *Chi-plot* mostrano anch'essi la presenza di dipendenza positiva tra le variabili coinvolte.

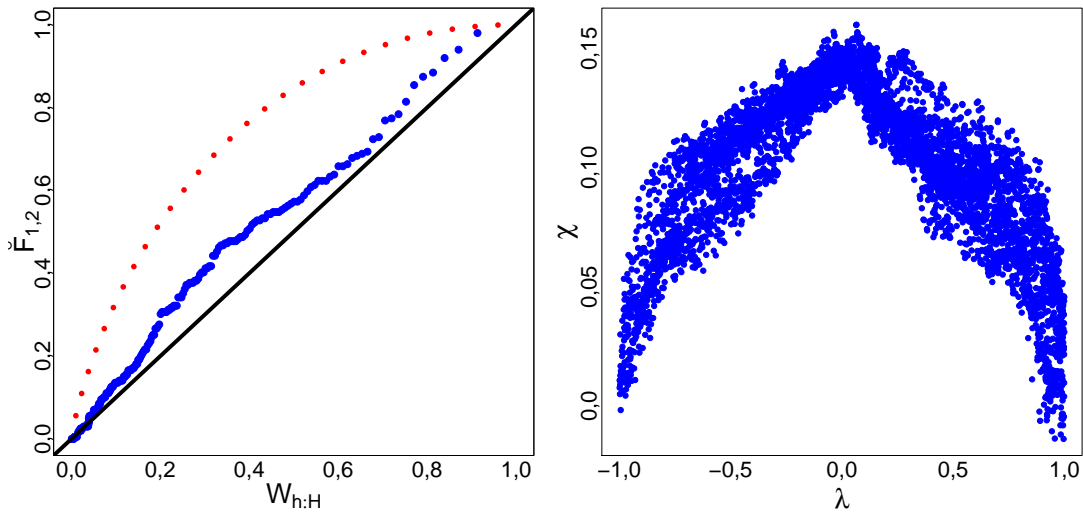


Figura 4.12: *K-plot* e *Chi-plot* per le variabili $\tilde{U}_{2|4}$ e $\tilde{U}_{3|4}$.

La copula stimata per tale arco risulta essere la copula Normale di parametro $\hat{\theta}^{SSP} = \hat{\rho}^{SSP} = 0,197$, a cui è associato un τ teorico di 0,126. Di seguito si riporta il grafico *Lambda*.

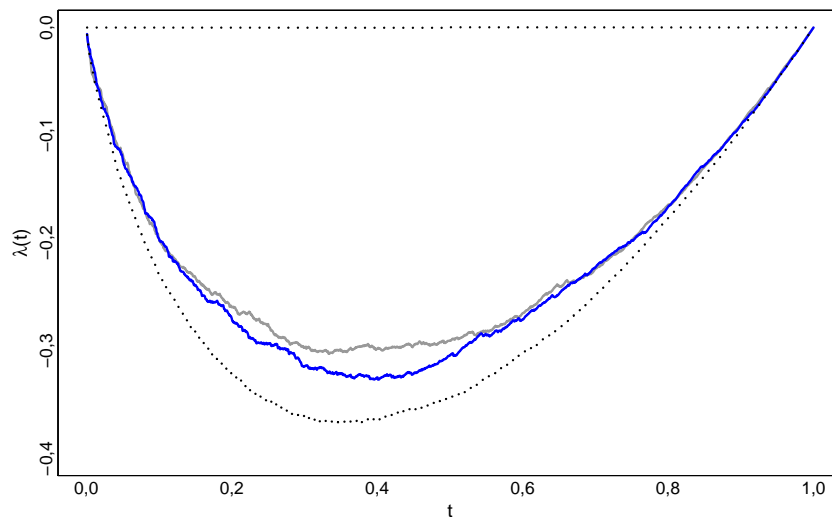


Figura 4.13: *Lambda-plot* delle variabili $\tilde{U}_{2|4}$ e $\tilde{U}_{3|4}$ (curva di colore grigio) e della copula stimata Normale di parametro $\hat{\theta}^{SSP} = \hat{\rho}^{SSP} = 0,197$ (curva di colore blu).

Al fine di valutare la validità dell'ipotesi semplificata della pair-copula construction, è stato implementato il test statistico CCC; la tabella seguente ne riporta l'esito.

Test CCC
valori soglia:
 $F_{\chi^2}^{-1}(\alpha/2) = 0,001,$
 $F_{\chi^2}^{-1}(1 - \alpha/2) = 5,024$

H0	Statistica test	P-value	Esito
$(\tilde{U}_{2 4}, \tilde{U}_{3 4}) \perp \tilde{U}_4$	0,9100	0,3401	non rifiuto H0

Tabella 4.17: Test di indipendenza tra la variabile $(\tilde{U}_{2|4}, \tilde{U}_{3|4})$ e la variabile \tilde{U}_4 .

Il risultato del test porta a non rifiutare l'ipotesi nulla di indipendenza tra la v.a. $(\tilde{U}_{2|4}, \tilde{U}_{3|4})$ e la variabile \tilde{U}_4 . E' quindi possibile concludere che la copula $c_{2,3|4}$ è nella forma semplificata:

$$c_{2,3|4}(F_{2|4}(u_2|u_4), F_{3|4}(u_3|u_4)) = c_{2,3|4}(F_{2|4}(u_2|u_4), F_{3|4}(u_3|u_4)|u_4)$$

L'ultimo albero della struttura D-vine, T_3 , possiede nodi N_3 coincidenti con gli archi dell'albero precedente e un solo arco che, esplicitandone l'insieme condizionato e condizionante delle variabili, risulta essere $\{1, 3|2, 4\}$, associato ai rami RCA e IEN condizionati rispetto alle Lob RCG e ADB. La copula bivariata coinvolta è $c_{1,3|2,4}(F_{1|2,4}(u_1|u_2, u_4), F_{3|2,4}(u_3|u_2, u_4))$. Per potere stimare la copula $c_{1,3|2,4}$ è necessario calcolarne gli argomenti. In modo ricorsivo, ricorrendo alla relazione (3.1.9), si ottiene:

$$F_{1|2,4}(u_1|u_2, u_4) = \frac{\partial C_{1,4|2}(F_{1|2}(u_1|u_2), (F_{4|2}(u_4|u_2)))}{\partial F_{4|2}(u_4|u_2)}$$

$$F_{3|2,4}(u_3|u_2, u_4) = \frac{\partial C_{3,2|4}(F_{3|4}(u_3|u_4), (F_{2|4}(u_2|u_4)))}{\partial F_{2|4}(u_2|u_4)}$$

Si noti che data la proprietà dell'interscambiabilità delle copule vale che $C_{3,2|4} = C_{2,3|4}$. La dipendenza tra le variabili coinvolte, $\tilde{U}_{1|2,4} = F_{1|2,4}(u_1|u_2, u_4)$ e $\tilde{U}_{3|2,4} = F_{3|2,4}(u_3|u_2, u_4)$, misurata tramite τ di Kendall è pari a 0,022. Nonostante il valore contenuto di τ , il test di indipendenza rifiuta l'ipotesi nulla.

Test di indipendenza
valori soglia:
 $\Phi^{-1}(1 - \alpha/2) = 1,960,$
 $\Phi^{-1}(\alpha/2) = -1,960$

H0	Statistica test	P-value	Esito
$\tilde{U}_{1 2,4} \perp \tilde{U}_{3 2,4}$	3,2509	0,0012	rifiuto H0

Tabella 4.18: Test di indipendenza tra le variabili $\tilde{U}_{1|2,4}$ e $\tilde{U}_{3|2,4}$.

I grafici del *K-plot* e del *Chi-plot*, riportati nel seguito, evidenziano una situazione di quasi indipendenza tra le variabili coinvolte.

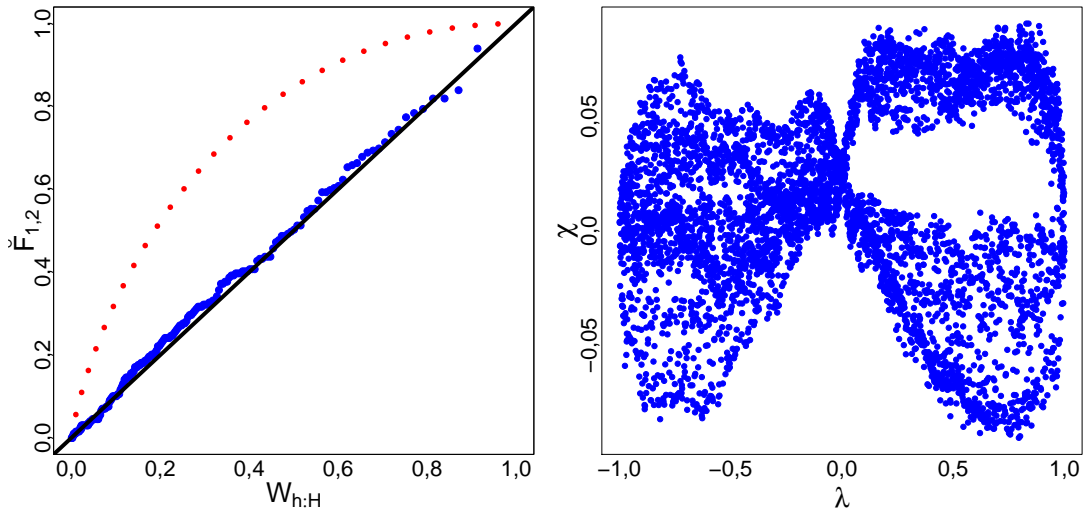


Figura 4.14: *K-plot* e *Chi-plot* per le variabili $\tilde{U}_{1|2,4}$ e $\tilde{U}_{3|2,4}$.

La stima della copula bivariata porta a selezionare la copula Clayton di parametro $\hat{\theta}^{SSP} = 0,090$, a cui corrisponde un τ teorico di 0,043. La dipendenza di coda inferiore, di contro, è quasi prossima a zero, $\lambda_L = 0,001$. Di seguito si riporta il *Lambda-plot* che permette di valutare la prossimità tra la funzione $\lambda(t)$ empirica con quella teorica della copula stimata.

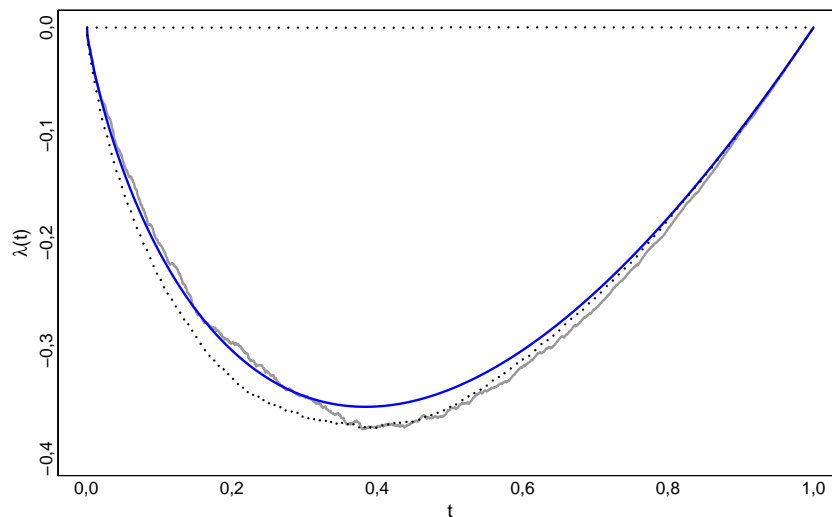


Figura 4.15: *Lambda-plot* delle variabili $\tilde{U}_{1|2,4}$ e $\tilde{U}_{3|2,4}$ (curva di colore grigio) e della copula stimata Clayton $\hat{\theta}^{SSP} = 0,090$ (curva di colore blu).

Infine, ricorrendo al constant conditional correlations test è possibile procedere ad analizzare l'adeguatezza dell'ipotesi semplificata della PCC. La tabella seguente mostra l'esito del test che porta a rigettare l'ipotesi di indipendenza tra le variabili.

Test CCC
valori soglia:
 $F_{\chi^2}^{-1}(\alpha/2) = 0,001,$
 $F_{\chi^2}^{-1}(1 - \alpha/2) = 5,024$

H0	Statistica test	P-value	Esito
$(\tilde{U}_{1 2,4}, \tilde{U}_{3 2,4}) \perp (\tilde{U}_2, \tilde{U}_4)$	32,7675	0,0000	rifiuto H0

Tabella 4.19: Test di indipendenza tra le variabili $(\tilde{U}_{1|2,4}, \tilde{U}_{3|2,4})$ e $(\tilde{U}_2, \tilde{U}_4)$.

Al fine di valutare la validità o meno dell'ipotesi semplificata, seguendo i risultati teorici mostrati in Stöber, Joe e Czado (2013), si è stimata la copula sottostante alle variabili condizionate e condizionanti. La procedura di stima porta a selezionare per la copula $c(F_{1|2,4}(u_1|u_2, u_4), F_{3|2,4}(u_3|u_2, u_4), u_2, u_4)$ la famiglia t-student. Come noto, tuttavia, la copula condizionata bivariata ottenuta da una copula multivariata t-student appartiene a sua volta alla stessa famiglia; la selezione della famiglia Clayton per la copula $c_{1,3|2,4}$ e della famiglia t-student per la copula multivariata palesa la difficoltà di stimare un adeguato modello di dipendenza in contesti multivariati ricorrendo alle famiglie di copule multivariate esistenti in letteratura. A fronte dei risultati ottenuti è lecito concludere che l'ipotesi semplificata della PCC non è rispettata; tuttavia, la selezione della famiglia t-student per la copula in $n = 4$ dimensioni porta a ritenere l'adozione dell'ipotesi semplificata come una buona approssimazione della copula non condizionata. In definitiva, quindi, la copula $c_{1,3|2,4}$ è un'approssimazione di quella nella forma non semplificata:

$$c_{1,3|2,4}(F_{1|2,4}(u_1|u_2, u_4), F_{3|2,4}(u_3|u_2, u_4)) \approx c_{1,3|2,4}(F_{1|2,4}(u_1|u_2, u_4), F_{3|2,4}(u_3|u_2, u_4)|u_2, u_4)$$

Dopo avere selezionato le famiglie di copule bivariate coinvolte e averne stimato i parametri tramite approccio stepwise semiparametrico (SSP), questi ultimi sono stati adottati come valori di partenza per la massimizzazione della funzione di verosimiglianza complessiva del D-vine. Tale operazione permette di stimare i parametri del modello tramite approccio semiparametrico (SP). La tabella seguente riassume le stime dei parametri ottenuti sia tramite approccio SSP che ricorrendo al metodo SP.

Arco	Copula	$\hat{\theta}^{SSP}$	$\hat{\delta}^{SSP}$	$\hat{\theta}^{SP}$	$\hat{\delta}^{SP}$
$c_{1,2}$	Normale	0,211	/	0,220	/
$c_{2,4}$	BB8 ruotata 180°	1,996	0,597	2,148	0,557
$c_{4,3}$	BB8 ruotata 180°	1,793	0,673	1,943	0,618
$c_{1,4 2}$	Clayton ruotata 180°	0,126	/	0,128	/
$c_{2,3 4}$	Normale	0,197	/	0,198	/
$c_{1,3 2,4}$	Clayton	0,090	/	0,157	/

Tabella 4.20: Parametri stimati tramite approccio SSP e SP per il D-vine selezionato.

Il valore di log-verosimiglianza del modello è pari a 899, mentre i valori di AIC e BIC sono rispettivamente di -1783 e -1726. Si noti che i valori stimati dei parametri tramite approccio SP non differiscono significativamente dai valori ottenuti tramite approccio SSP; ciò è prova del fatto che le famiglie di copule bivariate del modello sono adeguate rispetto al data set sottostante (Haff (2013)).

Infine, si è confrontata la bontà di adattamento ai dati del D-vine stimato con le copule multivariate Normale e t-student, che in letteratura costituiscono un punto di riferimento nello studio della dipendenza multivariata. La tabella seguente riporta i valori della log-verosimiglianza, di AIC e di BIC per le tre copule multivariate considerate.

Copula	Log-veros.	AIC	BIC
D-vine	899	-1783	-1726
C^{Ga}	784	-1485	-1442
C^t	762	-1511	-1461

Tabella 4.21: Confronto tra il D-vine selezionato e le copule Normale e t-student multivariate.

Come si evince dalla tabella precedente, la vine copula definita dal D-vine selezionato garantisce un adattamento migliore alla struttura di dipendenza multivariata sottostante ai dati oggetto di analisi.

Si noti, infine, che il modello D-vine è stato confrontato anche con le copule archimedee multivariate Clayton, Joe, Frank e Gumbel; l'adattamento di tali famiglie di copule ai dati sottostanti, tuttavia, è risultato particolarmente scadente.

4.2.2 Stima del C-vine

Analogamente a quanto svolto per i D-vine, di seguito si riportano le analisi sviluppate per selezionare la struttura C-vine che meglio modella le relazioni di dipendenza tra i dati oggetto di studio.

La tabella seguente riporta i valori di log-verosimiglianza, AIC e BIC di tutte e 12 le possibili configurazioni di C-vine esistenti considerando quattro variabili aleatorie. La prima colonna della tabella contiene i nodi dei tre alberi della struttura C-vine considerata.

#	Nodi degli alberi T_1, T_2 e T_3	AIC	BIC	Log-veros.
1	{RCG; {RCG, IEN}; {IEN, RCA RCG}}	-1772	-1714	894
2	{RCG; {RCG, RCA}; {RCA, IEN RCG}}	-1769	-1711	892
3	{IEN; {IEN, RCA}; {RCA, RCG IEN}}	-1764	-1699	891
4	{IEN; {IEN, RCA}; {RCA, RCG IEN}}	-1762	-1697	890
5	{RCA; {RCA, RCG}; {RCG, IEN RCA}}	-1711	-1653	863
6	{IEN; {IEN, ADB}; {ADB, RCA IEN}}	-1655	-1590	836
7	{RCG; {RCG, ADB}; {ADB, RCA RCG}}	-1644	-1586	830
8	{IEN; {IEN, RCG}; {RCG, RCA IEN}}	-1568	-1503	793
9	{ADB; {ADB, RCG}; {RCG, RCA ADB}}	-1478	-1420	747
10	{RCA; {RCA, IEN}; {IEN, RCG RCA}}	-1461	-1403	738
11	{ADB; {ADB, IEN}; {IEN, RCA ADB}}	-1454	-1382	737
12	{RCA; {RCA, ADB}; {ADB, RCG RCA}}	-1242	-1185	629

Tabella 4.22: Strutture C-vine ordinate per valori crescenti della grandezza AIC.

Il C-vine che mostra il migliore valore di AIC è quello definito, nei vari alberi, dai seguenti nodi {RCG; {RCG, IEN}; {IEN, RCA | RCG}}. Al fine di selezionare il modello migliore, sono stati considerati anche i risultati dei test statistici di Vuong e Clarke. Tali test sono stati utilizzati per confrontare i primi tre C-vine in termini di valori di AIC, le cui funzioni di densità sono riportate nel seguito:

$$\begin{aligned}
f_{1,2,3,4}^I(u_1, u_2, u_3, u_4) = & c_{2,1}(F_2(u_2), F_1(u_1)) \cdot c_{2,3}(F_2(u_2), F_3(u_3)) \cdot c_{2,4}(F_2(u_2), F_4(u_4)) \\
& \cdot c_{3,1|2}(F_{3|2}(u_3|u_2), F_{1|2}(u_1|u_2)) \cdot c_{3,4|2}(F_{3|2}(u_3|u_2), F_{4|2}(u_4|u_2)) \\
& \cdot c_{1,4|3,2}(F_{1|3,2}(u_1|u_3, u_2), F_{4|3,2}(u_4|u_3, u_2))
\end{aligned}$$

$$\begin{aligned}
f_{1,2,3,4}^{II}(u_1, u_2, u_3, u_4) = & c_{2,1}(F_2(u_2), F_1(u_1)) \cdot c_{2,3}(F_2(u_2), F_3(u_3)) \cdot c_{2,4}(F_2(u_2), F_4(u_4)) \\
& \cdot c_{1,3|2}(F_{1|2}(u_1|u_2), F_{3|2}(u_3|u_2)) \cdot c_{1,4|2}(F_{1|2}(u_1|u_2), F_{4|2}(u_4|u_2)) \\
& \cdot c_{3,4|1,2}(F_{3|1,2}(u_3|u_1, u_2), F_{4|1,2}(u_4|u_1, u_2))
\end{aligned}$$

$$\begin{aligned}
f_{1,2,3,4}^{III}(u_1, u_2, u_3, u_4) = & c_{4,1}(F_4(u_4), F_1(u_1)) \cdot c_{4,2}(F_4(u_4), F_2(u_2)) \cdot c_{4,3}(F_4(u_4), F_3(u_3)) \\
& \cdot c_{1,2|4}(F_{1|4}(u_1|u_4), F_{2|4}(u_2|u_4)) \cdot c_{1,3|4}(F_{1|4}(u_1|u_4), F_{3|4}(u_3|u_4)) \\
& \cdot c_{2,3|1,4}(F_{2|1,4}(u_2|u_1, u_4), F_{3|1,4}(u_3|u_1, u_4))
\end{aligned}$$

Nelle tabelle seguenti si riportano le caratteristiche stimate dei tre modelli C-vine, date dalle copule bivariate associate ad ogni arco della struttura vine e i relativi parametri stimati ($\hat{\theta}^{SP}$, $\hat{\delta}^{SP}$).

Nodi: {RCG; {RCG, IEN}; {IEN, RCA | RCG}}

Arco	Copula	$\hat{\theta}^{SP}$	$\hat{\delta}^{SP}$
$c_{2,1}$	Normale	0,217	/
$c_{2,3}$	Normale	0,220	/
$c_{2,4}$	BB8 ruotata 180°	1,992	0,603
$c_{3,1 2}$	Clayton	0,174	/
$c_{3,4 2}$	BB8 ruotata 180°	2,113	0,488
$c_{1,4 3,2}$	Clayton ruotata 180°	0,119	/
Log-verosimiglianza		894	
AIC		-1772	
BIC		-1714	

Tabella 4.23: Caratteristiche stimate del C-vine con nodi: {RCG; {RCG, IEN}; {IEN, RCA | RCG}}.

Nodi: {RCG; {RCG, RCA}; {RCA, IEN | RCG}}

Arco	Copula	$\hat{\theta}^{SP}$	$\hat{\delta}^{SP}$
$c_{2,1}$	Normale	0,217	/
$c_{2,3}$	Normale	0,220	/
$c_{2,4}$	BB8 ruotata 180°	1,992	0,603
$c_{1,3 2}$	Clayton	0,174	/
$c_{1,4 2}$	Clayton ruotata 180°	0,117	/
$c_{3,4 1,2}$	t-student	0,120	10,964
Log-verosimiglianza		892	
AIC		-1769	
BIC		-1711	

Tabella 4.24: Caratteristiche stimate del C-vine con nodi: {RCG; {RCG, RCA}; {RCA, IEN | RCG}}.

Nodi: {IEN; {IEN, RCA}; {RCA, RCG | IEN}}

Arco	Copula	$\hat{\theta}^{SP}$	$\hat{\delta}^{SP}$
$c_{4,1}$	Clayton ruotata 180°	0,155	/
$c_{4,2}$	BB8 ruotata 180°	2,199	0,543 /
$c_{4,3}$	BB8 ruotata 180°	1,978	0,604
$c_{1,2 4}$	Normale	0,207	/
$c_{1,3 4}$	BB8 ruotata 180°	1,508	0,796
$c_{2,3 1,4}$	Normale	0,170	/
Log-verosimiglianza		891	
AIC		-1764	
BIC		-1699	

Tabella 4.25: Caratteristiche stimate del C-vine con nodi: {IEN; {IEN, RCA}; {RCA, RCG | IEN}}.

Al fine di selezionare il modello più adeguato tra i tre in esame, si è fatto ricorso ai test statistici di Vuong e Clarke; la tabella seguente ne riporta gli esiti.

Vuong valori soglia: $\Phi^{-1}(\alpha/2) = -1,960,$ $\Phi^{-1}(1 - \alpha/2) = 1,960$				Clarke valori soglia: $F_{Bin(H,1/2)}^{-1}(\alpha/2) = 4902,$ $F_{Bin(H,1/2)}^{-1}(1 - \alpha/2) = 5098$		
H0	Statistica test (AIC adj.) (BIC adj.)	P-value	Esito	Statistica test (AIC adj.) (BIC adj.)	P-value	Esito
$f_{1,2,3,4}^I$	0,1706	0,4323	non rifiuto H0	4558	0,0000	$f_{1,2,3,4}^{II}$
\equiv	"	"	"	"	"	"
$f_{1,2,3,4}^{II}$	"	"	"	"	"	"
$f_{1,2,3,4}^I$	0,3703	0,3556	non rifiuto H0	4405	0,0000	$f_{1,2,3,4}^{III}$
\equiv	0,4890	0,3124	"	4413	0,0000	"
$f_{1,2,3,4}^{III}$	0,9168	0,1796	"	4450	0,0000	"
$f_{1,2,3,4}^{II}$	0,1390	0,4447	non rifiuto H0	4933	0,0918	non rifiuto H0
\equiv	0,2257	0,4107	"	4937	0,1056	"
$f_{1,2,3,4}^{III}$	0,5383	0,2952	"	4960	0,2148	"

Tabella 4.26: Test di Vuong e Clarke per i C-vine $f_{1,2,3,4}^I$, $f_{1,2,3,4}^{II}$ e $f_{1,2,3,4}^{III}$.

Gli esiti del test di Vuong portano a non rifiutare mai l'ipotesi nulla di equivalenza tra i tre modelli considerati. Di contro, il test di Clarke porta a preferire il modello C-vine $f_{1,2,3,4}^{II}$ rispetto a $f_{1,2,3,4}^I$ e il modello $f_{1,2,3,4}^{III}$ rispetto a $f_{1,2,3,4}^I$. A fronte di tali risultati e in considerazione del fatto che il modello $f_{1,2,3,4}^{II}$ possiede un valore di AIC migliore rispetto al terzo modello, si è deciso di selezionare come modello C-vine più adeguato la struttura $f_{1,2,3,4}^{II}$.

Si noti che l'applicazione dell'algoritmo di *Dißmann* per selezionare la struttura C-vine porterebbe a scegliere, tra le 12 possibili strutture, la terzultima in termini di AIC.

Di seguito, si riportano i dettagli delle analisi del processo di stima del modello C-vine prescelto. Si ricorda che l'analisi inferenziale non riguarda le variabili marginali, le quali sono rimappate tramite le rispettive pseudo-osservazioni. I parametri del modello, inoltre, sono inizialmente selezionati ricorsivamente tramite approccio SSP; successivamente, tali stime sono utilizzate come valori di partenza nel processo di stima dei parametri SP.

La struttura grafica del C-vine selezionato è riportata nella figura seguente ed è caratterizzata da tre alberi $\mathcal{V} = \{T_1, T_2, T_3\}$.

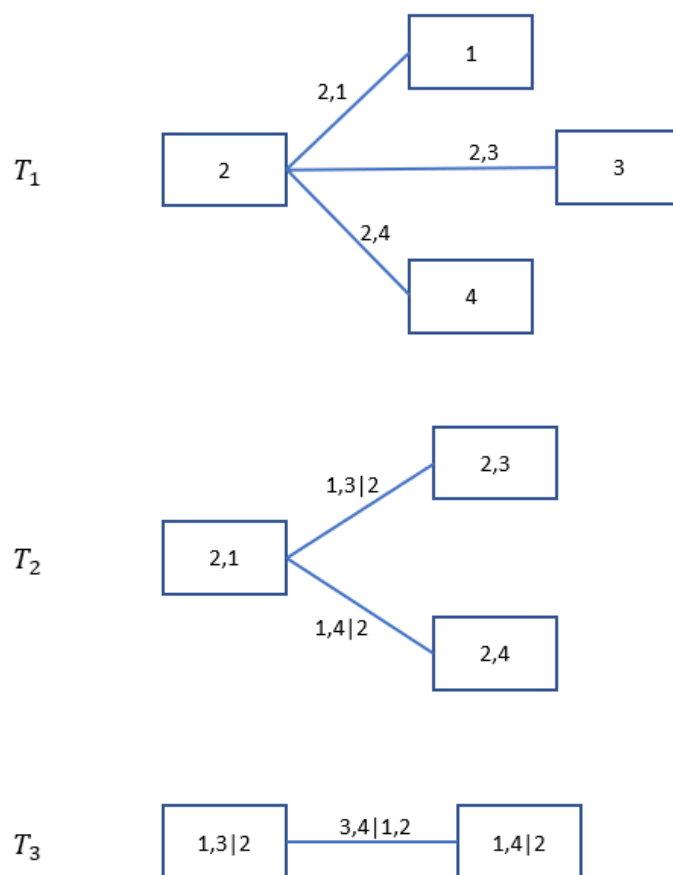


Figura 4.16: Struttura grafica del C-vine selezionato.

Il primo albero è costituito dai nodi $N_1 = \{2, 1, 3, 4\}$ e dagli archi $E_1 = \{\{2, 1\}; \{2, 3\}; \{2, 4\}\}$.

Considerando il primo arco dell'albero, $\{2,1\}$, corrispondente alle variabili marginali RCG e RCA, il valore del τ di Kendall è pari a 0,132 e il test di indipendenza, ad un livello di confidenza $\alpha = 5\%$, porta a rifiutare l'ipotesi H_0 :

Test di indipendenza

valori soglia:

$$\Phi^{-1}(\alpha/2) = -1,960,$$

$$\Phi^{-1}(1 - \alpha/2) = 1,960$$

H0	Statistica test	P-value	Esito
$\tilde{U}_2 \perp \tilde{U}_1$	19,6859	0,0000	rifiuto H0

Tabella 4.27: Test di indipendenza tra le variabili \tilde{U}_2 e \tilde{U}_1 .

Per i *K-plot* e *Chi-plot* si rimanda a figura 4.4 in quanto l'arco $\{1,2\}$ coincide con il primo arco dell'albero T_1 del D-vine selezionato. La copula selezionata, analogamente a quanto visto per il D-vine, è la copula Normale di parametro $\hat{\theta}^{SSP} = \hat{\rho}^{SSP} = 0,211$, a cui corrisponde un τ teorico di 0,136. Per il rispettivo

Lambda-plot si veda, invece, la figura 4.5.

Il secondo arco dell'albero T_1 coinvolge le variabili $\{2,3\}$, corrispondenti alle Lob RCG e IEN; il τ empirico è pari a 0,144. Il test di indipendenza, come mostrato nella tabella seguente, evidenzia la dipendenza statistica tra le variabili coinvolte.

Test di indipendenza

valori soglia:

$$\Phi^{-1}(\alpha/2) = -1,960,$$

$$\Phi^{-1}(1 - \alpha/2) = 1,960$$

H0	Statistica test	P-value	Esito
$\tilde{U}_2 \perp \tilde{U}_3$	21,5819	0,0000	rifiuto H0

Tabella 4.28: Test di indipendenza tra le variabili \tilde{U}_2 e \tilde{U}_3 .

I grafici del *K-plot* e *Chi-plot* palesano anch'essi la presenza di dipendenza positiva nei dati.

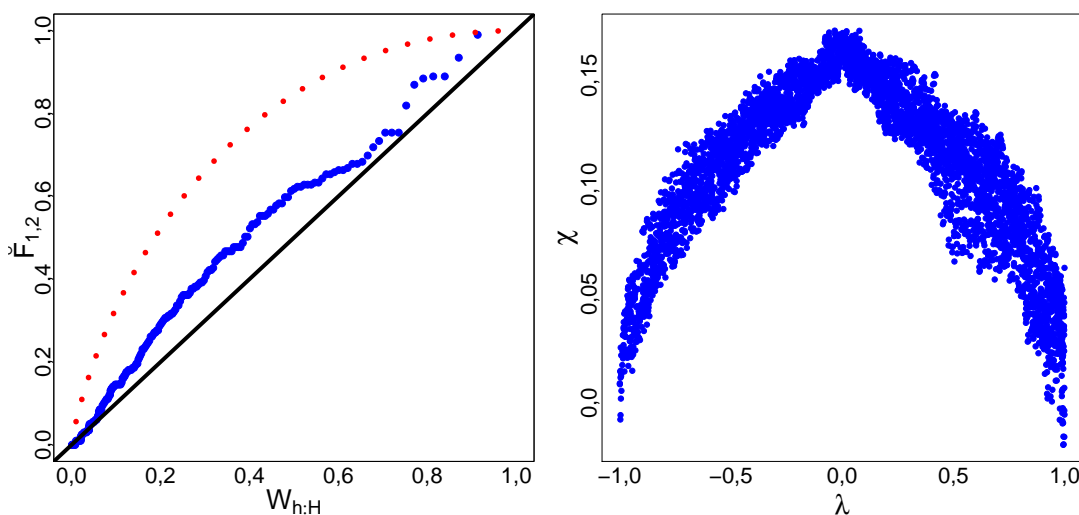


Figura 4.17: *K-plot* e *Chi-plot* per le variabili \tilde{U}_2 e \tilde{U}_3 .

La copula bivariata stimata associata all'arco $\{2,3\}$ è la copula Normale di parametro $\hat{\theta}^{SSP} = \hat{\rho}^{SSP} = 0,219$; il τ teorico determinato dalla copula è pari a 0,141. Di seguito si riporta il *Lambda-plot*.

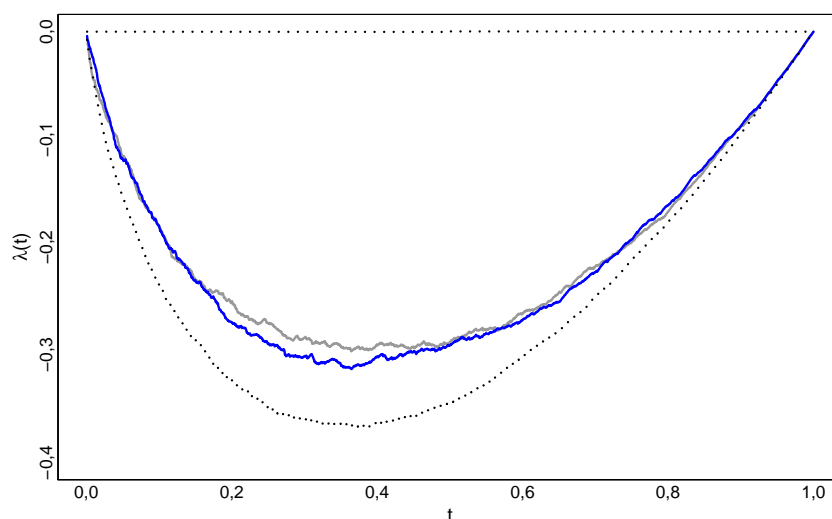


Figura 4.18: *Lambda-plot* delle variabili \tilde{U}_2 e \tilde{U}_3 (curva di colore grigio) e della copula stimata Normale di parametro $\hat{\theta}^{SSP} = \hat{\rho}^{SSP} = 0,219$ (curva di colore blu).

L'ultimo nodo del primo albero mette in relazione le variabili $\{2,4\}$, RCG e ADB, e coincide con il secondo nodo dell'albero T_1 del D-vine selezionato. La dipendenza espressa in termini del τ di Kendall è pari a 0,113 e il test di indipendenza porta a rifiutare l'ipotesi nulla di indipendenza tra le variabili coinvolte.

Test di indipendenza

valori soglia:

$$\Phi^{-1}(\alpha/2) = -1,960,$$

$$\Phi^{-1}(1 - \alpha/2) = 1,960$$

H0	Statistica test	P-value	Esito
$\tilde{U}_2 \perp \tilde{U}_4$	16,9924	0,0000	rifiuto H0

Tabella 4.29: Test di indipendenza tra le variabili \tilde{U}_2 e \tilde{U}_4 .

La copula stimata associata a tale arco risulta essere la copula BB8 ruotata di 180° di parametri $\hat{\theta}^{SSP} = 1,996$ e $\hat{\delta}^{SSP} = 0,597$. Per i grafici del *K-plot*, *Chi-plot* e *Lambda-plot* si rimanda alle figure 4.6 e 4.7 presentate nel paragrafo precedente relativo alle analisi svolte per selezionare la struttura D-vine.

Il secondo albero del C-vine ha nodi, N_2 , coincidenti con gli archi dell'albero precedente e archi, esplicitandone gli insiemi condizionati e condizionanti, definiti come $E_2 = \{\{1, 3|2\}; \{1, 4|2\}\}$. Le copule bivariate coinvolte in tale albero, quindi, risultano essere $c_{1,3|2}(F_{1|2}(u_1|u_2), F_{3|2}(u_3|u_2))$ e $c_{1,4|2}(F_{1|2}(u_1|u_2), F_{4,2}(u_4|u_2))$. Per potere stimare tali copule è necessario determinare gli argomenti delle stesse, dati dalle funzioni di ripartizione condizionate. Ricorrendo alla relazione (3.1.9) è possibile calcolare $F_{1|2}(u_1|u_2)$, $F_{3|2}(u_3|u_2)$ e $F_{4|2}(u_4|u_2)$:

$$F_{1|2}(u_1|u_2) = \frac{\partial C_{1,2}(F_1(u_1), F_2(u_2))}{\partial F_2(u_2)}$$

$$F_{3|2}(u_3|u_2) = \frac{\partial C_{3,2}(F_3(u_3), F_2(u_2))}{\partial F_2(u_2)}$$

$$F_{4|2}(u_4|u_2) = \frac{\partial C_{4,2}(F_4(u_4), F_2(u_2))}{\partial F_2(u_2)}$$

Si osservi che grazie alla proprietà dell'interscambiabilità delle copule, valgono le relazioni $C_{1,2} = C_{2,1}$, $C_{3,2} = C_{2,3}$ e $C_{4,2} = C_{2,4}$. Considerando il primo arco dell'albero T_2 , $\{1,3|2\}$, che coinvolge le Lob RCA e IEN condizionate al ramo RCG, il τ di Kendall calcolato sulle variabili $F_{1|2}$ e $F_{3|2}$ è pari a 0,083. Il test di indipendenza porta a rifiutare l'ipotesi nulla di indipendenza tra le variabili coinvolte.

Test di indipendenza

valori soglia:

$$\Phi^{-1}(\alpha/2) = -1,960,$$

$$\Phi^{-1}(1 - \alpha/2) = 1,960$$

H0	Statistica test	P-value	Esito
$\tilde{U}_{1 2} \perp \tilde{U}_{3 2}$	12,3948	0,0000	rifiuto H0

Tabella 4.30: Test di indipendenza tra le variabili $\tilde{U}_{1|2}$ e $\tilde{U}_{3|2}$.

Di seguito si riportano i grafici del *K-plot* e *Chi-plot*.

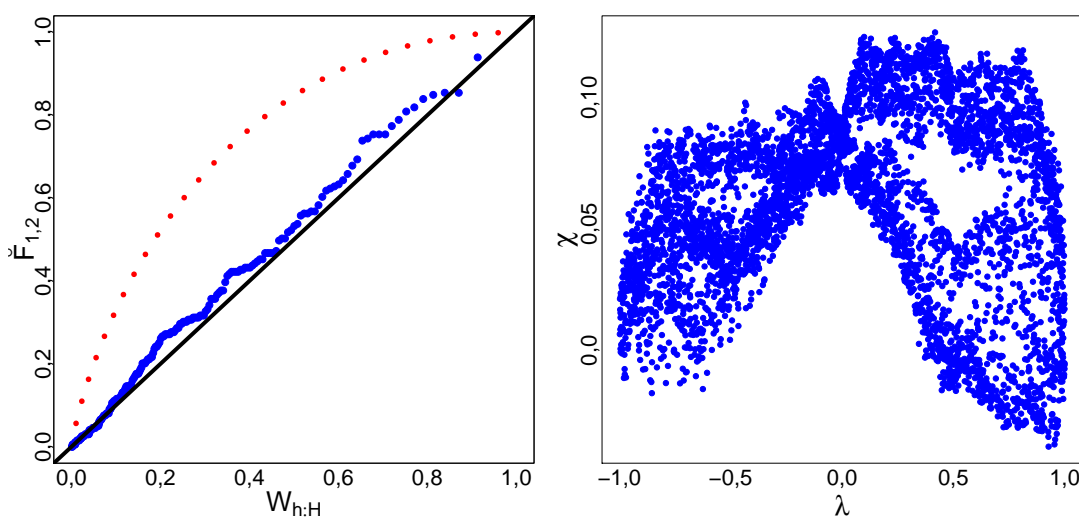


Figura 4.19: *K-plot* e *Chi-plot* per le variabili $\tilde{U}_{1|2}$ e $\tilde{U}_{3|2}$.

La copula bivariata stimata risulta essere la copula Clayton di parametro $\hat{\theta}^{SSP} = 0,163$, a cui corrisponde un τ teorico di 0,075. Il coefficiente di dipendenza di coda inferiore, invece, è pari a $\lambda_L = 0,014$. La figura seguente riporta il *Lambda-plot*.

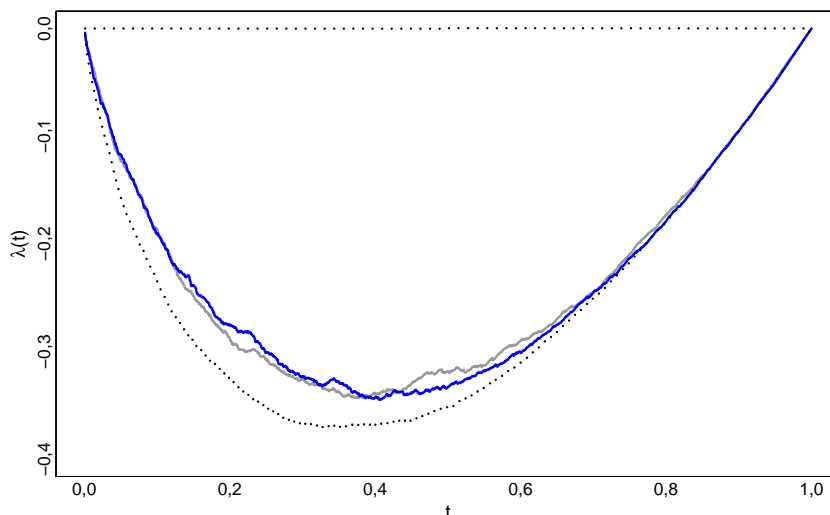


Figura 4.20: *Lambda-plot* delle variabili $\tilde{U}_{1|2}$ e $\tilde{U}_{3|2}$ (curva di colore grigio) e della copula stimata Clayton di parametro $\hat{\theta}^{SSP} = 0,163$ (curva di colore blu).

La validità dell'ipotesi semplificata alla base della pair-copula construction è inizialmente valutata tramite il test CCC. Come mostrato nella tabella seguente i risultati del test portano a rifiutare, ad un livello di confidenza $\alpha = 5\%$, l'ipotesi di indipendenza.

Test CCC

valori soglia:

$$F_{\chi^2}^{-1}(\alpha/2) = 0,001,$$

$$F_{\chi^2}^{-1}(1 - \alpha/2) = 5,024$$

H0	Statistica test	P-value	Esito
$(\tilde{U}_{1 2}, \tilde{U}_{3 2}) \perp \tilde{U}_2$	5,3803	0,0204	rifiuto H0

Tabella 4.31: Test CCC tra la variabile $(\tilde{U}_{1|2}, \tilde{U}_{3|2})$ e la variabile \tilde{U}_2 .

A fronte della dipendenza tra il vettore $(\tilde{U}_{1|2}, \tilde{U}_{3|2})$ e la variabile \tilde{U}_2 , per potere valutare la validità dell'ipotesi semplificata è necessario analizzare la copula trivariata sottostante alle variabili in esame. La stima della copula $c(F_{1|2}(u_1|u_2), F_{3|2}(u_3|u_2), u_2)$ mostra che la famiglia Normale possiede il migliore adattamento ai dati; la selezione della famiglia Clayton per la copula condizionata, $c_{1,3|2}$, tuttavia porta a concludere che l'incognito modello di dipendenza sottostante alle tre v.a. non è Normale, palesando la difficoltà a selezionare modelli di dipendenza multivariati

adeguati tramite le famiglie di copule sviluppate in letteratura. A fronte dei risultati ottenuti, quindi, è possibile concludere che nel contesto in esame l'ipotesi semplificata della PCC non è soddisfatta; l'indicazione che la copula trivariata più adeguata sia quella Normale, tuttavia, potrebbe indicare che lo scostamento dall'ipotesi semplificata non sia elevato. In definitiva, quindi, la copula coinvolta è un'approssimazione di quella nella forma semplificata:

$$c_{1,3|2}(F_{1|2}(u_1|u_2), F_{3|2}(u_3|u_2)) \approx c_{1,3|2}(F_{1|2}(u_1|u_2), F_{3|2}(u_3|u_2)|u_2)$$

Il secondo nodo dell'albero T_2 coinvolge le variabili $\{1,4|2\}$, associate ai rami RCA e ADB condizionati alla Lob RCG, e coincide con il primo nodo dell'albero T_2 del D-vine selezionato in precedenza. Il τ di Kendall tra le variabili $F_{1|2}$ e $F_{4|2}$ è pari a 0,051; il test di indipendenza rigetta l'ipotesi nulla di indipendenza tra le variabili coinvolte.

Test di indipendenza

valori soglia:

$$\Phi^{-1}(\alpha/2) = -1,960,$$

$$\Phi^{-1}(1 - \alpha/2) = 1,960$$

H0	Statistica test	P-value	Esito
$\tilde{U}_{1 2} \perp \tilde{U}_{4 2}$	7,6804	0,0000	rifuto H0

Tabella 4.32: Test di indipendenza tra le variabili $\tilde{U}_{1|2}$ e $\tilde{U}_{4|2}$.

Si rimanda alle figura 4.10, sviluppate per il D-vine, per i *K-plot* e *Chi-plot* associati alle variabili in esame.

Il processo di stima porta a selezionare la copula Clayton ruotata di 180° di parametro $\hat{\theta}^{SSP} = 0,126$, a cui è associato un τ teorico di 0,059. Il coefficiente di dipendenza di coda superiore, infine, è pari $\lambda_U = 0,004$. Per il *Lambda-plot* si rimanda alla figura 4.11.

Il test CCC implementato al fine di valutare l'indipendenza tra la variabile bivariata $(\tilde{U}_{1|2}, \tilde{U}_{4|2})$ e la variabile \tilde{U}_2 porta a rifiutare l'ipotesi semplificata alla base della PCC (si veda la tabella 4.15); per i dettagli delle analisi svolte per valutare l'ipotesi semplificata si rimanda a quanto sviluppato in precedenza per il primo nodo del dell'albero T_2 del D-vine. Tale analisi portano a concludere che la copula bivariata coinvolta è un'approssimazione di quella nella forma semplificata:

$$c_{1,4|2}(F_{1|2}(u_1|u_2), F_{4|2}(u_4|u_2)) \approx c_{1,4|2}(F_{1|2}(u_1|u_2), F_{4|2}(u_4|u_2)|u_2)$$

L'ultimo albero del C-vine, T_3 , possiede nodi coincidenti con gli archi dell'albero precedente, $N_3 = E_2$, ed un solo arco, dato da $E_3 = \{3,4|1,2\}$, associato alle linee di business IEN e ADB condizionate alle Lob RCA e RCG. La copula bivariata coinvolta è $c_{3,4|1,2}(F_{3|1,2}(u_3|u_1, u_2), F_{4|1,2}(u_4|u_1, u_2))$, i cui argomenti sono ottenuti grazie alla relazione (3.1.9):

$$F_{3|1,2}(u_3|u_1, u_2) = \frac{\partial C_{3,1|2}(F_{3|2}(u_3|u_2), F_{1|2}(u_1|u_2))}{\partial F_{1|2}(u_1|u_2)}$$

$$F_{4|1,2}(u_4|u_1, u_2) = \frac{\partial C_{4,1|2}(F_{4|2}(u_4|u_2), F_{1|2}(u_1|u_2))}{\partial F_{1|2}(u_1|u_2)}$$

Data la proprietà dell'interscambiabilità delle copule coinvolte, valgono le relazioni $C_{3,1|2} = C_{1,3|2}$ e $C_{4,1|2} = C_{1,4|2}$.

La dipendenza tra le variabili $\tilde{U}_{3|1,2} = F_{3|1,2}$ e $\tilde{U}_{4|1,2} = F_{4|1,2}$ è pari a $\tau = 0,087$; il test di indipendenza porta a rifiutare l'ipotesi nulla di indipendenza tra le variabili coinvolte.

Test di indipendenza

valori soglia:

$$\Phi^{-1}(\alpha/2) = -1,960,$$

$$\Phi^{-1}(1 - \alpha/2) = 1,960$$

H0	Statistica test	P-value	Esito
$\tilde{U}_{3 1,2} \perp \tilde{U}_{4 1,2}$	13,0477	0,0000	rifiuto H0

Tabella 4.33: Test di indipendenza tra le variabili ($\tilde{U}_{3|1,2}$ e $\tilde{U}_{4|1,2}$).

Di seguito si riportano i grafici del *Chi-plot* e del *K-plot* che mostrano la presenza di dipendenza positiva tra le variabili casuali in esame; in particolare, i punti del *K-plot* sono disposti al di sopra della diagonale principale, mentre i valori χ del *Chi-plot* sono principalmente concentrati sopra lo zero.

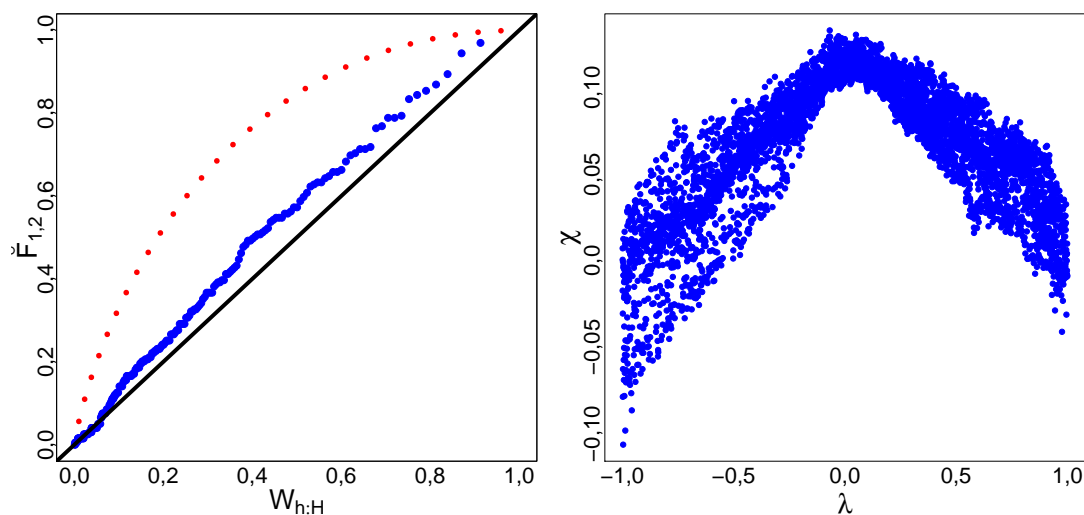


Figura 4.21: *K-plot* e *Chi-plot* per le variabili $\tilde{U}_{3|1,2}$ e $\tilde{U}_{4|1,2}$.

La copula bivariata selezionata coincide con la copula t-student di parametri $\hat{\theta}^{SSP} = \hat{\rho}^{SSP} = 0,120$ e $\hat{\delta}^{SSP} = \hat{m}^{SSP} = 11,048$. Il τ teorico associato a tale specificazione della copula è di 0,077, mentre i coefficienti di dipendenza di coda superiore e inferiore sono pari a $\lambda_L = \lambda_U = 0,010$. Di seguito si riporta il

Lambda-plot della copula in esame.

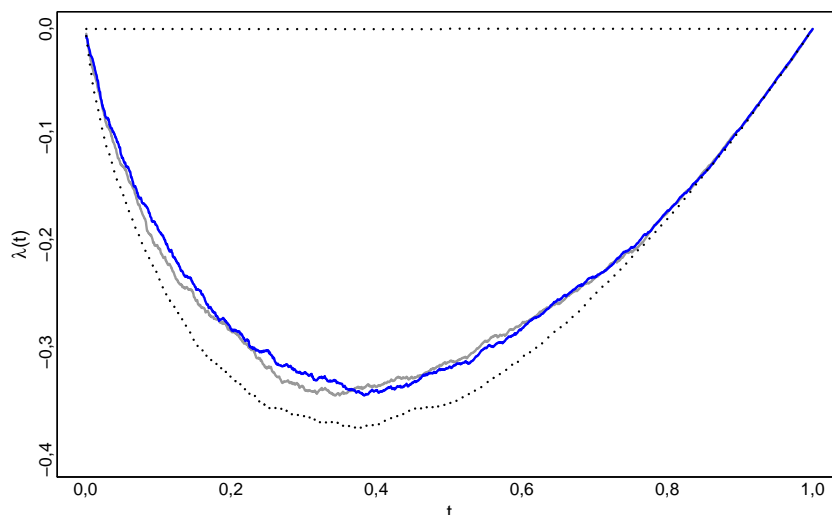


Figura 4.22: *Lambda-plot* delle variabili $\tilde{U}_{3|1,2}$ e $\tilde{U}_{4|1,2}$ (curva di colore grigio) e della copula stimata t-student di parametri $\hat{\theta}^{SSP} = \hat{\rho}^{SSP} = 0,120$ e $\hat{\delta}^{SSP} = \hat{m}^{SSP} = 11,048$ (curva di colore blu).

Il constant conditional correlations test, infine, è un primo strumento che permette di valutare l'adeguatezza dell'ipotesi semplificata sottostante alla PCC. L'esito del test, riportato nella seguente tabella, porta a rifiutare l'ipotesi di indipendenza tra le variabili condizionante e quelle condizionanti.

Test CCC

valori soglia:

$$F_{\chi^2}^{-1}(\alpha/2) = 0,001,$$

$$F_{\chi^2}^{-1}(1 - \alpha/2) = 5,024$$

H0	Statistica test	P-value	Esito
$(\tilde{U}_{3 1,2}, \tilde{U}_{4 1,2}) \perp (\tilde{U}_1, \tilde{U}_2)$	31,7758	0,0000	rifuto H0

Tabella 4.34: Test di indipendenza tra le variabili $(\tilde{U}_{3|1,2}, \tilde{U}_{4|1,2})$ e $(\tilde{U}_1, \tilde{U}_2)$.

A fronte dell'esito del test CCC, per potere valutare la validità dell'ipotesi semplificata è necessario stimare la copula sottostante alle 4 v.a. coinvolte. Il processo di stima porta a selezionare per la copula $c(F_{3|1,2}(u_3|u_1, u_2), F_{4|1,2}(u_4|u_1, u_2), u_1, u_2)$ la famiglia t-student. Tale selezione è in accordo con la stima della copula bivaricata $c_{3,4|1,2}$, che appartiene anch'essa alla famiglia t-student. I risultati di stima, quindi, portano ad affermare la validità dell'ipotesi semplificata:

$$c_{3,4|1,2}(F_{3|1,2}(u_3|u_1, u_2), F_{4|1,2}(u_4|u_1, u_2)) = c_{3,4|1,2}(F_{3|1,2}(u_3|u_1, u_2), F_{4|1,2}(u_4|u_1, u_2)|u_1, u_2)$$

I parametri delle copule bivariate stimati tramite approccio stepwise semi-parametrico (SSP) sono utilizzati, in seguito, come valori di partenza per la massimizzazione della funzione di verosimiglianza complessiva del C-vine. Tale procedura semiparametrica (SP) permette di stimare i valori dei parametri del modello.

La tabella seguente permette di confrontare le stime dei parametri ottenute tramite approccio SSP e SP; si noti che i valori stimati con i due approcci sono in accordo tra loro, a riprova dell'adeguata selezione delle famiglie di copule bivariate.

Arco	Copula	$\hat{\theta}^{SSP}$	$\hat{\delta}^{SSP}$	$\hat{\theta}^{SP}$	$\hat{\delta}^{SP}$
$c_{2,1}$	Normale	0,211	/	0,217	/
$c_{2,3}$	Normale	0,219	/	0,220	/
$c_{2,4}$	BB8 ruotata 180°	1,996	0,597	1,992	0,603
$c_{1,3 2}$	Clayton	0,163	/	0,174	/
$c_{1,4 2}$	Clayton ruotata 180°	0,126	/	0,117	/
$c_{3,4 1,2}$	t-student	0,120	11,048	0,120	10,964

Tabella 4.35: Parametri stimati tramite approccio SSP e SP per il C-vine selezionato.

Il valore di log-verosimiglianza del modello, considerando i parametri stimati tramite metodo SP, è pari a 892, mentre i valori di AIC e BIC sono rispettivamente pari a -1769 e -1711.

Analogamente a quanto svolto per il modello D-vine selezionato, si riporta il confronto, in termini di valori di log-verosimiglianza, AIC e BIC del modello, tra il C-vine selezionato e le copule multidimensionali ($n = 4$) Normale e t-student.

Copula	Log-veros.	AIC	BIC
C-vine	892	-1769	-1711
C^{Ga}	784	-1485	-1442
C^t	762	-1511	-1461

Tabella 4.36: Confronto tra il C-vine selezionato e le copule Normale e t-student multivariate.

I valori riportati nella tabella precedente mostrano che il modello C-vine è superiore rispetto alle copule Normale e t-student. Infine, analogamente a quanto svolto per il D-vine, la copula C-vine selezionata è stata confrontata con le copule archimedee multivariate Clayton, Joe, Frank e Gumbel, le quali hanno mostrato un adattamento ai dati particolarmente contenuto.

4.2.3 Selezione della vine copula

I valori di log-verosimiglianza, AIC e BIC dei modelli D-vine e C-vine selezionati nei paragrafi precedenti sono riassunti nella tabella seguente:

Copula	Log-veros.	AIC	BIC
D-vine	899	-1783	-1726
C-vine	892	-1769	-1711

Tabella 4.37: Confronto tra i modelli D-vine e C-vine selezionati.

La struttura D-vine è superiore rispetto al modello C-vine sia in termini di log-verosimiglianza che di AIC e BIC; si noti che entrambi i modelli possiedono otto parametri, quindi, al fine di confrontarli è sufficiente attenersi ai valori di log-verosimiglianza in quanto la penalità introdotta dagli indicatori AIC e BIC per tenere conto della parsimonia dei modelli, intesa come numero dei parametri, è analoga. Al fine di selezionare il modello di vine copula più appropriato, inoltre, sono stati implementati i test statistici di Vuong e Clarke che permettono di valutare tra due modelli quale sia quello statisticamente migliore. La tabella seguente riporta i risultati dei test.

H0	Vuong			Clarke		
	Statistica test (AIC adj.) (BIC adj.)	P-value	Esito	Statistica test (AIC adj.) (BIC adj.)	P-value	Esito
$f_{1,2,3,4}^{D-vine}$	0,7851	0,2162	non rifiuto H0	5301	0,0000	$f_{1,2,3,4}^{D-vine}$
=	"	"	"	"	"	"
$f_{1,2,3,4}^{C-vine}$	"	"	"	"	"	"

Tabella 4.38: Test di Vuong e Clarke sui modelli D-vine e C-vine selezionati.

Il test di Vuong non riesce a discriminare tra i due modelli, avvalorando l'ipotesi nulla di equivalenza tra gli stessi; di contro, il test di Clarke rigetta l'ipotesi nulla in favore del modello D-vine. Quest'ultimo, stante l'evidenza del test statistico di Clarke e il maggiore valore di log-verosimiglianza, è il modello prescelto con il quale si procede ad aggregare i rischi marginali del reserve risk delle quattro linee di business in esame.

Di seguito si riporta la matrice dei valori del τ di Kendall indotti dal D-vine selezionato e ottenuti tramite 100.000 simulazioni

τ	RCA	RCG	IEN	ADB
RCA	1			
RCG	0,14	1		
IEN	0,11	0,14	1	
ADB	0,08	0,12	0,12	1

Tabella 4.39: Valori del τ di Kendall ottenuti tramite simulazione di 100.000 valori dal D-vine selezionato.

4.3 La modellizzazione dei rischi marginali tramite Collective Risk Model

I metodi stocastici per valutare il rischio di riservazione hanno l'obiettivo di quantificare la variabilità o l'intera distribuzione di probabilità del rischio in esame con lo scopo di stimare il capitale da detenere a fini di solvibilità. In letteratura esistono molteplici modelli stocastici sviluppati al fine di modellizzare il non-life reserve risk. Un primo approccio fu proposto da Mack (1993, 1994, 1999), il quale permette di determinare la *prediction variance*, scomposta tra *process variance* e *estimation variance*, associata alla stima Chain-Ladder della riserva sinistri. In seguito, sono stati sviluppati ulteriori approcci, quali il bootstrap (England e Verrall (1999)) e i modelli lineari generalizzati (England e Verrall (2002)) al fine di valutare il rischio di riservazione. In anni recenti, inoltre, nel contesto del reserve risk sono stati introdotti modelli bayesiani; il vantaggio e la peculiarità di tali approcci consistono nel potere studiare la distribuzione di probabilità dei parametri del modello e potere introdurre in modo rigoroso all'interno del modello informazioni esterne. I principali modelli attuariali sviluppati coerentemente con il framework bayesiano sono il metodo Chain-Ladder (Sanchez, Verrall e Vilar (2011), Scollnik (2004) e Wüthrich (2007)), il Bornhuetter-Ferguson (England, Verrall e Wüthrich (2018), Scollnik (2004) e Wüthrich (2007)) e il modello Overdispersed Poisson (England, Verrall e Wüthrich (2018)). In Meyers (2009), inoltre, si presenta un approccio Collective Risk Model bayesiano basato sul metodo deterministico di stima della riserva sinistri Cape Code.

Nel presente caso studio il reserve risk marginale delle singole linee di business in esame viene analizzato seguendo un approccio Collective Risk Model (CRM).

Il metodo CRM per valutare in termini stocastici la riserva sinistri è un approccio introdotto e utilizzato da diversi autori, quali l'International Actuarial Association (2004), Meyers (2009) e Savelli e Clemente (2009, 2011). In seguito, l'approccio è stato esteso da Ricotta e Clemente (2016), dove i pagamenti incrementali oggetto di stima del triangolo di run-off sono modellizzati tramite un processo di Poisson composto. Gli autori, inoltre, considerano due fattori di disturbo, applicati rispettivamente alla componente del numero dei sinistri e alla componente di severity del modello.

L'approccio CRM seguito in tale caso studio è quello sviluppato da Ricotta e Luini (2019), in cui gli autori, colmando un *gap* esistente in letteratura, valutano la variabilità dei fattori di disturbo, atti a modellizzare l'incertezza associata alla stima dei parametri, nonché la dipendenza tra i parametri stessi, tramite un approccio bayesiano. L'obiettivo di tale procedura è quello di considerare l'intera informazione storica disponibile contenuta nei triangoli di run-off. Alla base del metodo bayesiano vengono utilizzate due consolidate metodologie della letteratura attuariale, quali il bootstrap applicato al metodo Chain-Ladder (England e Verrall (2002)) e la formula di Mack (Mack (1993)). L'algoritmo bayesiano è implementato facendo uso di distribuzioni a priori improprie (cd. *flat*) e delle distribuzioni a priori di Jeffreys. Entrambe le tipologie di distribuzioni a priori sono definite non informative e vengono adottate quando non è disponibile in modo esplicito informazione a priori; le distribuzioni a priori flat rispondono al principio di indifferenza postulato da Laplace, mentre le distribuzioni a priori di Jeffreys sono invarianti a riparametrazioni del modello.

Il metodo CRM in esame permette di modellizzare, tramite approccio simulativo Monte Carlo, la distribuzione della riserva sinistri considerando l'intero orizzonte temporale di vita delle passività assicurative (cd. *approccio a totale run-off*). La valutazione del reserve risk su un orizzonte temporale annuo (cd. *approccio one-year*), di contro, è perseguita ricorrendo all'approccio *re-reserving* (Diers (2009) e Ohlsson e Lauzenings (2009)) applicato coerentemente al metodo stocastico in questione. La distribuzione one-year della riserva sinistri permette, infine, di quantificare il requisito di capitale da detenere a fini di solvibilità come previsto dalla normativa Solvency II.

4.3.1 Il Collective Risk Model

Il Collective Risk Model sviluppato da Ricotta e Luini (2019), basato sulla Teoria del Rischio Collettiva, ha l'obiettivo di modellizzare in termini stocastici la riserva sinistri. Quest'ultima è rappresentata attraverso i triangoli di run-off, che nel presente caso studio possiedono un numero di generazioni (i) e antidurate (j) pari a $N = 12$; la dimensione dei triangoli, quindi, risulta essere 12×12 , con $i = 1, \dots, 12$ e $j = 1, \dots, 12$. Le grandezze alla base del modello sono rappresentate dall'importo e dal rispettivo numero dei sinistri che verranno pagati negli anni di calendario futuri. Tali dati sono stimati tramite metodo deterministico Frequency-Severity basato sui triangoli di run-off osservati degli importi e dei numeri, individuati con D e D^n (si veda paragrafo 4.1).

Il metodo CRM, considerando un orizzonte temporale a totale run-off, modella i pagamenti incrementali futuri, corrispondenti alle celle inferiori del triangolo di run-off degli importi incrementali, tramite la seguente variabile aleatoria:

$$\tilde{X}_{i,j} = \sum_{h=1}^{\tilde{K}_{i,j}} \tilde{p} \tilde{Z}_{i,j,h} \quad (4.3.1.1)$$

mentre la v.a. della riserva sinistri, \tilde{R} , data dalla somma di tutte le celle inferiori oggetto di stima del triangolo di run-off, risulta:

$$\tilde{R} = \sum_{i=1}^N \sum_{j=N-i+2}^N \tilde{X}_{i,j} \quad (4.3.1.2)$$

dove:

- $\tilde{K}_{i,j}$ individua la v.a. del numero dei sinistri di generazione i che verranno liquidati nell'anno di sviluppo j . Tale v.a. è modellizzata tramite un processo di Poisson misturato, dove la variabilità di parametro è introdotta attraverso la v.a. \tilde{q} , dotata di media unitaria e standard deviation $\sigma_{\tilde{q}}$, che rappresenta un fattore di disturbo. La v.a. del numero dei sinistri, quindi, risulta la seguente: $\tilde{K}_{i,j} \sim Po(\tilde{q}n_{i,j})$.
- $\tilde{Z}_{i,j,h}$ è la v.a. che rappresenta l'ammontare (cd. *severity*) del sinistro h -esimo che verrà liquidato nella cella (i, j) .
- \tilde{p} è la v.a. che permette di introdurre incertezza di parametro sulla severity del modello. La variabile \tilde{p} possiede media unitaria e standard deviation $\sigma_{\tilde{p}}$.

I fattori di disturbo del modello permettono di introdurre e gestire l'incertezza associata alla stima dei parametri del CRM; possedendo media unitaria, inoltre, non hanno alcun effetto sul valore atteso della riserva sinistri \tilde{R} . Il modello CRM è caratterizzato da due soli fattori di disturbo, applicati rispettivamente al numero e costo dei sinistri, che agiscono su tutto il triangolo inferiore di run-off; le variabili \tilde{q} e \tilde{p} , quindi, permettono di introdurre dipendenza, imputabile al processo liquidativo dei sinistri, tra le differenti generazioni e anni di sviluppo dei sinistri.

Le ipotesi alla base del modello CRM sono le seguenti:

- il numero dei sinistri ($\tilde{K}_{i,j}$), il costo dei sinistri ($\tilde{Z}_{i,j,h}$) e il fattore di disturbo \tilde{p} sono v.a. tra loro indipendenti in ogni cella (i, j) oggetto di stima del triangolo inferiore di run-off;
- i costi dei sinistri nelle diverse celle inferiori del triangolo di run-off sono v.a. indipendenti e nella stessa cella sono indipendenti e identicamente distribuite;
- il fattore di disturbo \tilde{q} è indipendente dalla severity dei sinistri in ogni cella.

In Ricotta e Clemente (2016) gli autori, sotto ipotesi di indipendenza tra le v.a. \tilde{q} e \tilde{p} , ricavano le formule analitiche della media, standard deviation e asimmetria della v.a. della riserva sinistri e mostrano come il valore atteso della variabile

\tilde{R} coincida con il valore di riserva stimato tramite il metodo deterministico sottostante. Gli autori, inoltre, evidenziano l'effetto non trascurabile dei fattori di disturbo sui momenti della riserva sinistri e propongono un approccio basato sulla formula di Mack per quantificarne la variabilità.

Nell'approccio CRM in esame, le v.a. \tilde{q} e \tilde{p} sono considerate tra loro dipendenti e non è più possibile ricavare analiticamente le espressioni dello scarto quadratico medio e dell'asimmetria della riserva sinistri; il valore atteso, di contro, non è affetto dalle v.a. \tilde{q} e \tilde{p} e risulta essere:

$$E(\tilde{R}|D; D^n) = \sum_{i=1}^N \sum_{j=N-i+2}^N n_{i,j} m_{i,j}$$

dove $n_{i,j}$ e $m_{i,j}$ individuano il numero atteso e costo medio dei sinistri che verranno liquidati nella generica cella (i, j) .

Al fine di implementare il modello CRM è necessario stimare per ogni cella (i, j) del triangolo inferiore di run-off le grandezze del numero incrementale atteso dei sinistri che verranno pagati ($n_{i,j}$) e il rispettivo costo medio ($m_{i,j}$); tali parametri sono stimati con il metodo deterministico Frequency-Severity, come mostrato nel paragrafo 4.1. Per implementare il metodo CRM, infine, è necessario adottare un'ipotesi distributiva per la grandezza della severity di ogni cella (i, j) ; a tal fine, è stata adottata la distribuzione Gamma, $\tilde{Z}_{h,i,j} \sim \text{Gamma}(\alpha_{i,j}; \beta_{i,j})$, i cui parametri sono stimati tramite la grandezza $m_{i,j}$ e facendo uso del coefficiente di variazione del costo dei sinistri calcolato per ogni antidurata del triangolo.

4.3.1.1 La stima bayesiana della variabilità dei fattori di disturbo

Le variabili aleatorie rappresentative dei fattori di disturbo introducono, nel contesto del CRM, l'incertezza associata alla stima dei parametri del modello, dati dal numero e importo dei sinistri. Le variabili \tilde{q} e \tilde{p} rappresentano, inoltre, la componente di rischio sistematico e non diversificabile del modello. Quanto proposto da Ricotta e Luini (2019), presentato nel seguito, ha l'obiettivo di sviluppare un approccio coerente con il framework bayesiano per stimare la standard deviation delle variabili \tilde{q} e \tilde{p} .

La statistica bayesiana, in contrasto con quella classica, permette di prendere in considerazione l'incertezza dei parametri di un modello statistico; la variabilità di parametro è introdotta tramite la distribuzione di probabilità a priori (*prior distribution*) che, congiuntamente con i dati osservati, permette di ottenere la distribuzioni a posteriori (*posterior distribution*) del parametro oggetto di analisi. Il teorema di Bayes determina la distribuzione a posteriori, $f(\theta|x)$, come segue:

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{f(x)}$$

dove:

- $f(x|\theta)$ è la densità campionaria dei dati osservati in funzione di un prescelto modello probabilistico; tale termine, visto come una funzione di θ per i valori osservati di x corrisponde alla funzione di verosimiglianza.
- $f(\theta)$ è la distribuzione a priori del parametro, che individua l'incertezza di parametro e rappresenta, inoltre, la conoscenza a priori sui valori del parametro.
- $f(x)$ è la distribuzione marginale dei dati, la quale non dipende da θ e risulta essere una costante di normalizzazione che permette di ottenere una distribuzione a posteriori non impropria.

Il termine $f(x)$, che ha il solo scopo di riscaldare la distribuzione a posteriori, è in genere omissa dalla formula di Bayes, la quale, in tale circostanza, diviene:

$$f(\theta|x) \propto f(x|\theta)f(\theta)$$

mostrando come la distribuzione a posteriori del parametro del modello sia proporzionale al prodotto tra la funzione di verosimiglianza e la distribuzione a priori. Le distribuzioni a priori vengono tipicamente classificate in distribuzioni non informative e distribuzioni informative. Le prime si rifanno al principio di indifferenza e hanno un effetto minimo rispetto ai dati sulla posterior distribution. Le distribuzioni informative, di contro, sono calibrate ricorrendo ai dati osservati. L'approccio bayesiano permette, inoltre, di fare inferenza sui futuri dati osservati (y) attraverso la *posterior predictive distribution*; questa è una media della distribuzione di probabilità di y condizionata rispetto al valore incognito di θ , ponderata per la distribuzione a posteriori di θ :

$$f(y|x) = \int f(y|\theta)f(\theta|x) d\theta$$

Il framework costituito dalla formula di Bayes, quindi, permette di ottenere la posterior predictive distribution, che contiene informazioni relative alle nuove osservazioni, e la posterior distribution che racchiude informazioni relative al parametro del modello. Considerando la distribuzione a posteriori è possibile sviluppare differenti analisi inferenziali, quali stime puntuali, intervallari o test di ipotesi.

Nel contesto del Collective Risk Model l'approccio bayesiano è utilizzato con lo scopo di quantificare la standard deviation dei fattori di disturbo. Seguendo la consueta assunzione della Teoria del Rischio Collettiva, si ipotizza che le v.a. \tilde{q} e \tilde{p} siano delle variabili Gamma a media unitaria:

$$\tilde{q} \sim \text{Gamma}(h; h) \quad \tilde{p} \sim \text{Gamma}(k; k)$$

I fattori di disturbo possiedono media pari a 1, data dal rapporto tra i parametri, e standard deviation definita rispettivamente come $\sigma_{\tilde{q}} = 1/\sqrt{h}$ e $\sigma_{\tilde{p}} = 1/\sqrt{k}$. I

valori di $\sigma_{\tilde{q}}$ e $\sigma_{\tilde{p}}$, quindi, permettono di individuare l'unico parametro caratteristico delle v.a. \tilde{q} e \tilde{p} da cui ne dipendono tutti i momenti.

Nel seguito, le grandezze $\sigma_{\tilde{q}}$ e $\sigma_{\tilde{p}}$ sono considerate aleatorie⁴ e analizzate coerentemente con il framework bayesiano. In particolare, si assume che le v.a. $\tilde{\sigma}_{\tilde{q}}$ e $\tilde{\sigma}_{\tilde{p}}$, definite per valori positivi, seguano una distribuzione Gamma:

$$\tilde{\sigma} \sim \text{Gamma}(\tilde{A}; \tilde{B})$$

dove le grandezze $\tilde{A} \sim f(a)$ e $\tilde{B} \sim f(b)$, definite iperparametri, sono le v.a. tramite cui è introdotta nel modello informazione a priori. In tale contesto, la valutazione della standard deviation dei fattori di disturbo è effettuata ricorrendo alla formula di Bayes al fine di ottenere la distribuzione a posteriori dei parametri da cui le v.a. $\tilde{\sigma}_{\tilde{q}}$ e $\tilde{\sigma}_{\tilde{p}}$ dipendono:

$$f(a, b|\sigma) \propto f(\sigma|a, b)f(a)f(b) \quad (4.3.1.1.1)$$

Il valore atteso della distribuzione a posteriori, infine, è utilizzato per calibrare la v.a. che rappresenta la standard deviation dei fattori di disturbo:

$$\tilde{\sigma} \sim \text{Gamma}\left(E(\tilde{A}|\tilde{\sigma}); E(\tilde{B}|\tilde{\sigma})\right) \quad (4.3.1.1.2)$$

Nell'equazione (4.3.1.1.1) si ipotizza che le distribuzioni $f(a)$ e $f(b)$ siano tra loro indipendenti; si noti che tale assunzione non inficia l'approccio bayesiano in questione in quanto la relazione (4.3.1.1.2) prevede di calibrare le standard deviation dei fattori di disturbo tramite i valori attesi della distribuzione a posteriori, considerando, quindi, separatamente le distribuzioni a posteriori marginali dei singoli iperparametri.

I dati osservati che rientrano nella funzione di verosimiglianza della relazione (4.3.1.1.1), $f(\sigma|a, b)$, sono ottenuti tramite la formula di Mack implementata con un approccio bootstrap. In particolare, come già proposto in Ricotta e Clemente (2016), l'estimation variance della formula di Mack, che misura la variabilità associata alla stima dei parametri, è utilizzata per quantificare lo scarto quadratico medio dei fattori di disturbo. In tale contesto, la metodologia bootstrap come sviluppata da England e Verrall (2002) è applicata ai triangoli di run-off del numero cumulato e costo medio medio cumulato dei sinistri; ad ogni iterazione, il rapporto tra la radice quadrata dell'estimation variance di Mack e il valore della rispettiva grandezza in esame stimata tramite Chain-Ladder quantifica la variabilità relativa associata alla stima dei parametri. Tale variabilità relativa, intesa come coefficiente di variazione, inoltre, coincide con lo scarto quadratico medio della v.a. \tilde{q} e \tilde{p} in quanto tali variabili possiedono media unitaria. In dettaglio, ogni step simulativo di tale procedura, definita *Mack-Bootstrap*, considerando i triangoli dei numeri cumulati e costi medi cumulati, consta dei seguenti passaggi.

⁴Le grandezze $\sigma_{\tilde{q}}$ e $\sigma_{\tilde{p}}$, in quanto v.a., saranno denotate con la tilde; inoltre, al fine di semplificare la notazione, quando saranno eseguite considerazioni di carattere generale verrà omissa il pedice \tilde{q} o \tilde{p} .

- 1) Determinare i fattori di sviluppo del metodo Chain-Ladder del triangolo delle grandezze cumulate osservate.
- 2) Dai link ratio e dai dati osservati sull'ultima diagonale del triangolo determinare ricorsivamente le grandezze cumulate del triangolo superiore; tramite differenza, in seguito, calcolare le grandezze incrementali.
- 3) Calcolare i residui del modello tra le grandezze incrementali del passo precedente e i dati osservati dei triangoli di run-off.
- 4) Ricampionare con reimmissione i residui creando il triangolo di run-off dei residui e, in seguito, da questi calcolare le grandezze cumulate.
- 5) Implementare il metodo Chain-Ladder e la formula di Mack per calcolare le grandezze di interesse: il rapporto tra la radice quadrata dell'estimation variance e la stima Chain-Ladder individua la standard deviation del fattore di disturbo in esame.

I dati utilizzati nella funzione di verosimiglianza della formula di Bayes, quindi, sono ottenuti ricorrendo a due metodologie ampiamente utilizzate in letteratura, quali il bootstrap e la formula di Mack applicate ai triangoli di run-off del numero cumulato e costo medio cumulato. Al fine di potere considerare tutta l'informazione disponibile, inoltre, l'approccio Mack-Bootstrap è implementato sui triangoli di run-off di differenti dimensioni; partendo dai triangoli di run-off di dimensioni $N \times N$, associati all'ultimo anno di bilancio disponibile, l'approccio è applicato ai triangoli ottenuti eliminando una diagonale alla volta. La procedura, quindi, è applicata ai triangoli di run-off di differenti anni contabili. Considerando gli ultimi l anni di calendario, i triangoli di run-off hanno dimensione pari a $(N-l+1 \times N-l+1), \dots, (N \times N)$; la procedura Mack-Bootstrap, quindi, permette di ottenere per ogni triangolo storico considerato la distribuzione campionaria delle grandezze $\tilde{\sigma}_{\tilde{q}}$ e $\tilde{\sigma}_{\tilde{p}}$. I valori attesi di tali distribuzioni campionarie, che individuano la variabilità media della stima associata ai parametri, sono adottati come stima del valore vero ed incognito delle v.a. $\tilde{\sigma}_{\tilde{q}}$ e $\tilde{\sigma}_{\tilde{p}}$ e costituiscono i dati in cui le funzioni di verosimiglianza $f(\sigma|a, b)$, basate sul modello distributivo Gamma, sono calcolate. Gli l valori di $E(\tilde{\sigma}_{\tilde{q}})$ e $E(\tilde{\sigma}_{\tilde{p}})$, quindi, sono utilizzati come dati con cui è calcolata la funzione di verosimiglianza; tali grandezze sono ipotizzate essere indipendenti ed identicamente distribuite. Si noti che la prima assunzione, utile nel calcolo della verosimiglianza, non è completamente rispettata nella pratica in quanto i dati sono ottenuti dai triangoli storici che condividono delle medesime celle. Anche l'ipotesi di identica distribuzione non è necessariamente soddisfatta, in quanto il modello in esame non presenta condizioni volte a garantire tale condizione.

Le distribuzioni a priori associate ai parametri, definiti per valori positivi, sono scelte quali distribuzioni non informative, al fine di evitare qualsiasi sorta di expert judgment. In particolare, le analisi sono svolte considerando sia le prior uniformi che le distribuzioni a priori di Jeffreys; queste ultime dipendono dal

modello prescelto per la funzione di verosimiglianza e sono invarianti rispetto a riparametrizzazione dei parametri del modello. La forma di tali distribuzioni a priori associate ad un modello Gamma a due parametri la si trova in Moala, Ramos e Achcar (2013) e, considerando una generica v.a. $\text{Gamma}(\alpha; \beta)$, è pari a:

$$\pi_J(\alpha, \beta) = \sqrt{\alpha\psi'(\alpha) - 1/\beta}$$

dove ψ' individua la derivata della funzione trigamma, con $\psi(k) = \frac{\Gamma'(k)}{\Gamma(k)}$.

Al fine di ottenere i dati con cui valutare la funzione di verosimiglianza, l'approccio Mack-Bootstrap descritto in precedenza è stato applicato ai triangoli di run-off del numero cumulato dei sinistri e dei rispettivi costi medi cumulati delle 4 linee di business oggetto di studio. I triangoli storici considerati afferiscono agli ultimi nove anni di calendario ($l = 9$) ed hanno dimensione $(N \times N)$ da 4×4 a 12×12 . Il numero di iterazioni utilizzate nella procedura bootstrap al fine di ottenere le distribuzioni campionarie di $\tilde{\sigma}_{\tilde{q}}$ e $\tilde{\sigma}_{\tilde{p}}$ è pari a 10.000. Le tabelle seguenti riportano per le Lob considerate i valori attesi e i quantili ai livelli 5% e 95% delle distribuzioni campionarie delle standard deviation dei due fattori di disturbo.

Lob: RCA

Dimensione (N)		4	5	6	7	8	9	10	11	12
$\tilde{\sigma}_{\tilde{q}}$	Media	4,16%	2,86%	2,16%	2,11%	1,95%	1,98%	2,25%	2,69%	2,88%
	$q(5\%)$	1,33%	1,15%	1,05%	1,18%	1,21%	1,28%	1,49%	1,82%	2,11%
	$q(95\%)$	7,78%	5,14%	3,65%	3,32%	2,95%	2,92%	3,32%	3,98%	4,12%
$\tilde{\sigma}_{\tilde{p}}$	Media	3,03%	3,19%	2,49%	2,39%	2,02%	2,11%	1,97%	2,22%	2,17%
	$q(5\%)$	0,79%	1,29%	1,22%	1,27%	1,13%	1,24%	1,23%	1,40%	1,49%
	$q(95\%)$	7,13%	6,38%	4,58%	4,18%	3,45%	3,56%	3,18%	3,61%	3,23%

Tabella 4.40: Lob RCA - Valore atteso, quantili al 5% e 95% delle v.a. $\tilde{\sigma}_{\tilde{q}}$ e $\tilde{\sigma}_{\tilde{p}}$ relativi ai triangoli storici di dimensione da 4×4 a 12×12 .

Lob: RCG

Dimensione (N)		4	5	6	7	8	9	10	11	12
$\tilde{\sigma}_{\tilde{q}}$	Media	4,49%	3,79%	2,99%	3,54%	3,62%	3,40%	3,71%	4,35%	4,89%
	$q(5\%)$	1,14%	1,80%	1,65%	2,07%	2,32%	2,24%	2,28%	2,78%	3,28%
	$q(95\%)$	10,53%	6,20%	4,78%	5,74%	5,58%	5,21%	6,20%	6,87%	7,38%
$\tilde{\sigma}_{\tilde{p}}$	Media	2,63%	2,31%	2,24%	1,98%	2,29%	2,08%	1,96%	2,12%	2,24%
	$q(5\%)$	0,83%	0,82%	1,04%	1,01%	1,20%	1,21%	1,17%	1,34%	1,50%
	$q(95\%)$	5,72%	4,78%	4,20%	3,55%	4,07%	3,50%	3,21%	3,38%	3,40%

Tabella 4.41: Lob RCG - Valore atteso, quantili al 5% e 95% delle v.a. $\tilde{\sigma}_{\tilde{q}}$ e $\tilde{\sigma}_{\tilde{p}}$ relativi ai triangoli storici di dimensione da 4×4 a 12×12 .

		Lob: IEN									
Dimensione (N)		4	5	6	7	8	9	10	11	12	
$\tilde{\sigma}_{\tilde{q}}$	Media	5,99%	5,64%	4,32%	4,30%	4,15%	4,28%	3,45%	3,76%	3,78%	
	$q(5\%)$	1,89%	2,65%	2,41%	2,60%	2,65%	2,67%	2,30%	2,46%	2,55%	
	$q(95\%)$	11,22%	9,46%	6,63%	6,22%	5,83%	6,53%	4,87%	5,70%	5,69%	
$\tilde{\sigma}_{\tilde{p}}$	Media	2,63%	5,01%	4,82%	4,86%	4,68%	4,93%	4,24%	4,65%	4,14%	
	$q(5\%)$	0,83%	1,91%	2,23%	2,74%	2,75%	2,93%	2,91%	3,15%	2,96%	
	$q(95\%)$	5,72%	9,66%	7,91%	7,47%	7,34%	8,16%	6,14%	7,22%	5,95%	

Tabella 4.42: Lob IEN - Valore atteso, quantili al 5% e 95% delle v.a. $\tilde{\sigma}_{\tilde{q}}$ e $\tilde{\sigma}_{\tilde{p}}$ relativi ai triangoli storici di dimensione da 4×4 a 12×12 .

		Lob: ADB									
Dimensione (N)		4	5	6	7	8	9	10	11	12	
$\tilde{\sigma}_{\tilde{q}}$	Media	4,74%	3,49%	3,70%	3,65%	3,45%	3,64%	3,63%	3,75%	4,33%	
	$q(5\%)$	1,17%	1,41%	1,79%	1,96%	1,99%	2,10%	2,20%	2,26%	2,53%	
	$q(95\%)$	8,79%	6,18%	6,11%	5,76%	5,26%	5,93%	5,73%	6,12%	7,52%	
$\tilde{\sigma}_{\tilde{p}}$	Media	2,82%	2,79%	3,63%	3,40%	4,51%	4,94%	5,36%	5,06%	5,58%	
	$q(5\%)$	0,92%	1,34%	2,06%	2,16%	2,77%	2,90%	3,22%	2,93%	3,23%	
	$q(95\%)$	5,61%	4,66%	5,81%	5,15%	7,02%	8,26%	8,69%	8,47%	10,03%	

Tabella 4.43: Lob ADB - Valore atteso, quantili al 5% e 95% delle v.a. $\tilde{\sigma}_{\tilde{q}}$ e $\tilde{\sigma}_{\tilde{p}}$ relativi ai triangoli storici di dimensione da 4×4 a 12×12 .

Dopo avere scelto le distribuzioni a priori e avere ottenuto i dati su cui valutare la funzione di verosimiglianza è possibile calcolare la distribuzione a posteriori. Quest'ultima, nel caso in esame, non possiede forma analitica chiusa; al fine di ottenere un campione (denominato in letteratura anche *chain*, ossia *catena*) da tale distribuzione, si è fatto ricorso all'algoritmo di Metropolis-Hastings⁵. Tale procedura Markov Chain Monte Carlo (MCMC) permette di campionare dalla distribuzione a posteriori attraverso una funzione proporzionale rispetto alla vera funzione di densità, evitando, quindi, di calcolarne la costante di normalizzazione. In particolare, l'algoritmo adottato è un Random Walk Metropolis-Hastings; a fronte di opportune analisi di convergenza, è stato, quindi, possibile ottenere una catena stazionaria di valori con cui valutare la distribuzione a posteriori (si veda Appendice B).

Grazie alle distribuzioni a posteriori è possibile calibrare le v.a. $\tilde{\sigma}_{\tilde{q}}$ e $\tilde{\sigma}_{\tilde{p}}$ i cui parametri, come mostrato nella relazione (4.3.1.1.2), risultano essere il vettore dei valori attesi della distribuzione a posteriori. Considerando le 4 Lob, i valori attesi delle distribuzioni a posteriori associati alle v.a. della standard deviaton del numero e costo dei sinistri risultano simili per entrambe le tipologie di distribuzioni a priori utilizzate (si veda Appendice B). La calibrazione delle v.a.

⁵Una trattazione completa del tema la si trova, ad esempio, in Gamerman e Lopes (2006) e Robert e Casella (2010).

$\tilde{\sigma}_{\tilde{q}}$ e $\tilde{\sigma}_{\tilde{p}}$, quindi, non risente in modo sostanziale dalla famiglia di distribuzione a priori utilizzata. Nelle tabelle seguenti si riportano, per le Lob considerate, i valori attesi, il coefficiente di variazione e l'indice di asimmetria delle variabili $\tilde{\sigma}_{\tilde{q}}$ e $\tilde{\sigma}_{\tilde{p}}$.

Lob: RCA

Variabile aleatoria	$\tilde{\sigma}_{\tilde{q}}$		$\tilde{\sigma}_{\tilde{p}}$	
	Uniforme	Jeffreys	Uniforme	Jeffreys
Prior				
Valore atteso	2,547%	2,560%	2,394%	2,399%
Coeff. di variazione	20,632%	23,801%	14,296%	16,493%
Asimmetria	0,413	0,476	0,286	0,330

Tabella 4.44: Lob RCA - Valore atteso, coefficiente di variazione e indice di asimmetria, al variare delle distribuzioni prior considerate, delle v.a. $\tilde{\sigma}_{\tilde{q}}$ e $\tilde{\sigma}_{\tilde{p}}$.

Lob: RCG

Variabile aleatoria	$\tilde{\sigma}_{\tilde{q}}$		$\tilde{\sigma}_{\tilde{p}}$	
	Uniforme	Jeffreys	Uniforme	Jeffreys
Prior				
Valore atteso	3,859%	3,866%	2,204%	2,205%
Coeff. di variazione	12,543%	14,469%	7,512%	8,671%
Asimmetria	0,251	0,289	0,150	0,173

Tabella 4.45: Lob RCG - Valore atteso, coefficiente di variazione e indice di asimmetria, al variare delle distribuzioni prior considerate, delle v.a. $\tilde{\sigma}_{\tilde{q}}$ e $\tilde{\sigma}_{\tilde{p}}$.

Lob: INE

Variabile aleatoria	$\tilde{\sigma}_{\tilde{q}}$		$\tilde{\sigma}_{\tilde{p}}$	
	Uniforme	Jeffreys	Uniforme	Jeffreys
Prior				
Valore atteso	4,398%	4,410%	4,430%	4,441%
Coeff. di variazione	15,217%	17,571%	15,420%	17,784%
Asimmetria	0,304	0,351	0,308	0,356

Tabella 4.46: Lob INE - Valore atteso, coefficiente di variazione e indice di asimmetria, al variare delle distribuzioni prior considerate, delle v.a. $\tilde{\sigma}_{\tilde{q}}$ e $\tilde{\sigma}_{\tilde{p}}$.

Lob: ADB

Variabile aleatoria	$\tilde{\sigma}_{\tilde{q}}$		$\tilde{\sigma}_{\tilde{p}}$	
	Uniforme	Jeffreys	Uniforme	Jeffreys
Prior				
Valore atteso	3,817%	3,821%	4,212%	4,234%
Coeff. di variazione	8,788%	10,150%	21,737%	25,040%
Asimmetria	0,176	0,203	0,435	0,501

Tabella 4.47: Lob ADB - Valore atteso, coefficiente di variazione e indice di asimmetria, al variare delle distribuzioni prior considerate, delle v.a. $\tilde{\sigma}_{\tilde{q}}$ e $\tilde{\sigma}_{\tilde{p}}$.

La calibrazione delle variabili $\tilde{\sigma}_{\tilde{q}}$ e $\tilde{\sigma}_{\tilde{p}}$ permette di stimare i momenti dei fattori di disturbo presenti nel modello CRM. Le variabili aleatorie \tilde{q} e \tilde{p} sono ipotizzate essere delle variabili Gamma a parametri uguali, in particolare:

$$\tilde{q} \sim \text{Gamma} \left(\frac{1}{\tilde{\sigma}_{\tilde{q}}^2}; \frac{1}{\tilde{\sigma}_{\tilde{q}}^2} \right) \quad \tilde{p} \sim \text{Gamma} \left(\frac{1}{\tilde{\sigma}_{\tilde{p}}^2}; \frac{1}{\tilde{\sigma}_{\tilde{p}}^2} \right)$$

dove,

$$- \tilde{\sigma}_{\tilde{q}} \sim \text{Gamma} \left(E(\tilde{A}_{\tilde{q}}^{post}); E(\tilde{B}_{\tilde{q}}^{post}) \right);$$

$$- \tilde{\sigma}_{\tilde{p}} \sim \text{Gamma} \left(E(\tilde{A}_{\tilde{p}}^{post}); E(\tilde{B}_{\tilde{p}}^{post}) \right);$$

in cui con $E(\tilde{A}^{post})$ e $E(\tilde{B}^{post})$ si sono indicati i valori attesi marginali della distribuzione a posteriori, $E(\tilde{A}|\tilde{\sigma})$ e $E(\tilde{B}|\tilde{\sigma})$.

I momenti delle v.a. \tilde{q} e \tilde{p} dipendono rispettivamente dalle variabili misturanti $\tilde{\sigma}_{\tilde{q}}$ e $\tilde{\sigma}_{\tilde{p}}$, che tuttavia non ne modificano il valore atteso che è pari all'unità. Considerando per semplicità la sola v.a. $\tilde{\sigma}_{\tilde{q}}$, valendo considerazioni analoghe per la v.a. $\tilde{\sigma}_{\tilde{p}}$, la varianza del fattore di disturbo risulta essere (si veda Appendice C):

$$\text{Var}(\tilde{q}) = \frac{E(\tilde{A}_{\tilde{q}}^{post}) [E(\tilde{A}_{\tilde{q}}^{post}) + 1]}{E(\tilde{B}_{\tilde{q}}^{post})^2}$$

Il coefficiente di variazione, pari alla radice quadrata della varianza, è:

$$\text{CV}(\tilde{q}) = \frac{\sqrt{E(\tilde{A}_{\tilde{q}}^{post}) [E(\tilde{A}_{\tilde{q}}^{post}) + 1]}}{E(\tilde{B}_{\tilde{q}}^{post})}$$

Infine, l'indice di asimmetria, pari al rapporto tra il momento centrale terzo e la standard deviation al cubo, risulta essere:

$$\gamma(\tilde{q}) = \frac{2 [E(\tilde{A}_{\tilde{q}}^{post}) + 2] [E(\tilde{A}_{\tilde{q}}^{post}) + 3]}{E(\tilde{B}_{\tilde{q}}^{post}) \sqrt{E(\tilde{B}_{\tilde{q}}^{post}) [E(\tilde{B}_{\tilde{q}}^{post}) + 1]}}$$

Le tabelle seguenti riportano, per le quattro Lob oggetto di studio, le caratteristiche dei fattori disturbo al variare delle distribuzioni a priori considerate nell'algoritmo bayesiano.

Lob: RCA

Prior	Uniforme		Jeffreys	
Fattore di disturbo	\tilde{q}	\tilde{p}	\tilde{q}	\tilde{p}
Valore atteso	1	1	1	1
Coeff. di variazione	2,601%	2,631%	2,418%	2,432%
Asimmetria	0,061	0,065	0,052	0,054

Tabella 4.48: Lob RCA - Valore atteso, coefficiente di variazione e asimmetria dei fattori di disturbo \tilde{q} e \tilde{p} .**Lob: RCG**

Prior	Uniforme		Jeffreys	
Fattore di disturbo	\tilde{q}	\tilde{p}	\tilde{q}	\tilde{p}
Valore atteso	1	1	1	1
Coeff. di variazione	3,889%	3,906%	2,210%	2,213%
Asimmetria	0,083	0,085	0,045	0,046

Tabella 4.49: Lob RCG - Valore atteso, coefficiente di variazione e asimmetria dei fattori di disturbo \tilde{q} e \tilde{p} .**Lob: IEN**

Prior	Uniforme		Jeffreys	
Fattore di disturbo	\tilde{q}	\tilde{p}	\tilde{q}	\tilde{p}
Valore atteso	1	1	1	1
Coeff. di variazione	4,449%	4,477%	4,482%	4,511%
Asimmetria	0,097	0,101	0,098	0,102

Tabella 4.50: Lob IEN - Valore atteso, coefficiente di variazione e asimmetria dei fattori di disturbo \tilde{q} e \tilde{p} .**Lob: ADB**

Prior	Uniforme		Jeffreys	
Fattore di disturbo	\tilde{q}	\tilde{p}	\tilde{q}	\tilde{p}
Valore atteso	1	1	1	1
Coeff. di variazione	3,832%	3,840%	4,311%	4,364%
Asimmetria	0,079	0,080	0,103	0,110

Tabella 4.51: Lob ADB - Valore atteso, coefficiente di variazione e asimmetria dei fattori di disturbo \tilde{q} e \tilde{p} .

Considerando il fattore di disturbo \tilde{q} , che è la variabile strutturale associata al numero dei sinistri, la linea di business che mostra il valore di standard deviation più elevato, per entrambe le tipologie di prior, è il ramo incendio. Il ramo meno affetto dal fattore disturbo, invece, risulta essere RCA. Considerazioni analoghe valgono per la variabile casuale \tilde{p} ; anche in tale circostanza la Lob incendio risulta

quella caratterizzata dalla maggiore magnitudo del fattore di disturbo. Il ramo a cui corrisponde la minore variabilità di parametro, invece, risulta essere la linea di business RCG.

4.3.1.2 La stima bayesiana della correlazione lineare tra i fattori di disturbo

Il Collective Risk Model assume che la variabile aleatoria del numero dei sinistri e quella che individua i rispettivi importi siano tra loro indipendenti in ogni cella (i, j) del triangolo inferiore di run-off. Tale ipotesi, tuttavia, nella pratica non è rispettata a priori a causa della dipendenza indotta dal metodo deterministico sottostante alla stima dei parametri (in tale contesto costituito dal Frequency-Severity). Di seguito si presenta un approccio bayesiano per quantificare il coefficiente di correlazione lineare tra il numero e il costo dei sinistri stimato sulle variabili aleatorie dei fattori di disturbo (si veda Ricotta e Luini (2019)). In tale contesto, quindi, le v.a. \tilde{q} e \tilde{p} sono utilizzate per stimare e introdurre la dipendenza tra le grandezze caratteristiche del modello CRM, date dalle variabili $\tilde{K}_{i,j}$ e $\tilde{Z}_{h,i,j}$.

Analogamente a quanto sviluppato nel paragrafo 4.3.1.1, il processo di stima della dipendenza si basa sull'utilizzo della formula di Mack e dell'approccio bootstrap; quest'ultimo tuttavia, ricampionando in modo congiunto le celle che occupano la stessa posizione nei triangoli del numero cumulato e del costo medio cumulato dei sinistri, permette di tenere implicitamente conto della dipendenza esistente tra i due triangoli di run-off. L'obiettivo, quindi, consiste nell'ottenere le distribuzioni marginali delle v.a. $\tilde{\sigma}_{\tilde{q}}$ e $\tilde{\sigma}_{\tilde{p}}$ che tengano implicitamente conto della dipendenza esistente tra il data set del numero cumulato dei sinistri e del costo medio cumulato degli stessi. In seguito, la correlazione lineare quantificata tramite l'approccio bayesiano è utilizzata per costruire una copula Normale bidimensionale di variabili marginali \tilde{q} e \tilde{p} , calibrate nel paragrafo precedente.

Nel contesto bayesiano, il coefficiente di correlazione lineare di Pearson è interpretato come una grandezza aleatoria e distribuita in accordo con una v.a. Beta; si ipotizza, quindi, che la dipendenza tra il numero e il costo dei sinistri sia definita nell'intervallo $[0, 1]$:

$$\tilde{\rho} \sim \text{Beta}(\tilde{C}; \tilde{D})$$

Le v.a. \tilde{C} e \tilde{D} rappresentano le distribuzioni a priori del modello. In base alla formula di Bayes, la distribuzione a posteriori dei parametri da cui la v.a. $\tilde{\rho}$ dipende, è la seguente:

$$f(c, d|\rho) \propto f(\rho|c, d)f(c)f(d) \quad (4.3.1.2.1)$$

In seguito, la v.a. $\tilde{\rho}$ è calibrata ricorrendo al vettore del valore atteso della distribuzione a posteriori,

$$\tilde{\rho} \sim \text{Beta} \left(E(\tilde{C}|\tilde{\rho}); E(\tilde{D}|\tilde{\rho}) \right) \quad (4.3.1.2.2)$$

Infine, la media della v.a. della relazione (4.3.1.2.2) è adottata come parametro caratteristico della copula Normale utilizzata per correlare le variabili marginali \tilde{q} e \tilde{p} .

Le grandezze su cui è valutata la funzione di verosimiglianza della relazione (4.3.1.2.1) coincidono con i coefficienti di correlazione lineare, ottenuti ricorrendo all'approccio Mack-Bootstrap dipendente, calcolati tra le distribuzioni delle variabili $\tilde{\sigma}_{\tilde{q}}$ e $\tilde{\sigma}_{\tilde{p}}$ ottenute sugli $l = 9$ triangoli di run-off storici. Per le distribuzioni a priori della formula di Bayes si è fatto uso sia della distribuzione uniforme che della distribuzione di Jeffreys; in quest'ultima circostanza, la distribuzione a priori associata ad un modello $\text{Beta}(\alpha, \beta)$ risulta essere (Yang e Berger (1997)):

$$\pi_J(\alpha, \beta) = \sqrt{[\psi'(\alpha) - \psi'(\alpha + \beta)][\psi'(\beta) - \psi'(\alpha + \beta)] - [-\psi'(\alpha + \beta)]^2}$$

dove ψ' individua la derivata della funzione trigamma.

La tabella seguente riporta, per le quattro Lob considerate, i valori del coefficiente di correlazione ottenuti tramite l'approccio bootstrap dipendente esposto in precedenza e calcolati considerando i 9 triangoli di run-off storici.

Coefficiente di correlazione lineare di Pearson									
Dimensione (N)	4	5	6	7	8	9	10	11	12
RCA	0,428	0,367	0,461	0,577	0,546	0,405	0,313	0,110	0,103
RCG	0,369	0,108	0,221	0,028	0,082	0,040	0,029	0,183	0,236
IEN	0,641	0,043	0,256	0,294	0,254	0,344	0,212	0,235	0,239
ADB	0,132	0,046	0,140	0,036	0,004	0,084	0,029	0,052	0,091

Tabella 4.52: Correlazione lineare, per le Lob considerate, tra le v.a. $\tilde{\sigma}_{\tilde{q}}$ e $\tilde{\sigma}_{\tilde{p}}$ associate ai triangoli di run-off storici di dimensione da 4×4 a 12×12 .

La valutazione della distribuzione a posteriori è effettuata ricorrendo all'algoritmo MCMC Metropolis-Hastings. In particolare, l'algoritmo adottato è caratterizzato da una distribuzione *proposal* Random Walk; dopo opportune analisi di convergenza, è stato possibile valutare la distribuzioni a posteriori (si veda l'Appendice B per ulteriori dettagli). In seguito, come mostrato nella relazione (4.3.1.2.2), il vettore delle medie della distribuzione a posteriori è utilizzato per calibrare la v.a. $\tilde{\rho}$.

La tabella seguente riporta, per le Lob oggetto di studio, il coefficiente di correlazione lineare tra le due variabili sistematiche del modello CRM, mentre i grafici mostrano gli scatterplot delle distribuzioni bivariate dei fattori di disturbo delle quattro Lob.

Valore atteso della v.a. $\tilde{\rho}$		
Prior	Uniforme	Jeffreys
RCA	0,359	0,362
RCG	0,138	0,144
IEN	0,277	0,281
ADB	0,065	0,068

Tabella 4.53: Stima della correlazione lineare, per le Lob considerate, tra le v.a. \tilde{q} e \tilde{p} .

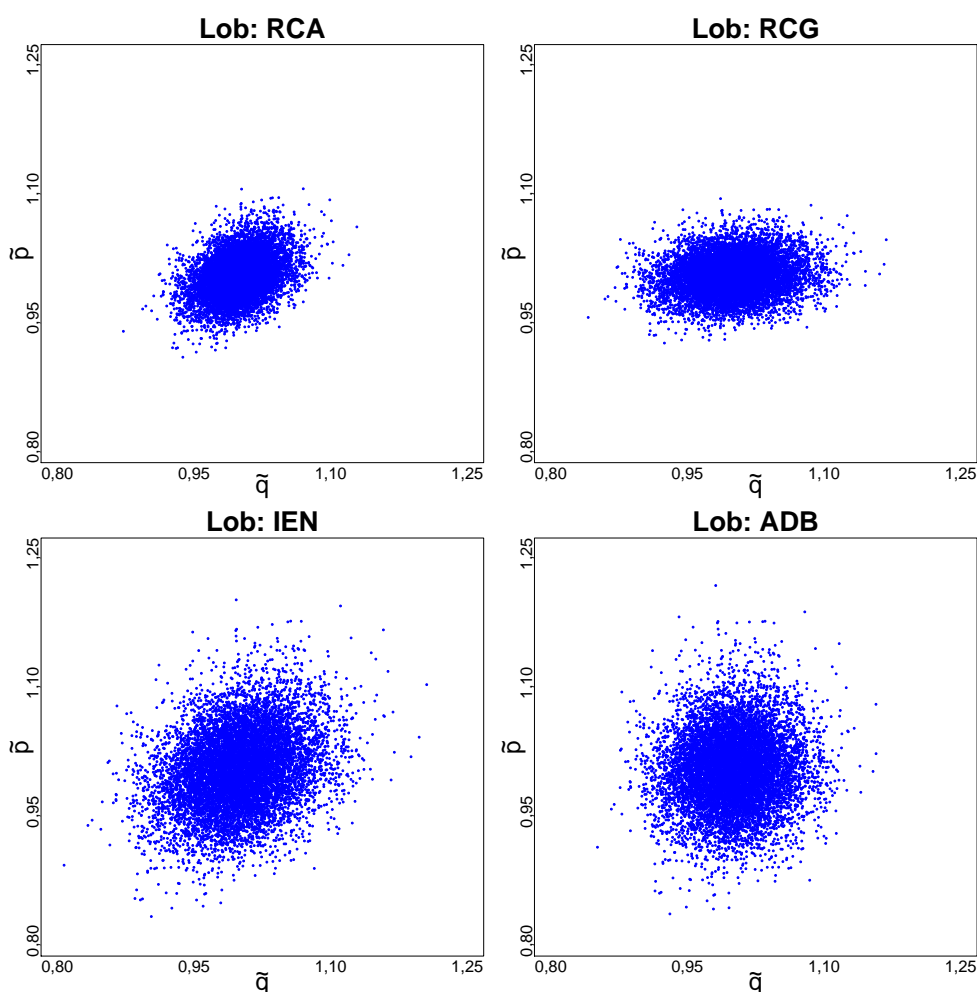


Figura 4.23: Scatterplot, ottenuti con 10.000 simulazioni, delle distribuzioni bidimensionali dei fattori di disturbo. Al fine di confrontare tra le varie Lob la dispersione dei fattori di disturbo marginali, i grafici sono riportati sulla stessa scala: $\tilde{q}, \tilde{p} \in [0, 80; 1, 25]$.

Come si evince dalla tabella precedente, per tutte le Lob considerate, i valori attesi della variabile casuale $\tilde{\rho}$ determinati considerando le due tipologie di distribuzioni a priori, uniforme e di Jeffreys, sono prossimi tra loro. Il ramo RCA mostra il maggiore grado di correlazione, mentre la linea di business ADB possiede il valore di correlazione lineare più contenuto.

I fattori di disturbo presenti nel Collective Risk Model, quindi, sono modellizzati per ogni Lob tramite una distribuzione bivariata meta-Gaussiana (si veda figura 4.23); le variabili marginali date dalle v.a. \tilde{q} e \tilde{p} calibrate come mostrato nel paragrafo precedente sono congiunte tra di loro tramite una copula Gaussiana di parametro ρ , stimato in precedenza.

4.3.1.3 Le distribuzioni marginali del reserve risk

La stima bayesiana dei fattori di disturbo ottenuta nei paragrafi 4.3.1.1 e 4.3.1.2 permette di implementare l'approccio Collective Risk Model; in particolare, nel seguito i fattori di disturbo considerati, nonché la loro correlazione, sono calibrate ricorrendo alle sole distribuzioni a priori uniformi. Le formule (4.3.1.1) e (4.3.1.2) permettono di valutare, per le singole Lob, in termini stocastici la riserva sinistri considerando un orizzonte temporale a totale run-off (\tilde{R}). Le distribuzioni marginali rappresentative del rischio di riservazione delle quattro linee di business, inoltre, sono valutate considerando un orizzonte temporale di 12 mesi futuri (\tilde{R}^{OY}). La distribuzione one-year del reserve risk è ottenuta tramite approccio re-reserving; il requisito di capitale da detenere a fini di solvibilità (SCR), infine, è ottenuto come differenza tra il quantile al livello di confidenza 99,5% della distribuzione, opportunamente scontato, e il valore corrente della best estimate sinistri, pari al valore attuale atteso della distribuzione.

La valutazione stocastica della riserva sinistri è effettuata, sia per l'orizzonte temporale a totale run-off che per quello one-year, tramite simulazioni Monte Carlo. A differenza di quanto mostrato in Ricotta e Clemente (2016), la dipendenza tra i fattori di disturbo non permette di ricavare, ad eccezione del valore atteso, espressioni analitiche per le caratteristiche della riserva sinistri. Al fine di individuare il numero adeguato di simulazioni da utilizzare per valutare la distribuzione della riserva sinistri è stata sviluppata un'analisi preliminare di convergenza del metodo Monte Carlo per l'algoritmo CRM considerando le quattro linee di business. In particolare, sono stati eseguiti $M = 100$ run simulativi distinti della v.a. \tilde{R} con un numero crescente di simulazioni adottate (H); partendo da $H_1 = 1.000$ simulazioni e con incrementi di 1.000 si è giunti fino ad un numero di realizzazioni simulate pari a $H_{100} = 100.000$. Gli step simulativi, quindi, in totale sono pari a $N = H_{100}/1.000 = 100$ per ogni run M -esimo. Per ognuno di questi, sono state calcolate le caratteristiche di interesse della v.a. \tilde{R} , quali il valore atteso, la standard deviation e l'asimmetria. Il numero complessivo di esecuzioni Monte Carlo, quindi, determina una matrice di M righe e N colonne, $M \times N = 100 \times 100 = 10.000$, contenente la caratteristica in esame della v.a. della riserva sinistri. La standard deviation campionaria calcolata sugli N vettori delle caratteristiche simulate al variare del numero di simulazioni permette di valutare l'errore relativo associato alla stima Monte Carlo commesso al crescere del numero di simulazioni. In particolare, l'errore relativo, in tale contesto, è inteso come il coefficiente di variazione del vettore N_i considerato, con $i = 1, \dots, 100$, dove lo scarto quadratico medio è pari alla standard deviation campionaria calcolata

sul vettore e , analogamente, il valore atteso è definito come la media del vettore in esame. Le figure seguenti riportano, a titolo esemplificativo per la sola Lob RCA, l'output dell'analisi di convergenza del Monte Carlo descritta precedentemente e permettono di apprezzare la diminuzione della variabilità della grandezza considerata al crescere del numero H di simulazioni effettuate.

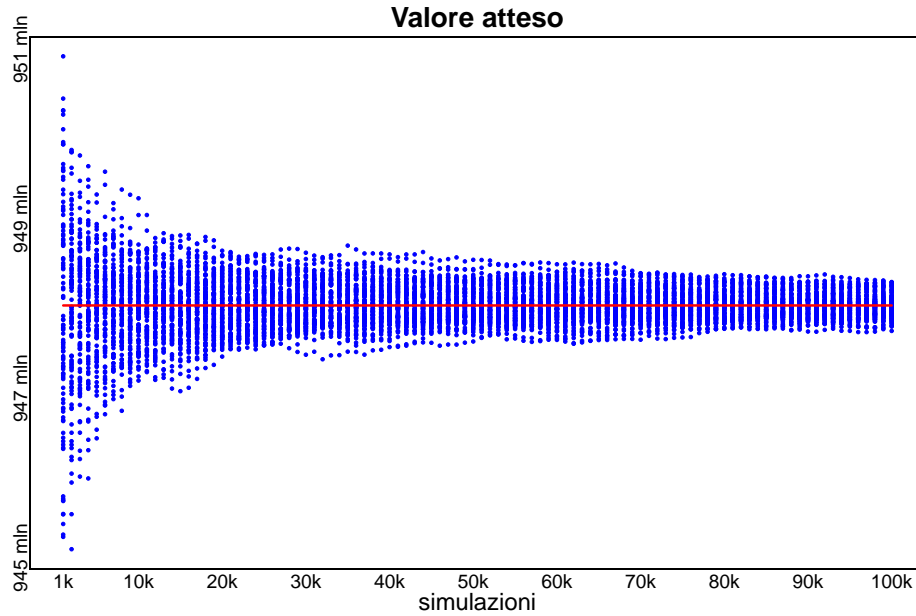


Figura 4.24: $M \times N = 10.000$ punti simulati del valore atteso della v.a. \tilde{R} per la Lob RCA; la linea orizzontale rossa corrisponde alla media del vettore N_{100} .

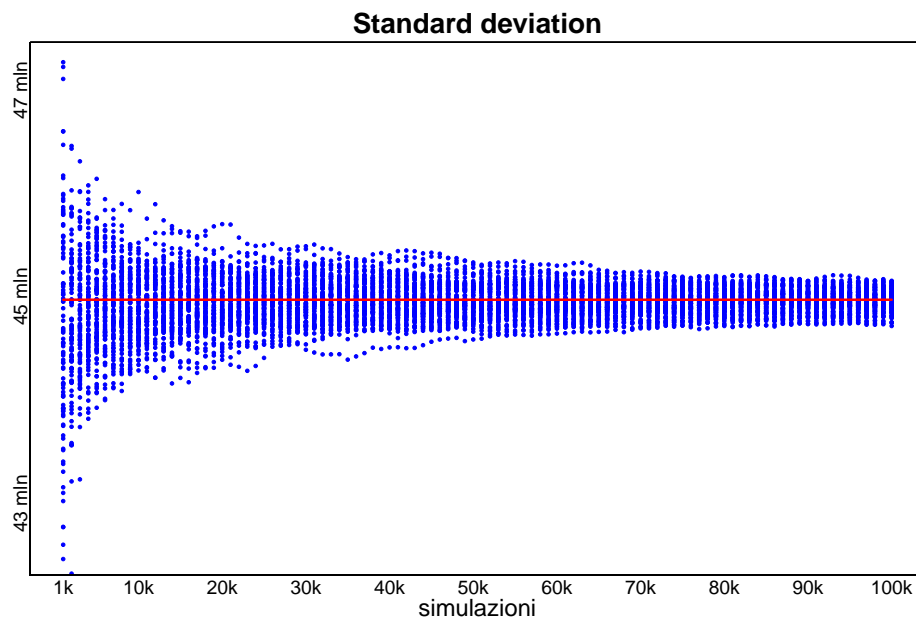


Figura 4.25: $M \times N = 10.000$ punti simulati dello scarto quadratico medio della v.a. \tilde{R} per la Lob RCA; la linea orizzontale rossa corrisponde alla media del vettore N_{100} .

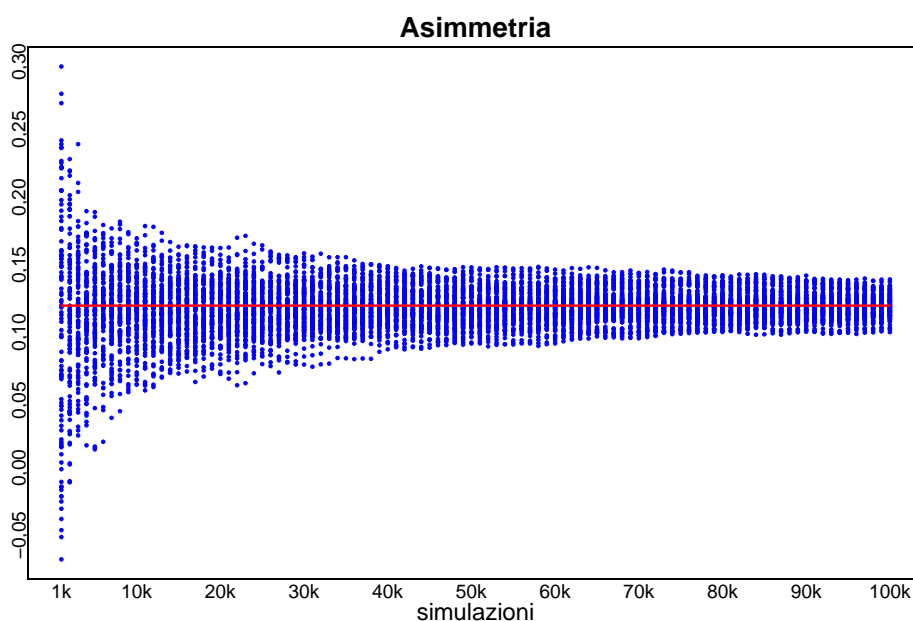


Figura 4.26: $M \times N = 10.000$ punti simulati dell'indice di asimmetria della v.a. \tilde{R} per la Lob RCA; la linea orizzontale rossa corrisponde alla media del vettore N_{100} .

La tabella seguente riporta per le quattro Lob considerate l'errore relativo valutato con $H_1 = 1.000$ e $H_{100} = 100.000$ simulazioni.

Lob	H	Valore atteso	Standard deviation	Asimmetria
RCA	1.000	0,14%	2,40%	72,30%
	100.000	0,01%	0,23%	6,88%
RCG	1.000	0,23%	2,48%	55,39%
	100.000	0,02%	0,23%	5,21%
IEN	1.000	0,35%	2,19%	28,71%
	100.000	0,03%	0,27%	2,72%
ADB	1.000	0,44%	2,19%	23,50%
	100.000	0,04%	0,27%	2,01%

Tabella 4.54: Errore relativo, basato su $M = 100$ run, associato alla media, standard deviation e asimmetria della variabile \tilde{R} ottenute tramite simulazioni Monte Carlo per $H_1 = 1.000$ e $H_{100} = 100.000$ simulazioni.

Si noti come l'errore relativo determinato con 100.000 simulazioni è circa 10 volte inferiore a quello calcolato con 1.000 simulazioni; come noto, infatti, la precisione dello stimatore Monte Carlo aumenta al crescere delle simulazioni di un fattore pari a \sqrt{H} . A fronte dei risultati presentati in precedenza, ai fini delle analisi seguenti è stato prescelto un numero di simulazioni pari a 100.000.

Di seguito si riportano, per le quattro Lob considerate, le caratteristiche simulate della riserva sinistri valutata a totale run-off e considerando un orizzonte temporale di 12 mesi.

Lob	Orizzonte temporale	Valore atteso	Standard deviation	Coeff. di variazione	Asimmetria
RCA	Tot. run-off	948.062	45.040	4,75%	0,113
	One-year	948.074	33.876	3,57%	0,172
RCG	Tot. run-off	332.050	23.997	7,23%	0,151
	One-year	331.998	23.145	6,97%	0,291
IEN	Tot. run-off	136.900	15.354	11,22%	0,293
	One-year	136.828	12.625	9,23%	0,398
ADB	Tot. run-off	41.549	5.819	14,01%	0,480
	One-year	41.539	5.095	12,27%	0,679

Tabella 4.55: Valore atteso, standard deviation, coefficiente di variazione e asimmetria della riserva sinistri delle quattro Lob valutata a totale run-off (\tilde{R}) e con un orizzonte temporale one-year (\tilde{R}^{OY}) (gli importi monetari sono espressi in migliaia di euro).

La variabilità relativa della distribuzione della riserva sinistri valutata a totale run-off è maggiore rispetto a quella della riserva valutata su un orizzonte temporale annuo; di contro, quest'ultima possiede un'asimmetria più marcata rispetto alla riserva sinistri valutata considerando l'intera durata di vita dei sinistri. La variabilità della distribuzione della riserva one-year rappresenta una percentuale consistente di quella a totale run-off; il rapporto tra il coefficiente di variazione one-year rispetto a quello a totale run-off, infatti, si attesta rispettivamente al 75%, 96%, 82% e 87% per le Lob RCA, RCG, IEN e ADB.

Il confronto tra le diverse linee di business mostra come la variabilità relativa decresca al crescere della dimensione della riserva sinistri; in particolare, al crescere del parametro dimensionale della Lob, dato dal numero dei sinistri riservati (si veda tabella 4.1), aumenta l'effetto di diversificazione che agisce sulla componente di rischio non ascrivibile ai fattori di disturbo, che rappresentano la quota di rischio sistematico all'interno del processo CRM (Ricotta e Clemente (2016)). Anche l'indice di asimmetria della riserva sinistri mostra un andamento analogo, risultando decrescente rispetto alla dimensione del ramo.

La tabella seguente mostra, per le diverse Lob valutate su un orizzonte annuo, il quantile al livello di confidenza 99,5%, il Tail VaR al livello 99%, il Solvency Capital Requirement e il rapporto tra il SCR e il valore della best estimate, definito SCR ratio.

Lob	q_{99,5%}	TVaR_{99%}	SCR	SCR ratio
RCA	1.042.212	1.045.725	94.137	9,93%
RCG	396.814	399.680	64.816	19,52%
IEN	174.292	176.054	37.464	27,38%
ADB	58.177	59.086	16.638	40,05%

Tabella 4.56: Quantile 99,5%, Tail VaR al 99%, SCR e SCR ratio della riserva sinistri per le quattro Lob valutata su un orizzonte temporale annuo (gli importi monetari sono espressi in migliaia di euro).

Il valore di SCR ratio è crescente al diminuire della dimensione del ramo a causa della maggiore variabilità e asimmetria della distribuzione della riserva; il ramo RCA mostra il minore valore di SCR ratio, pari al 9,93%, mentre il ramo ADB, che è quello caratterizzato dalla minore dimensione, possiede un SCR ratio pari al 40,05%, risultando essere la Lob con il maggiore assorbimento di capitale.

La normativa europea Solvency II fornisce, attraverso la standard formula, i parametri e la formula di calcolo per determinare il requisito di capitale associato al premium e reserve risk. Questo è definito come il prodotto tra una costante ($k = 3$), che funge da moltiplicatore, la variabilità relativa associata al premium e reserve risk delle Lob aggregate tra loro, definita σ , e la misura di volume dei rami (si veda Capitolo 1):

$$SCR_{nl,prem,res} = 3 \cdot \sigma_{nl} \cdot V_{nl}$$

Applicando la precedente formula al solo contesto del rischio di riservazione e considerando una singola Lob è, quindi, possibile quantificare il valore del SCR marginale per ogni singolo ramo coerentemente con la logica della standard formula. La tabella 4.58 permette di confrontare, per le singole Lob, le peculiarità del modello CRM con quelle della standard formula; in particolare, le grandezze considerate sono le seguenti.

- 1) Il valore di SCR definito dal modello CRM:

$$SCR^{CRM} = q_{99,5\%}(\tilde{R}^{OY}) - E(\tilde{R}^{OY}) = k^{CRM} \cdot \sigma^{CRM} \cdot E(\tilde{R}^{OY})$$

dove $E(\tilde{R}^{OY})$ individua la best estimate della riserva sinistri, definita come il valore atteso della distribuzione one-year, σ^{CRM} e k^{CRM} il coefficiente di variazione e il moltiplicatore determinati dal modello CRM.

- 2) Il valore di SCR ratio determinato del modello CRM:

$$SCR\ ratio^{CRM} = \frac{SCR^{CRM}}{E(\tilde{R}^{OY})} = k^{CRM} \cdot \sigma^{CRM}$$

- 3) Il moltiplicatore definito implicitamente dall'approccio CRM:

$$k^{CRM} = \frac{SCR^{CRM}}{\sigma^{CRM} \cdot E(\tilde{R}^{OY})}$$

4) Il valore di SCR e SCR ratio definiti dalla standard formula:

$$SCR^{SF} = 3 \cdot \sigma^{SF} \cdot E(\tilde{R}^{OY})$$

$$SCR\ ratio^{SF} = \frac{SCR^{SF}}{E(\tilde{R}^{OY})} = 3 \cdot \sigma^{SF}$$

dove σ^{SF} rappresenta la variabilità relativa delle singole Lob definita dagli Atti Delegati, che per le Lob in esame risulta essere pari al 9% per RCA, 11% per RCG e 10% per i rami IEN e ADB, che nella normativa Solvency II sono considerati congiuntamente. Per una maggiore chiarezza espositiva, di seguito si riporta il confronto tra i valori di σ ottenuti con il modello CRM e quelli previsti dalla normativa Solvency II.

Lob	σ^{CRM}	σ^{SF}
RCA	3,57%	9,00%
RCG	6,97%	11,00%
IEN	9,23%	10,00%
ADB	12,27%	10,00%

Tabella 4.57: Confronto, per le Lob considerate, tra i coefficienti di variazione ottenuti tramite metodo CRM e quelli previsti dalla standard formula.

Lob	CRM			SF		
	SCR	SCR ratio	k	SCR	SCR ratio	k
RCA	94.137	9,93%	2,78	255.980	27,00%	3,00
RCG	64.816	19,52%	2,80	109.559	33,00%	3,00
IEN	37.464	27,38%	2,97	41.048	30,00%	3,00
ADB	16.638	40,05%	3,27	12.462	30,00%	3,00

Tabella 4.58: Confronto, per le Lob considerate, tra i valori di SCR, SCR ratio e del moltiplicatore ottenuti considerando il modello CRM e la standard formula (gli importi monetari sono espressi in migliaia di euro).

I valori di SCR e SCR ratio ottenuti con il modello CRM risultano per le Lob RCA, RCG e IEN inferiori rispetto a quelli calcolati tramite la standard formula. In particolare, per tali Lob, sia la variabilità relativa che il rispettivo moltiplicatore sono più contenuti rispetto a quelli previsti dalla normativa comunitaria. La Lob ADB, di contro, a causa della limitata dimensione del portafoglio, mostra un valore di σ e di k derivanti dal modello CRM più elevati rispetto a quelli previsti dalla standard formula.

La figura seguente mostra, per le quattro Lob oggetto di studio, le distribuzioni one-year della riserva sinistri ottenute tramite il modello CRM; gli istogrammi,

inoltre, evidenziano il valore di SCR, determinato come distanza tra la media e il quantile al livello di confidenza 99,5%.

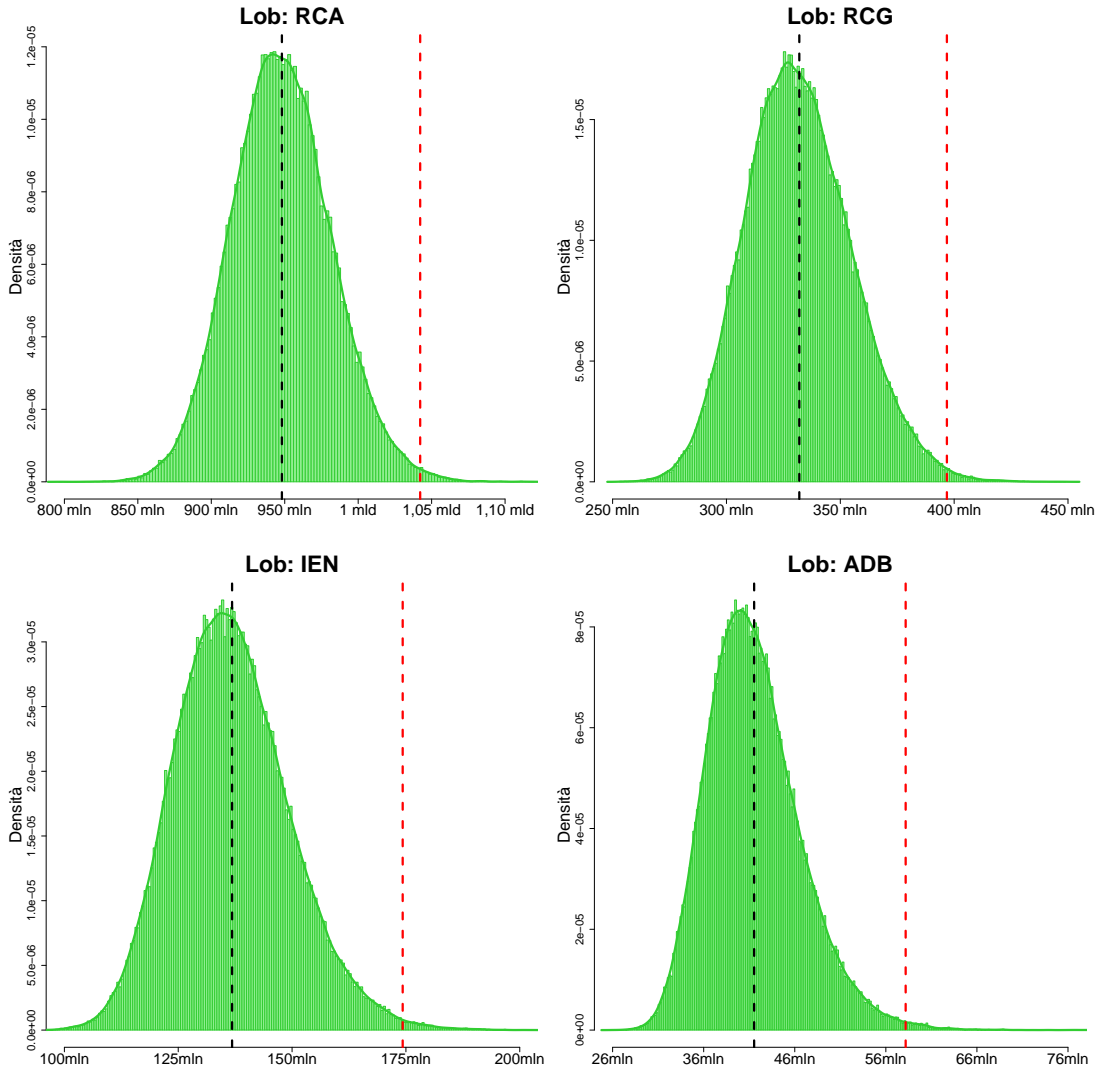


Figura 4.27: Distribuzioni one-year della riserva sinistri per le quattro Lob considerate; le linee verticali di colore nero e rosso individuano rispettivamente il valore atteso e il quantile al livello di confidenza 99,5%.

4.4 La stima del requisito patrimoniale di solvibilità complessivo per il reserve risk

Dopo avere quantificato il reserve risk delle singole Lob è possibile aggregare tali distribuzioni utilizzando la vine copula selezionata nel paragrafo 4.2.3, la quale modella la struttura di dipendenza esistente tra i rischi marginali. L'approccio CRM congiuntamente con la vine copula si qualificano, coerentemente con il contesto normativo Solvency II, come un modello interno parziale per quantificare il requisito patrimoniale di solvibilità afferente al rischio di riservazione danni di una compagnia di assicurazione.

Il SCR complessivo è valutato sulla variabile casuale data dalla somma dei quattro rischi marginali, modellizzati tramite CRM, sui quali è stata introdotta la struttura di dipendenza della vine copula selezionata. In particolare, simulando H realizzazioni casuali nell'intervallo $[0, 1]$ dal D-vine stimato e ricorrendo alla funzione quantilica dei quattro rischi marginali del reserve risk one-year modellizzati tramite CRM (\tilde{R}_i^{OY} , con $i = 1, \dots, 4$), si ottengono, per costruzione, quattro v.a. marginali che mantengono le caratteristiche distributive delle variabili marginali originarie ma con una struttura di dipendenza identica a quella del D-vine:

$$F_{\tilde{R}_i^{OY}}^{-1}(u_{i,h}) = r_{i,h}^{OY} \quad \text{con} \quad i = 1, \dots, 4; \quad h = 1, \dots, H; \quad u \in [0, 1]$$

dove,

- $F_{\tilde{R}_i^{OY}}$ individua la funzione di ripartizione della v.a. \tilde{R}_i^{OY} che rappresenta il reserve risk della i -esima Lob;
- $F_{\tilde{R}_i^{OY}}^{-1}$ rappresenta l'inversa della funzione di ripartizione della variabile \tilde{R}_i^{OY} , definita funzione quantilica;
- $u_{i,h}$ costituisce, per la variabile marginale i , la realizzazione casuale h -esima ottenuta dal D-vine;
- $r_{i,h}^{OY}$ è la realizzazione della v.a. \tilde{R}_i^{OY} associata all'estrazione casuale del D-vine $u_{i,h}$.

Il numero di simulazioni adottato è pari a $H = 100.000$ e coincide con il numero di simulazioni utilizzate per implementare il modello CRM; il risultato di tale procedura quindi è costituito da una matrice di quattro colonne, corrispondenti alle Lob in esame, e 100.000 righe, che rappresentano la numerosità campionaria delle singole variabili marginali.

Di seguito, a fini di chiarezza espositiva, si riporta la tabella 4.39 presentata in precedenza, che mostra la matrice del τ di Kendall tra le variabili marginali.

τ	RCA	RCG	IEN	ADB
RCA	1			
RCG	0,14	1		
IEN	0,11	0,14	1	
ADB	0,08	0,12	0,12	1

Tabella 4.59: Valori del τ di Kendall tra le variabili del reserve risk delle quattro Lob ottenuti tramite simulazione di 100.000 valori dal D-vine selezionato.

Le tabelle seguenti riportano il ρ di Spearman e il coefficiente di correlazione lineare indotto dal D-vine tra le variabili marginali \tilde{R}_i^{OY} .

ρ_S	RCA	RCG	IEN	ADB
RCA	1			
RCG	0,21	1		
IEN	0,16	0,21	1	
ADB	0,12	0,18	0,17	1

Tabella 4.60: Valori del ρ di Spearman tra le variabili del reserve risk delle quattro Lob ottenuti tramite simulazione di 100.000 valori dal D-vine selezionato.

ρ	RCA	RCG	IEN	ADB
RCA	1			
RCG	0,22	1		
IEN	0,16	0,22	1	
ADB	0,14	0,16	0,16	1

Tabella 4.61: Valori del coefficiente di correlazione lineare tra le variabili del reserve risk delle quattro Lob ottenuti tramite simulazione di 100.000 valori dal D-vine selezionato.

Il requisito patrimoniale di solvibilità relativo al rischio di riservazione complessivo delle quattro linee di business è calcolato sulla variabile casuale unidimensionale data dalla somma delle quattro marginali \tilde{R}_i^{OY} con struttura di dipendenza coincidente con quella del D-vine prescelto:

$$\begin{aligned}
 SCR\left(\sum_{i=1}^4 \tilde{R}_i^{OY}\right) &= q_{99,5\%}\left(\sum_{i=1}^4 \tilde{R}_i^{OY}\right) - E\left(\sum_{i=1}^4 \tilde{R}_i^{OY}\right) \\
 &= q_{99,5\%}\left(\sum_{i=1}^4 \tilde{R}_i^{OY}\right) - \sum_{i=1}^4 E\left(\tilde{R}_i^{OY}\right)
 \end{aligned}$$

La v.a. rappresentativa del reserve risk complessivo ottenuta aggregando con la vine copula le variabili marginali è caratterizzata da un coefficiente di variazione pari al 3,49% e indice di asimmetria di 0,136. La tabella seguente riporta il valore atteso, la standard deviation, la variabilità relativa e l'asimmetria della variabile casuale rappresentativa del reserve risk complessivo.

	Valore atteso	Standard deviation	Coeff. di variazione	Asimmetria
Reserve risk	1.458.037	50.842	3,49%	0,136

Tabella 4.62: Valore atteso, standard deviation, coefficiente di variazione e asimmetria del reserve risk complessivo (gli importi monetari sono espressi in migliaia di euro).

Il valore di SCR complessivo risulta pari a 136.562 migliaia di euro, a cui corrisponde un SCR ratio, definito come rapporto tra il SCR e il suo valore atteso, del 9,37%. Il moltiplicatore (k) applicato alla standard deviation che implicitamente porta al valore di SCR totale è pari a 2,69, a fronte di un moltiplicatore di 3 previsto dalla normativa europea. La tabella seguente riporta alcune caratteristiche della variabile aleatoria del rischio di riservazione totale.

	$q_{99,5\%}$	TVaR _{99%}	SCR	SCR ratio	k
Reserve risk	1.595.278	1.615.022	136.562	9,37%	2,69

Tabella 4.63: Quantile al livello di confidenza 99,5%, Tail VaR al livello 99%, SCR, SCR ratio e moltiplicatore implicitamente definito dal valore di SCR (gli importi monetari sono espressi in migliaia di euro).

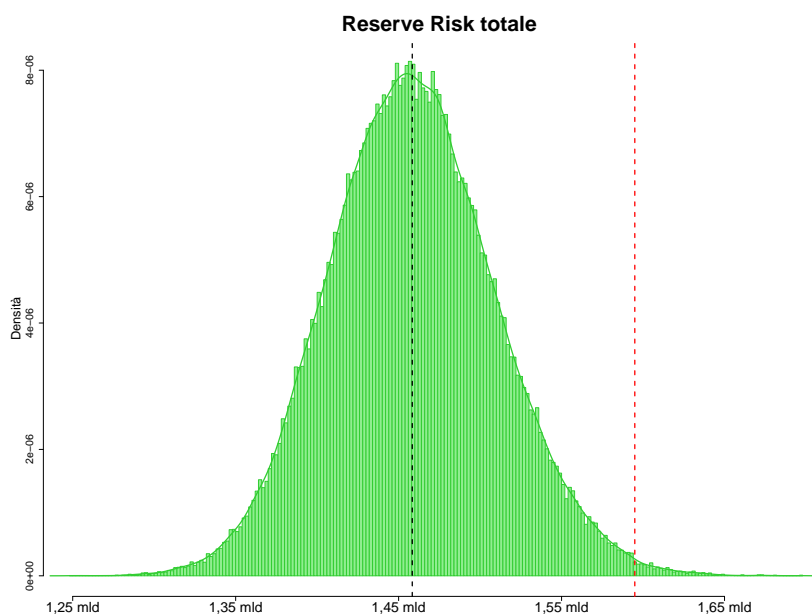


Figura 4.28: Distribuzione del reserve risk complessivo; le linee verticali di colore nero e rosso individuano rispettivamente il valore atteso e il quantile 99,5%.

Considerando il caso limite di rischi marginali tra loro perfettamente dipendenti in senso positivo, quindi tra loro comonotoni, il SCR complessivo coincide con la somma dei singoli SCR marginali. In tale circostanza l'assicuratore non gode di alcun effetto di diversificazione tra i rischi, infatti:

$$\begin{aligned} SCR\left(\sum_{i=1}^4 \tilde{R}_i^{OY}\right) &= q_{99,5\%}\left(\sum_{i=1}^4 \tilde{R}_i^{OY}\right) - E\left(\sum_{i=1}^4 \tilde{R}_i^{OY}\right) \\ &= \sum_{i=1}^4 q_{99,5\%}\left(\tilde{R}_i^{OY}\right) - \sum_{i=1}^4 E\left(\tilde{R}_i^{OY}\right) = \sum_{i=1}^4 SCR\left(\tilde{R}_i^{OY}\right) \end{aligned}$$

in quanto, come noto, nella circostanza di v.a. comonotone i quantili godono della proprietà dell'additività.

Il valore di SCR nella situazione di perfetta dipendenza positiva tra i rischi permette di calcolare il seguente indice di diversificazione, che quantifica la quota di capitale diversificato rispetto alla situazione di assenza totale di compensazione tra i rischi:

$$\text{Indice Div.} = \frac{\sum_{i=1}^4 SCR\left(\tilde{R}_i^{OY}\right) - SCR\left(\sum_{i=1}^4 \tilde{R}_i^{OY}\right)}{\sum_{i=1}^4 SCR\left(\tilde{R}_i^{OY}\right)} = 1 - \frac{SCR\left(\sum_{i=1}^4 \tilde{R}_i^{OY}\right)}{\sum_{i=1}^4 SCR\left(\tilde{R}_i^{OY}\right)}$$

Nel caso studio in esame, la somma dei singoli SCR ammonta a 213.056 migliaia di euro e l'indice di diversificazione associato al SCR complessivo determinato precedentemente è del 35,90%.

La normativa europea Solvency II aggrega i singoli SCR marginali attraverso la formula della standard deviation della somma più variabili casuali (si rimanda al Capitolo 1 per ulteriori dettagli),

$$\begin{aligned} SCR_{nl\text{ prem res}} &= 3 \cdot \sigma_{nl} \cdot V_{nl} \\ &= 3 \cdot \frac{1}{V_{nl}} \cdot \sqrt{\sum_{s,t} CorrS_{s,t} \sigma_s V_s \sigma_t V_t} \cdot V_{nl} \\ &= 3 \cdot \sqrt{\sum_{s,t} CorrS_{s,t} \sigma_s V_s \sigma_t V_t} \end{aligned}$$

dove $\sigma_s V_s$ individua lo scarto quadratico medio del premium e reserve risk della generica Lob s e $CorrS_{s,t} = \rho_{s,t}$ rappresenta il coefficiente di correlazione lineare tra le Lob s e t .

Coerentemente con tale metodologia, e considerando solo il contesto del reserve

risk, è possibile considerare gli SCR marginali derivanti dal metodo CRM e aggregare tali grandezze tramite un valore di correlazione lineare comune ($\bar{\rho}$) tra tutte le Lob:

$$SCR \left(\sum_{i=1}^4 \tilde{R}_i^{OY} \right) = \sqrt{\sum_{i=1}^4 SCR \left(\tilde{R}_i^{OY} \right)^2 + \bar{\rho} \sum_{\substack{i,j \\ (i \neq j)}} SCR \left(\tilde{R}_i^{OY} \right) SCR \left(\tilde{R}_j^{OY} \right)} \quad (4.4.1)$$

Tale analisi permette di quantificare i valori del requisito patrimoniale di solvibilità e SCR ratio complessivi, nonché l'indice di diversificazione al variare del parametro $\bar{\rho}$.

La tabella e il grafico seguenti mostrano i valori del SCR, SCR ratio e dell'indice di diversificazione ottenuti aggregando gli SCR marginali tramite la formula (4.4.1) al variare del valore medio di correlazione lineare tra le linee di business. Si noti che date $n \geq 3$ variabili aleatorie, il coefficiente di correlazione lineare medio possiede il limite inferiore (Wickens (1995)),

$$\bar{\rho}_{min} \geq -\frac{1}{(n-1)}$$

che nel caso in esame, date 4 variabili aleatorie, risulta essere $\bar{\rho}_{min} \geq -1/3$.

$\bar{\rho}$	SCR	SCR ratio	Indice Div.
1,0	213.056	14,61%	0,00%
0,9	205.737	14,11%	3,44%
0,8	198.149	13,59%	7,00%
0,7	190.258	13,05%	10,70%
0,6	182.025	12,48%	14,56%
0,5	173.401	11,89%	18,61%
0,4	164.326	11,27%	22,87%
0,3	154.720	10,61%	27,38%
0,2	144.476	9,91%	32,19%
0,1	133.448	9,15%	37,36%
0,0	121.422	8,33%	43,01%
-0,1	108.067	7,41%	49,28%
-0,2	92.809	6,37%	56,44%
-1/3	67.282	4,61 %	68,42%

Tabella 4.64: Solvency Capital Requirement, SCR ratio e indice di diversificazione al variare del coefficiente di correlazione medio ($\bar{\rho}$) utilizzato per aggregare i valori di SCR marginali.

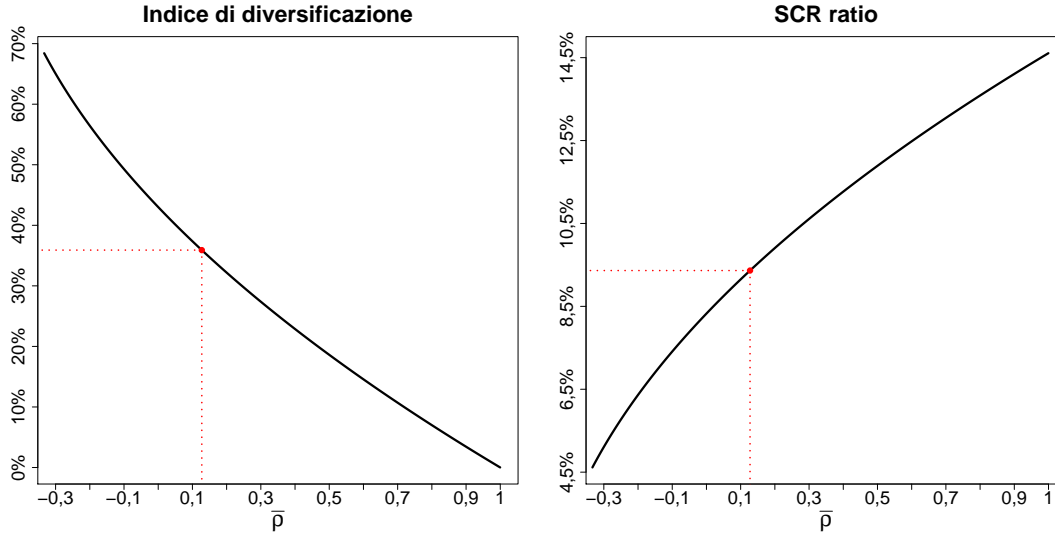


Figura 4.29: Indice di diversificazione e SCR ratio al variare del parametro $\bar{\rho}$; le linee tratteggiate di colore rosso evidenziano i valori determinati dalla struttura di aggregazione del D-vine.

Invertendo la relazione (4.4.1), inoltre, è possibile determinare il valore di correlazione lineare medio implicitamente determinato dalla struttura di dipendenza del D-vine:

$$\bar{\rho} = \frac{\left[SCR \left(\sum_{i=1}^4 \tilde{R}_i^{OY} \right) \right]^2 - \sum_{i=1}^4 SCR \left(\tilde{R}_i^{OY} \right)^2}{\sum_{\substack{i,j \\ (i \neq j)}} SCR \left(\tilde{R}_i^{OY} \right) SCR \left(\tilde{R}_j^{OY} \right)}$$

dove,

- $SCR \left(\tilde{R}_i^{OY} \right)$ individua il valore di SCR marginale del ramo i ottenuto tramite modello CRM;
- $SCR \left(\sum_{i=1}^4 \tilde{R}_i^{OY} \right)$ rappresenta il valore di SCR totale ottenuto aggregando i rischi marginali \tilde{R}_i^{OY} tramite la struttura di dipendenza del D-vine.

A fronte del valore di SCR ottenuto aggregando i rischi marginali tramite vine copula, il valore di correlazione lineare medio implicito è pari a $\bar{\rho}_{D-vine} = 0,13$.

L'applicazione della standard formula per il calcolo del requisito patrimoniale di solvibilità complessivo porta a dei risultati che si discostano in modo materiale dal SCR determinato tramite il modello CRM e la struttura di dipendenza della vine copula. Considerando i valori marginali di SCR ottenuti nel paragrafo precedente tramite standard formula (si veda tabella 4.58) e procedendo all'aggregazione di questi ultimi tramite la matrice di correlazione lineare prevista dagli

Atti Delegati si giunge ad un valore complessivo di SCR di 343.083 migliaia di euro, che corrisponde ad un valore di SCR ratio del 23,53%; la correlazione lineare media implicita risulta $\bar{\rho} = 0,40$. In particolare, la matrice di correlazione utilizzata nel processo di aggregazione è la seguente:

ρ	RCA	RCG	IEN	ADB
RCA	1			
RCG	0,50	1		
IEN	0,25	0,25	1	
ADB	0,25	0,25	0,50	1

Tabella 4.65: Matrice di correlazione prevista dagli Atti Delegati (allegato IV) per le Lob oggetto di studio. La correlazione associata al ramo ADB è stata mutuata della Lob IEN; il valore di ρ tra le linee IEN e ADB, infine, è stato ipotizzato pari a 0,50.

dove il ramo RCA coincide con la Lob 1 prevista dagli Atti Delegati, il ramo RCG corrisponde alla linea di business 5 prevista dalla normativa europea, il ramo di attività IEN è stato assimilato alla Lob 4, mentre per il ramo ADB, che nella normativa europea è incorporato nella Lob 4, sono stati considerati i valori di correlazione analoghi a quelli della Lob 4. Nella matrice di correlazione precedente è stato, infine, ipotizzato un valore di correlazione tra le linee IEN e ADB pari a 0,50. Si noti, inoltre, che l'ipotesi effettuata sul coefficiente di correlazione tra le Lob IEN e ADB ha un impatto contenuto sul valore complessivo di SCR a causa del peso limitato degli SCR di tali linee di business rispetto alle Lob RCA e RCG. Ipotizzare un valore di correlazione tra i rami IEN e ADB pari a 0 o a 1, infatti, determina rispettivamente un SCR totale di 342.337 e 343.828 migliaia di euro.

Al fine di effettuare un'analisi di sensibilità sul SCR complessivo ottenuto con il modello interno parziale, è stato calcolato il requisito patrimoniale di solvibilità aggregando gli SCR marginali ottenuti con l'approccio CRM ricorrendo alla matrice di correlazione lineare fornita dagli Atti Delegati (tabella 4.65). In tale circostanza il SCR complessivo risulta pari a 160.525 migliaia di euro, a cui corrisponde un SCR ratio dell'11,01%. La tabella seguente permette di confrontare i valori di SCR, SCR ratio e dell'indice di diversificazione ottenuti con il modello interno, la standard formula e tramite l'approccio *ibrido*, in cui gli SCR marginali sono calcolati con il modello interno mentre l'aggregazione dei rischi segue la standard formula.

	SCR	SCR ratio	Indice Div.
Modello Interno	136.562	9,37%	35,90%
SCR_i^{CRM}, P^{SF}	160.525	11,01%	24,66%
Standard formula	343.083	23,53%	18,13%

Tabella 4.66: SCR, SCR ratio e indice di diversificazione calcolati tramite modello interno, standard formula e ricorrendo agli SCR marginali del modello CRM e matrice di correlazione (P^{SF}) definita dagli Atti Delegati (gli importi monetari sono espressi in migliaia di euro).

Il confronto tra il SCR complessivo calcolato tramite modello interno e quello calcolato utilizzando gli SCR marginali definiti dall'approccio CRM e con struttura aggregativa della standard formula permette di quantificare e isolare l'effetto della sola struttura di dipendenza adottata per aggregare i rischi. In particolare, la differenza tra il SCR del modello interno e quello calcolato con l'approccio ibrido è pari a circa 24 milioni di euro a vantaggio del modello interno, corrispondente ad un incremento di SCR rispetto al modello interno di circa il 18%. Tale variazione, inoltre, corrisponde a un incremento assoluto del SCR ratio di 1,6 punti percentuali (da 9,37% a 11,01%). La standard formula, di contro, determinata un SCR totale di circa 207 milioni di euro più elevato rispetto al modello interno, pari a un incremento di capitale del 151%. Tale variazione, inoltre, equivale a un incremento assoluto di SCR ratio rispetto al modello interno del 14,2% (da 9,37% a 23,53%).

In definitiva, il SCR derivante dalla standard formula è di 207 milioni più elevato del SCR del modello interno, di cui 24 milioni, pari al 12%, imputabili alla differente struttura di dipendenza tra i rischi, mentre il restante 88% è imputabile alla differente modellizzazione dei rischi marginali. Il risparmio di capitale, quindi, che si ottiene dal modello interno rispetto alla standard formula è in larga parte imputabile al diverso valore che assumo gli SCR marginali nelle due casistiche; in particolare, l'elemento principale che porta a tale risultato è la notevole differenza tra il requisito di capitale dei rami RCA e RCG calcolato con il modello CRM e la standard formula; l'elevata dimensione delle Lob RCA e RCG, infatti, porta il modello CRM a determinare un SCR contenuto rispetto a quello che si avrebbe con la standard formula, in cui è assente ogni legame con il parametro dimensionale dell'assicuratore. Si noti, inoltre, che la differenza tra la standard formula e il CRM è tale da portare ad un valore di SCR totale calcolato in conformità con la standard formula maggiore rispetto a quello che si otterrebbe nel caso di comonotonicità tra gli SCR marginali ottenuti tramite approccio CRM.

La struttura di dipendenza tra i rischi, modellizzata tramite vine copula, ha un effetto limitato sul SCR totale rispetto ai rischi marginali, anche a causa della elevata magnitudo che il SCR del ramo RCA possiede rispetto agli altri SCR marginali. Il requisito di capitale di tale Lob, infatti, corrisponde al 44% della somma dei singoli SCR marginali, a fronte di un peso relativo di SCR dei rami RCG, IEN e ADB rispettivamente del 30%, 18% e 8%. Il peso preponderante del requisito di capitale del ramo RCA, quindi, determina un SCR totale e SCR

ratio prossimi ai valori marginali di tale Lob, facendo risultare, nel caso specifico, contenuto l'effetto della struttura di aggregazione adottata. L'indice di diversificazione in tale circostanza, come mostrato in precedenza, è pari al 35,90%, che è un valore inferiore rispetto a quello che si otterrebbe nella situazione di rischi marginali con identico peso relativo a livello di portafoglio.

Al fine di esemplificare tale risultato, è utile osservare quanto sviluppato nel seguente esempio in cui si considera la struttura di dipendenza della vine copula selezionata e per i rischi marginali delle quattro Lob si ipotizza una identica distribuzione.

In particolare ogni ramo sia definito da una v.a. $\text{Gamma}(\alpha; \beta)$ di media pari ad un quarto di quella dell'intero portafoglio (circa 365 milioni) e coefficiente di variazione (CV) del 10%. I parametri, quindi, ricorrendo al metodo dei momenti sono definiti come segue: $\alpha = 1/(CV)^2$, $\beta = \alpha/\text{media}$. In tale circostanza, ogni Lob possiede SCR pari a circa 101 milioni e il valore di SCR totale aggregato tramite D-vine risulta di 237 milioni. La standard formula, di contro, determina dei valori marginali di SCR, pari a 98 milioni per la Lob RCA, 120 milioni per RCG e 109 milioni per IEN e ADB rispettivamente; il valore di SCR totale, infine, è pari a 309 milioni di euro.

In tale contesto a fronte di valori simili tra gli SCR marginali delle distribuzioni Gamma e quelli ottenuti con la standard formula e stante l'identica distribuzione dei rischi marginali Gamma, che determina un identico valore di SCR per ogni Lob, si ha che la differenza tra i due SCR totali, pari a 73 milioni, sia imputabile per un 66% alla diversa struttura di aggregazione tra D-vine e standard formula, mentre il restante 34% afferisca alla diversa modellizzazione degli SCR marginali. In tale circostanza, quindi, l'effetto sul SCR complessivo imputabile alla struttura di aggregazione dei rischi risulta preponderante rispetto all'effetto imputabile alla diversa modellizzazione degli SCR marginali. Al crescere del peso percentuale del SCR di una singola Lob rispetto al totale degli SCR marginali, minore diventa l'effetto che la struttura di aggregazione dei rischi ha sul SCR totale; quest'ultimo, inoltre, diviene sempre più determinato dal singolo rischio marginale, riducendo sempre più il beneficio di diversificazione tra i rischi.

Il grafico seguente mostra, per l'esempio considerato, con sottostante la struttura D-vine, l'effetto sul SCR totale imputabile alla variazione del peso percentuale del SCR di una singola Lob rispetto alla somma totale degli SCR marginali. In particolare, il grafico è costruito fissando a livello di portafoglio l'ammontare della somma dei singoli SCR e variando il SCR di una Lob a scapito degli SCR delle restanti tre Lob. Nel dettaglio, il SCR della Lob 1 è stato fatto variare attribuendogli un peso percentuale dal 99% al 5% a livello complessivo rispetto alla somma dei singoli SCR. Si noti che data una v.a. $\text{Gamma}(\alpha; \beta)$, è possibile incrementare il valore di SCR di un fattore ϵ riscaldando il parametro β per la grandezza $1/\epsilon$.

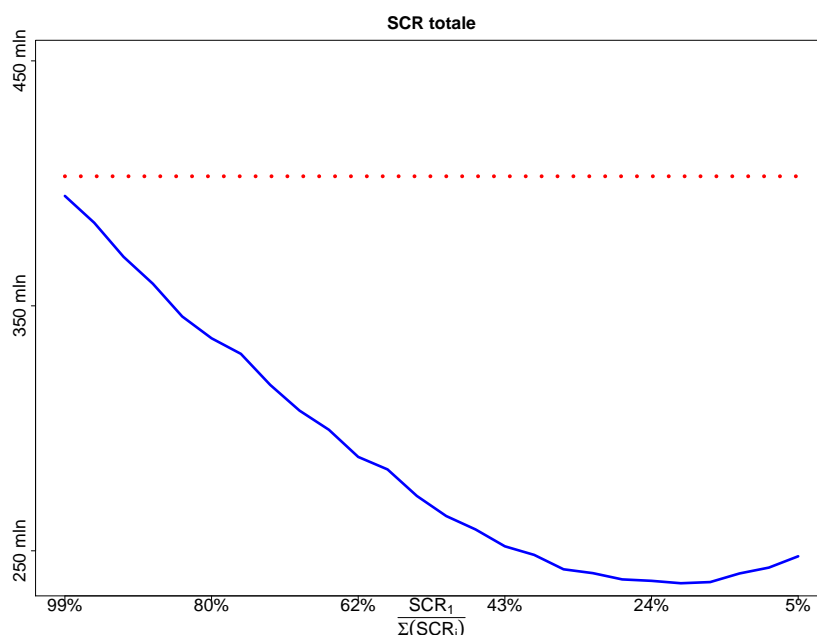


Figura 4.30: Andamento del SCR totale, ottenuto aggregando 4 linee di business tramite vine copula, al variare del peso percentuale del SCR marginale di una singola Lob e tenendo costante la somma degli SCR marginali. La linea di colore blu individua il SCR della somma della v.a., mentre la linea rossa tratteggiata rappresenta il valore della somma dei quattro SCR marginali.

Il grafico precedente mostra che al crescere della concentrazione del portafoglio verso il SCR di una sola Lob, il valore di SCR totale converge al valore di SCR di tale ramo, a cui è imputabile, inoltre, la quasi totalità del valore della somma dei singoli SCR marginali. Dati gli SCR di $i = 1, \dots, n$ Lob, quando il SCR di una Lob tende ad assumere un peso preponderante rispetto agli altri SCR marginali, $SCR_i \rightarrow 100\%$, vale che il SCR della somma delle v.a. tende al SCR marginale i -esimo, il quale inoltre tende alla somma dei singoli SCR marginali.

In tale situazione, inoltre, il beneficio di diversificazione tende a zero e la struttura di dipendenza utilizzata per aggregare i rischi risulta avere anch'essa un impatto ininfluenza sul SCR aggregato. Si verifica, di contro, una situazione opposta nel caso in cui gli SCR marginali delle singole Lob abbiano lo stesso peso percentuale; nel grafico in esame, infatti, si evidenzia il maggiore beneficio di diversificazione nel valore di $SCR_1 / \sum_i SCR = 0,25$, in quanto, per costruzione gli SCR delle altre tre Lob sono stati calibrati tali da essere identici e quindi tutti gli SCR marginali pesano 0,25. In tale circostanza, inoltre, la struttura di dipendenza sottostante ai rischi marginali risulta avere l'effetto massimo sul SCR totale.

Le considerazioni esposte in precedenza valgono, in termini generali, indipendentemente dalla struttura di aggregazione utilizzata tra i rischi. Nell'ottica della standard formula, utilizzando quindi il coefficiente di correlazione lineare, si ha, anche in tale circostanza, che al crescere della magnitudo del SCR marginale di una Lob il beneficio di diversificazione e quindi l'effetto della struttura di aggregazione tra i rischi tende a svanire. Il grafico seguente esemplifica tale fenomeno

considerando il SCR aggregato di due sole Lob al variare del peso relativo degli SCR in portafoglio; in particolare si ha che $SCR_1 + SCR_2 = 100$ e il valore di SCR_1 è fatto variare nell'intervallo $[0, 100]$, attribuendo, quindi, rispettivamente all'altro SCR marginale un valore di $SCR_2 = 100 - SCR_1$. L'analisi è condotta per i seguenti valori di correlazione, $\rho = 0$, $\rho = 0,25$, $\rho = 0,50$, $\rho = 0,75$ e $\rho = 1$.

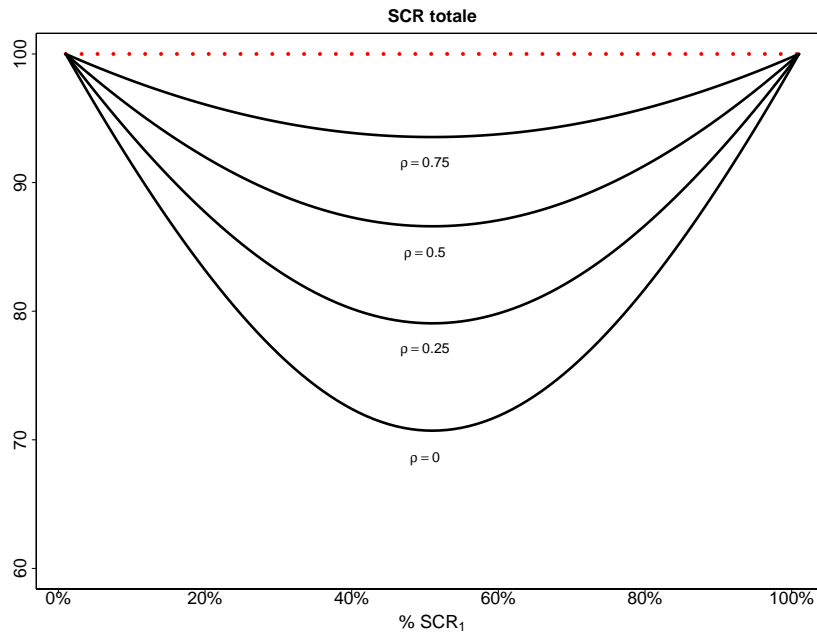


Figura 4.31: Andamento del SCR totale aggregando due SCR marginali con correlazione lineare e facendo variare il peso percentuale dei singoli SCR. Le linee nere individuano il SCR totale calcolato per i valori $\rho = 0$, $\rho = 0,25$, $\rho = 0,50$, $\rho = 0,75$; la linea tratteggiata rossa rappresenta la somma dei due SCR marginali ($\rho = 1$).

Conclusioni

Il rischio di riservazione nel contesto delle assicurazioni danni rappresenta uno dei rischi caratteristici dell'attività assicurativa a cui sono esposti gli assicuratori. Le analisi sviluppate nella presente tesi propongono un modello volto a valutare il reserve risk complessivo afferente a molteplici linee di attività; l'approccio, inoltre, nel contesto normativo europeo definito da *Solvency II* si qualifica come un possibile modello interno parziale.

Quanto formulato risponde all'esigenza pratica di costruire un modello multivariato, atto a valutare il reserve risk complessivo di un assicuratore, che tenga conto sia delle specificità delle differenti linee di business che della relativa struttura di dipendenza. L'approccio sviluppato, inoltre, supera il problema, tipico del contesto in esame, rappresentato dalla limitata numerosità dei dati; nella prassi attuariale, infatti, la valutazione della dipendenza del reserve risk tra differenti rami di attività è basata su indicatori sintetici rappresentativi del rischio in esame, quali il run-off ratio; tali grandezze, tuttavia, sono costruite, in genere, considerando intervalli temporali annuali. Il metodo sviluppato, di contro, si basa su un approccio di ricampionamento congiunto implementato sui triangoli di run-off che permette di ottenere un data set sintetico di dimensione arbitraria con cui calibrare la struttura di dipendenza.

Nell'approccio proposto la quantificazione del non-life reserve risk è perseguita analizzando separatamente sia i singoli rischi marginali, rappresentativi del rischio di riservazione delle varie Lob, che la relativa struttura di dipendenza sottostante alle variabili marginali.

La dipendenza insita tra i rischi marginali è studiata ricorrendo ad una particolare classe di copule multivariate, definite *vine copule*. Tale costrutto metodologico prevede di descrivere la dipendenza multivariata tra n ($n > 2$) fonti di rischio ricorrendo a sole copule bivariate, definite *pair-copule*; l'utilizzo di copule bidimensionali permette di cogliere molteplici tipologie di dipendenza e modellizzare in modo estremamente flessibile l'intera struttura multivariata. Le vine copule, quindi, consentono di superare la rigidità delle famiglie di copule multidimensionali proposte in letteratura, quali quelle ellittiche e archimedee. Le vine copule, sviluppate e utilizzate in letteratura principalmente nel contesto finanziario, sono adottate nel contesto in esame per descrivere la dipendenza sottostante alle 4 Lob oggetto di analisi. In particolare, la vine copula è calibrata su un data set sintetico che contiene la dipendenza implicita sottostante ai triangoli di run-off. L'uso di un metodo di ricampionamento parametrico bootstrap *congiunto* applicato ai

triangoli di run-off delle diverse linee di business consente di ottenere un data set, di numerosità arbitraria, di valori ricostruiti di riserva sinistri dipendenti tra loro. Tale approccio, dunque, estrae la dipendenza implicitamente contenuta nei dati storici dei triangoli di run-off. La dipendenza così ottenuta è funzione non solo delle osservazioni storiche, intese come realizzazioni dell'ignoto processo generatore dei dati, ma anche del metodo attuariale deterministico di stima della riserva sottostante alla procedura di ricampionamento bootstrap; questa peculiarità rappresenta una potenziale limitazione del metodo stesso. Tale specificità, tuttavia, permette di applicare l'approccio di ricampionamento congiunto a qualsiasi metodo deterministico di stima della riserva sinistri basato sui triangoli di run-off. Si noti, inoltre, che anche l'utilizzo della serie storica del run-off ratio per quantificare la dipendenza tra più Lob risente del metodo di stima utilizzato per valutare la riserva sinistri.

Stante il numero di rischi marginali pari a 4, al fine di selezionare il modello migliore è stato possibile stimare tutte le possibili configurazioni di vine copule esistenti con 4 variabili aleatorie; la struttura che meglio si adatta al data set in esame risulta essere una vine copula appartenente alla classe dei D-vine. La copula selezionata, in seguito, è stata confrontata con le famiglie di copule multivariate ellittiche e archimedee, palesando la superiorità della vine copula prescelta.

Le analisi sviluppate per stimare la vine copula, inoltre, si sono soffermate sull'ipotesi semplificata alla base della *pair-copula construction*, che consiste nel considerare copule bivariate non condizionate. Il caso in esame ha mostrato la difficoltà insita nel valutare la validità di tale assunzione, che passa necessariamente attraverso la stima di copule multidimensionali ($n > 2$); tale circostanza, al crescere della dimensione del problema, fa venire meno uno dei fondamentali vantaggi delle vine copule, rappresentato dall'utilizzo di sole copule bidimensionali. La superiorità in termini di *fitting* delle vine copule rispetto alle classiche strutture di dipendenza multivariate rende, tuttavia, l'ipotesi semplificata alla base della pair-copula construction un'assunzione non stringente del modello.

I rischi marginali rappresentativi del reserve risk dei singoli segmenti di attività sono modellizzati tramite un approccio Collective Risk Model (CRM). La peculiarità di tale modello, colmando una mancanza della letteratura, consiste nel quantificare la variabilità di parametro del modello, introdotta in tale contesto dai fattori di disturbo, tramite una metodologia bayesiana che permette di calibrare i fattori sistematici sulla base dell'intera informazione storica contenuta nei triangoli di run-off. L'approccio bayesiano, inoltre, è adottato per quantificare il grado di dipendenza lineare tra i parametri caratteristici del modello, dati dalla *frequency e severity* dei sinistri. Il metodo CRM permette di ottenere, tramite simulazioni Monte Carlo, la distribuzione di probabilità dei rischi marginali del reserve risk considerando l'intera durata di vita delle passività; di contro, in coerenza con quanto previsto da Solvency II, l'approccio re-reserving permette di valutare il rischio di riservazione considerando un orizzonte temporale annuo.

I parametri del modello CRM sono stimati ricorrendo al metodo deterministico Frequency-Severity basato sulla meccanica Chain-Ladder applicata ai triangoli di

run-off storici del numero e costo dei sinistri; tale metodo, inoltre, è alla base dell'approccio di ricampionamento congiunto utilizzato per estrarre la dipendenza insita tra i triangoli di run-off delle differenti Lob. Nel complesso, quindi, il modello interno parziale è basato su un unico metodo deterministico di stima della riserva sinistri, garantendo una coerenza dell'intero approccio proposto.

In conclusione, il modello sviluppato nella tesi, atto a valutare il rischio di riservazione complessivo afferente ad una compagnia di assicurazioni, risulta essere un valido approccio in quanto permette di considerare le caratteristiche dei singoli rischi marginali, nonché le peculiarità della struttura di dipendenza sottostante alle linee di attività considerate. L'intero modello, inoltre, è costruito interamente utilizzando i triangoli di run-off del numero e del costo dei sinistri. I rischi marginali, modellizzati con una metodologia basata sulla Teoria del Rischio Collettiva, considerano l'incertezza di parametro, valutata tramite approccio bayesiano, basandosi sull'intera informazione storica disponibile, evitando qualsiasi sorta di expert judgment. La dipendenza sottostante ai dati, infine, è quantificata ricorrendo ad un approccio di ricampionamento congiunto dei triangoli di run-off ed è modellizzata tramite le vine copule, permettendo di costruire una copula multivariata che descrive la struttura di dipendenza in modo assai più flessibile e adeguato rispetto alle famiglie di copule multivariate esistenti in letteratura.

Bibliografia

- Aas, K. (2016). «Pair-Copula Construction for Financial Applications: A Review». In: *Econometrics* 4.43. DOI: [10.3390/econometrics4040043](https://doi.org/10.3390/econometrics4040043).
- Aas, K. et al. (2009). «Pair-Copula Construction of Multiple dependence». In: *Insurance: Mathematics and Economics* 44.2, pp. 182–198.
- Acar, E.F., C. Genest e J. Nešlehová (2012). «Beyond simplified pair-copula construction». In: *Journal of Multivariate Analysis* 110, pp. 74–90.
- Allen, D.E., M. Ashraf et al. (2013). «Financial Dependence Analysis: Applications of Vine Copulas». In: *Statistica Neerlandica* 67.4. DOI: [10.1111/stan.12015](https://doi.org/10.1111/stan.12015).
- Allen, D.E., M. McAleer e A.K. Singh (2017). «Risk Measurement and Risk Modelling Using Applications of Vine Copulas». In: *Sustainability* 9.1762. DOI: [10.3390/su9101762](https://doi.org/10.3390/su9101762).
- Balakrishnan, N. e Chin-Diew Lai (2009). *Continuous Bivariate Distributions*. Springer.
- Bedford, T. e R.M. Cooke (2001). «Probability Density Decomposition for Conditionally Dependent Random Variables Modeled by Vines». In: *Annals of Mathematics and Artificial Intelligence* 32, pp. 245–268.
- (2002). «Vines—a new graphical model for dependent random variables». In: *Ann. Statist.* 30.4, pp. 1031–1068. DOI: [10.1214/aos/1031689016](https://doi.org/10.1214/aos/1031689016). URL: <https://doi.org/10.1214/aos/1031689016>.
- Berg, D. e K. Aas (2009). «Models for construction of multivariate dependence: A comparison study». In: *The European Journal of Finance*.
- Brechmann, E. e U. Schepsmeier (2013). «Modeling Dependence with C- and D-Vine Copulas: The R Package CDVine». In: *Journal of Statistical Software, Articles* 52.3, pp. 1–27. ISSN: 1548-7660. DOI: [10.18637/jss.v052.i03](https://doi.org/10.18637/jss.v052.i03). URL: <https://www.jstatsoft.org/v052/i03>.
- CEIOPS (2006). *Quantitative Impact Study 2, Technical Specification*. PI-08/06.
- Clarke, K.A. (2007). «A Simple Distribution-Free test for Nonnested Model Selection». In: *Political analysis Advance Acces*.
- Commissione Europea (2015). *Regolamento delegato (UE) 2015/35*. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2015:012:TOC>.
- (2019). *Regolamento delegato (UE) 2019/981*. <https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32019R0981&from=EN>.

- Cooke, R.M. (1997). «Markov and Entropy Properties of Tree- and Vines- Dependent Variables». In: *Proceedings of the ASA Section on Bayesian Statistical Science*.
- Cuadras, C. M. (2002). «On the Covariance between Functions». In: *Journal of Multivariate Analysis* 81, pp. 19–27. DOI: [10.1006/jmva.2001.2000](https://doi.org/10.1006/jmva.2001.2000).
- Czado, C., S. Jeske e M. Hofmann (2013). «Selection strategies for regular vine copulae». In: *Journal de la Société Française de Statistique* 154, pp. 11–129.
- De Felice, M. e F. Moriconi (2016). «On the Estimation of the Undertaking-Specific Parameters and the related Hypothesis Testing». In: *University of Perugia Working paper*.25.
- Demarta, S. e A.J. McNeil (2005). «The t Copula and Related Copulas». In: *International Statistical Review* 73.1, pp. 111–129. DOI: [10.1111/j.1751-5823.2005.tb00254.x](https://doi.org/10.1111/j.1751-5823.2005.tb00254.x). eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1751-5823.2005.tb00254.x>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1751-5823.2005.tb00254.x>.
- Diers, D. (2009). «Stochastic re-reserving in multi-year internal models - An approach based on simulations». In: *ASTIN Colloquium*.
- Dißmann, J. et al. (2013). «Selecting and estimating regular vine copulae and application to financial returns». In: *Computational Statistics & Data Analysis* 59, pp. 52–69.
- Efron, B. e R.J. Tibshirani (1993). *An introduction to the bootstrap*. USA: Chapman & Hall.
- Eling, M. e K. Jung (2018). «Risk aggregation in non-life insurance: standard models vs. internal models». In: *University of St.Gallen - Working papers on risk management and insurance* 215.
- Embrechts, P., F. Lindskog e E. J. McNeil (2003). «Modelling Dependence With Copulas and Applications to Risk Management». In: *Handbook of Heavy Tailed Distributions in Finance*. Elsevier Science B.V. Cap. 8, pp. 329–384.
- England, P. e R. Verrall (1999). «Analytic and bootstrap estimates of prediction errors in claims reserving». In: *Insurance: Mathematics and Economics* 25, pp. 281–293.
- (2002). «Stochastic Claims reserving in general insurance». In: *British Actuarial Journal* 8, pp. 443–544.
- England, P., R. Verrall e M.V. Wüthrich (2018). «Bayesian Overdispersed Poisson Model and the Bornhuetter-Ferguson Claims Reserving Method». In: *Annals of Actuarial Science* 6.2, pp. 258–283.
- Fang, H. B., K.T. Fang e S. Kotz (2002). «The Meta-elliptical Distributions with Given Marginals». In: *Journal of Multivariate Analysis* 82, pp. 1–16. DOI: [10.1006/jmva.2001.2017](https://doi.org/10.1006/jmva.2001.2017).
- Fang, K.T., S. Kotz e K.W. Ng (1989). *Symmetric Multivariate and Related Distributions*. Vol. 36. Monographs on Statistics & Applied Probability. Chapman e Hall/CRC.
- Fisher, N. e P. Switzer (1985). «Chi-plots for assessing dependence». In: *Biometrika* 72.2, pp. 253–265.

- Friedland, J. (2010). *Estimating unpaid claims using basic techniques*. Casualty Actuarial Society.
- Gamerman, D. e H.F. Lopes (2006). *Markov Chain Monte Carlo - stochastic Simulation for Bayesian Inference*. Texts in Statistical Science. Chapman & Hall/CRC.
- Geidosh, M. e M Fischer (2016). «Applications of Vine Copulas to Credit portfolio risk Modeling». In: *Journal of Risk and Financial Management* 9.4. DOI: [10.3390/jrfm9020004](https://doi.org/10.3390/jrfm9020004).
- Genest, C. e A.C. Favre (2007). «Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask». In: *Journal of Hydrologic Engineering*.
- Genest, C., J.F. Quessy e B. Rémillard (2006). «Goodness-of-fit procedures for copula models based on the probability integral transformation». In: *Scandinavian Journal of Statistics* 33.2, pp. 337–366.
- Green, P.J. (1995). «Reversible jump Markov chain Monte Carlo computation and Bayesian model determination». In: *Biometrika* 82, pp. 711–732.
- Gruber, L.F. e C. Czado (2015). «Sequential Bayesian Model Selection of regular Vine Copulas». In: *Bayesian Analysis* 10.4, pp. 937–963.
- (2018). «Bayesian Model Selection of Regular Vine Copulas». In: *Bayesian Analysis* 13.4, pp. 1111–1135.
- Haff, I.H. (2012). «Comparison of estimators for pair-copula constructions». In: *Journal of Multivariate Analysis* 110, pp. 91–105.
- (2013). «Parameter estimation for pair-copula constructions». In: *Bernoulli* 19.2, pp. 462–491.
- Haff, I.H., K. Aas e A. Frigessi (2010). «On the simplified pair-copula construction - Simply useful or too simplistic?». In: *Journal of Multivariate Analysis* 101, pp. 1296–1310.
- International Actuarial Association (2004). «A Global Framework for Insurer Solvency Assessment». In: *Research Report of the Insurer Solvency Assessment Working Party*.
- Joe, H. (1996). «Families of m -Variate Distributions with Given Margins and $m(m-1)/2$ Bivariate Dependence Parameters». In: *Distributions with Fixed Marginals and related Topics*, pp. 120–141.
- (1997). *Multivariate Models and Multivariate Dependence Concepts*. Chapman & Hall.
- Killiches, M., D. Kraus e C. Czado (2016). «Examination and visualisation of the simplifying assumption for vine copulas in three dimensions». In: *arXiv:1602.05795v2*.
- (2018). «Using model distances to investigate the simplifying assumption, model selection and truncation levels for vine copulas». In: *arXiv:1610.08795v3*.
- Kollo, T. e A. Roos (2003). «On Kotz-Type Elliptical Distributions». In: vol. 2. Series in Biostatistics. Contemporary Multivariate Analysis and Design of Experiments. Yale University School of Medicine, USA: World Scientific. Cap. 8, pp. 159–170.

- Kotz, S., N. Balakrishnan e N. L. Johnson (2000). *Continuous Multivariate Distributions, Volume 1: Models and Applications*. 2^a ed. John Wiley & Sons.
- Kotz, S., T. J. Kozubowski e K. Podgórski (2001a). «An Asymmetric Multivariate Laplace Distribution». In: DOI: [10.1007/978-1-4612-0173-1_6](https://doi.org/10.1007/978-1-4612-0173-1_6).
- (2001b). *The Laplace Distribution and Generalizations*. Birkhäuser.
- Kraus, D. e C. Czado (2016). «D-vine copula based quantile regression». In: *arXiv:1510.04161v4*.
- (2017). «Growing simplified vine copula trees: improving Dißmann’s algorithm». In: *arXiv:1703.05203v1*.
- Kurowicka, D. e R. Cooke (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling*. Chichester: Wiley Series in Probability e Statistics.
- Kurowicka, D. e H. Joe (2011). *Dependence Modeling*. Singapore: World Scientific.
- Kurz, M.S. e F. Spanhel (2017). «Testing the simplifying assumption in high-dimensional vine copulas». In: *arXiv:1706.02338v1*.
- Mack, T. (1993). «Distribution-Free Calculation of the Standard Error of Chain Ladder Reserve Estimates». In: *ASTIN Bulletin* 23, pp. 213–225.
- (1994). «Measuring the variability of Chain-Ladder reserve estimates». In: *Casualty Actuarial Society Forum* 23, pp. 213–225.
- (1999). «The standard error of Chain-Ladder reserve estimates: recursive calculation and inclusion of a tail factor». In: *Astin Bulletin* 29.2, pp. 361–366.
- McNeil, J.A., R. Frey e P. Embrechts (2005). *Quantitative Risk Management: Concepts, Techniques And Tools*. Princeton University Press.
- Mejdoub, H. e M. Ben Arab (2017). «A multivariate analysis for risk capital estimation in insurance industry: vine copulas». In: *Asian Economic and Social Society* 5.2, pp. 100–119. ISSN: 2313-8343. DOI: [10.18488/journal.107.2017.52.100.119](https://doi.org/10.18488/journal.107.2017.52.100.119).
- Mesfioui, M. e J.-F. Quessy (2008). «Dependence Structure of conditional Archimedean copulas». In: *Journal of Multivariate Analysis* 99, pp. 372–385.
- Meyers, G. (2009). «Stochastic loss reserving with the Collective Risk Model». In: *Casualty Actuarial Society* 3.2, pp. 239–269.
- Min, A. e C. Czado (2010). «Bayesian Inference for Multivariate Copulas Using Pair-Copula Constructions». In: *Journal of Financial Econometrics* 8.4, pp. 511–546.
- (2011). «Bayesian model selection for D-vine pair-copula constructions». In: *Canadian Journal of Statistics* 39.2, pp. 239–258.
- Moala, F.A., P.L. Ramos e J.A. Achcar (2013). «Bayesian Inference for Two-Parameter Gamma Distribution Assuming Different Noninformative Priors». In: *Revista Colombiana de Estadística* 36.2, pp. 319–336.
- Morales-Nápoles, O., R.M. Cooke e D. Kurowicka (2008). «The number of vines and regular vines on n nodes». In: *Discrete Applied Mathematics*.
- Nelsen, R.B. (2006). *An Introduction to Copulas*. Springer.
- Nikoloulopoulos, A.K., H. Joe e H. Li (2012). «Vine copulas with asymmetric tail dependence and applications to financial return data». In: *Computational*

- Statistics & Data Analysis* 56.11, pp. 3659–3673. ISSN: 0167-9473. DOI: [10.1016/j.csda.2010.07.016](https://doi.org/10.1016/j.csda.2010.07.016).
- Ohlsson, E. e J. Lauzeningks (2009). «The one-year non-life insurance risk». In: *Insurance: Mathematics and Economics* 45.2, pp. 203–208.
- Palaro, H.P. e L.K. Hotta (2006). «Using Conditional Copula to Estimate Value at Risk». In: *Journal of Data Science* 4, pp. 93–115.
- Parlameto Europeo (2009). *Direttiva 2009/138/EC "Solvency II"*. <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32009L0138>.
- Ricotta, A. e G.P. Clemente (2016). «An Extension of Collective Risk Model for Stochastic Claim Reserving». In: *Journal of Applied Finance & Banking* 6.5, pp. 45–62.
- Ricotta, A. e E. Luini (2019). «Bayesian Estimation of Structure Variables in the Collective Risk Model for Reserve Risk». In: *Journal of Applied Finance & Banking* 9.2, pp. 23–54.
- Robert, P.C. e G. Casella (2010). *Introducing Monte Carlo Methods with R*. Use R!: Springer.
- Sanchez, J.R., R. Verrall e J.L. Vilar (2011). «Bayesian and Credibility Estimation for the Chain Ladder Reserving Method». In: *Anales del Instituto de Actuarios Españoles* 17, pp. 51–74.
- Savelli, N. e G.P. Clemente (2009). «A Collective Risk Model for Claims Reserve Distribution». In: *Atti del XVI Convegno di Teoria del Rischio*, pp. 59–87.
- (2011). «Stochastic claim reserving based on CRM for Solvency II purposes». In: *ASTIN Colloquium*.
- Schirmacher, D. e E. Schirmacher (2008). «Multivariate dependence modelling using pair-copulas». In: *Technical report - Society of Actuaries: 2008 Enterprise Risk Management Symposium*.
- Schmidt, R. (2002). «Tail Dependence for Elliptically Contoured Distributions». In: *Math. Methods of Operations Research* 55.2, pp. 301–327.
- Scollnik, D.P.M. (2004). «Bayesian Reserving Models Inspired by Chain Ladder Methods and Implemented Using WinBUGS». In: *Actuarial Research Clearing House* 2.
- Solaro, N. (2004). «Random variate generation from Multivariate Exponential Power distribution». In: *Statistica & Applicazioni* II.2.
- Spanhel, F. e M.S. Kurz (2019). «Simplified vine copula models: Approximations based on the simplifying assumption». In: *Electronic Journal of Statistics* 13, pp. 1254–1291. ISSN: 1935-7524.
- Stöber, J., H. Joe e C. Czado (2013). «Simplified pair copula construction - Limitations and extensions». In: *Journal of Multivariate Analysis* 119, pp. 101–118.
- Tsukahara, H. (2005). «Semiparametric estimation in copula models». In: *The Canadian Journal of Statistics* 33.2, pp. 357–375.
- Vatter, T. e V. Chavez-Demoulin (2015). «Generalized additive models for conditional dependence structures». In: *Journal of Multivariate Analysis* 141, pp. 147–167.

- Vatter, T. e T. Nagler (2017). «Generalized Additive Models for Pair-Copula Constructions». In: *arXiv:1608.01593v2*.
- Vuong, Q.H. (1989). «Likelihood ratio tests for model selection and non-nested hypotheses». In: *Econometrica* 57.2, pp. 307–333.
- Wang, W. e M.T. Wells (2000). «Model selection and semiparametric inference for bivariate failure-time data». In: *Journal of the American Statistical Association* 95.449, pp. 62–72.
- White, H. (1982). «Maximum likelihood estimation of misspecified models». In: *Econometrica* 50.1.
- Wickens, T.D. (1995). *The Geometry of Multivariate Statistics*. Texts in Statistical Science. New Jersey, USA: Lawrence Erlbaum Associates, Publishers Hillsdale.
- Wüthrich, M. V. e M. Merz (2008). «Modelling Claims Development Result for Solvency Purposes». In: *Casualty Actuarial Society, E-Forum*.
- Wüthrich, M.V. (2007). «Using a Bayesian approach for claims reserving». In: *Variance* 1, pp. 292–301.
- Yang, R. e J.O. Berger (1997). «A catalog of noninformative priors». In: *Technical Report - Duke University Department of Statistical Science* 42.

Appendice A

Triangoli di run-off

	Lob: RCA											
	1	2	3	4	5	6	7	8	9	10	11	12
1995	90.067	172.910	212.210	231.542	243.584	251.983	259.580	266.259	271.388	276.171	279.620	282.789
1996	94.268	180.233	217.706	235.818	247.703	256.563	263.890	269.408	273.492	278.852	281.159	
1997	106.206	200.180	241.360	261.836	274.421	284.111	290.876	297.804	304.956	307.860		
1998	117.416	230.887	277.824	299.085	313.447	322.892	331.298	336.131	341.425			
1999	128.599	251.054	302.709	327.952	342.172	353.943	362.530	368.449				
2000	135.867	266.767	326.973	350.790	366.929	378.100	389.607					
2001	141.404	273.200	331.061	363.144	382.570	397.688						
2002	143.228	281.353	349.738	381.303	402.011							
2003	147.762	290.903	365.319	404.966								
2004	160.395	323.418	400.155									
2005	169.184	335.258										
2006	173.136											

Tabella A.1: Lob RCA - Importo cumulato dei sinistri pagati (valori in migliaia di euro).

Lob: RCA												
	1	2	3	4	5	6	7	8	9	10	11	12
1995	117.700	163.823	169.086	171.040	171.961	172.462	172.783	172.987	173.141	173.297	173.402	173.482
1996	115.158	155.439	160.417	162.150	162.957	163.422	163.694	163.879	164.053	164.184	164.268	
1997	119.342	158.690	163.682	165.296	166.026	166.408	166.659	166.875	167.063	167.155		
1998	124.560	167.209	172.332	173.889	174.518	174.866	175.116	175.270	175.388			
1999	127.263	170.716	176.835	178.626	179.428	179.895	180.208	180.417				
2000	124.719	164.882	170.894	172.851	173.821	174.360	174.777					
2001	121.134	160.485	167.580	170.168	171.409	172.186						
2002	118.105	160.096	168.100	171.306	172.846							
2003	117.546	156.956	165.878	169.386								
2004	113.574	151.751	160.088									
2005	112.949	150.930										
2006	104.473											

Tabella A.2: Lob RCA - Numero cumulato dei sinistri pagati.

Lob: RCG												
	1	2	3	4	5	6	7	8	9	10	11	12
1995	8.732	22.416	32.621	38.556	44.101	49.515	54.233	57.859	61.915	65.339	68.707	70.964
1996	9.954	26.410	37.650	44.513	50.579	56.416	61.851	66.751	71.864	76.689	79.972	
1997	10.834	28.021	38.585	46.402	52.053	57.505	62.561	68.277	73.710	77.759		
1998	11.924	31.306	44.457	52.073	57.549	62.979	68.597	75.117	79.398			
1999	12.043	31.174	43.411	50.012	56.120	62.107	68.574	73.872				
2000	13.917	33.503	46.641	54.820	61.982	69.533	77.134					
2001	14.213	34.872	46.881	54.422	63.641	72.595						
2002	13.495	33.832	47.643	58.510	69.256							
2003	12.235	28.714	40.231	49.056								
2004	11.282	26.339	35.739									
2005	10.070	23.065										
2006	9.139											

Tabella A.3: Lob RCG - Importo cumulato dei sinistri pagati (valori in migliaia di euro).

Lob: RCG												
	1	2	3	4	5	6	7	8	9	10	11	12
1995	9.509	15.152	19.361	19.841	20.200	20.454	20.594	20.665	20.767	20.829	20.877	20.917
1996	10.134	15.846	19.747	20.272	20.621	20.862	21.004	21.110	21.292	21.358	21.415	
1997	10.145	15.475	18.751	19.244	19.522	19.692	19.824	19.939	20.058	20.129		
1998	10.541	15.937	19.598	20.044	20.247	20.405	20.523	20.657	20.737			
1999	9.931	14.960	18.177	18.552	18.844	19.031	19.197	19.317				
2000	10.646	15.098	18.084	18.621	18.986	19.273	19.467					
2001	10.147	14.716	17.372	17.859	18.358	18.721						
2002	9.273	13.853	16.963	17.846	18.523							
2003	8.111	11.472	13.821	14.446								
2004	6.657	9.269	10.900									
2005	5.602	7.804										
2006	4.596											

Tabella A.4: Lob RCG - Numero cumulato dei sinistri pagati.

Lob: IEN												
	1	2	3	4	5	6	7	8	9	10	11	12
1995	51.960	100.685	112.147	118.394	119.570	120.258	120.844	121.419	121.637	122.010	122.358	122.517
1996	51.134	95.515	105.322	108.106	109.113	110.123	110.958	112.205	112.495	113.320	113.456	
1997	58.466	107.760	117.294	119.278	119.918	120.828	121.361	121.805	122.224	123.019		
1998	60.291	110.994	120.171	122.503	124.509	125.306	126.399	127.032	127.426			
1999	64.462	120.157	137.189	139.820	141.470	143.239	145.309	146.295				
2000	81.262	148.723	162.472	165.443	167.060	168.087	169.061					
2001	88.382	153.724	167.596	173.531	176.069	177.289						
2002	73.959	124.355	133.713	140.681	142.656							
2003	63.433	111.633	123.802	128.658								
2004	83.803	161.052	177.799									
2005	65.467	110.504										
2006	69.693											

Tabella A.5: Lob IEN - Importo cumulato dei sinistri pagati (valori in migliaia di euro).

Lob: IEN												
	1	2	3	4	5	6	7	8	9	10	11	12
1995	33.123	40.478	40.965	41.165	41.205	41.221	41.231	41.239	41.241	41.248	41.253	41.255
1996	30.471	36.109	36.522	36.606	36.637	36.658	36.662	36.682	36.687	36.699	36.702	
1997	32.047	37.643	38.009	38.059	38.076	38.090	38.098	38.105	38.109	38.123		
1998	31.200	36.367	36.684	36.738	36.777	36.789	36.802	36.812	36.815			
1999	31.118	36.476	37.116	37.175	37.217	37.245	37.275	37.292				
2000	36.388	42.000	42.435	42.512	42.555	42.575	42.589					
2001	36.933	42.222	42.761	42.913	42.985	43.010						
2002	29.749	33.903	34.250	34.345	34.411							
2003	24.615	28.213	28.675	28.811								
2004	28.946	33.851	34.428									
2005	21.320	24.112										
2006	20.514											

Tabella A.6: Lob IEN - Numero cumulato dei sinistri pagati.

Lob: ADB												
	1	2	3	4	5	6	7	8	9	10	11	12
1995	21.995	35.119	37.082	37.510	37.752	37.960	38.140	38.455	38.542	38.649	38.912	39.025
1996	25.143	38.832	41.120	41.640	42.005	42.148	42.262	42.428	42.571	42.880	42.934	
1997	25.796	41.447	43.308	44.023	44.218	44.608	44.824	45.032	45.216	45.849		
1998	28.305	44.870	46.421	46.923	48.071	48.307	48.597	48.768	48.898			
1999	33.280	50.635	52.392	52.889	53.237	53.539	53.698	53.785				
2000	37.505	57.527	59.671	60.411	61.271	61.605	61.895					
2001	37.872	56.788	58.993	59.938	60.780	61.234						
2002	37.017	54.418	59.408	60.509	60.970							
2003	33.394	49.642	51.550	52.455								
2004	34.472	51.245	52.829									
2005	34.006	49.278										
2006	34.197											

Tabella A.7: Lob ADB - Importo cumulato dei sinistri pagati (valori in migliaia di euro).

Lob: ADB												
	1	2	3	4	5	6	7	8	9	10	11	12
1995	16.908	19.097	19.191	19.206	19.216	19.223	19.227	19.234	19.236	19.238	19.244	19.246
1996	18.068	19.989	20.097	20.115	20.128	20.132	20.135	20.139	20.143	20.149	20.150	
1997	17.051	19.014	19.094	19.114	19.120	19.129	19.134	19.139	19.143	19.156		
1998	17.663	19.528	19.588	19.601	19.612	19.617	19.622	19.626	19.628			
1999	19.373	21.218	21.292	21.305	21.315	21.322	21.325	21.327				
2000	20.252	22.092	22.168	22.190	22.217	22.226	22.232					
2001	19.084	20.776	20.872	20.899	20.915	20.928						
2002	17.955	19.540	19.628	19.668	19.686							
2003	15.627	16.967	17.049	17.078								
2004	14.358	15.535	15.596									
2005	13.355	14.401										
2006	12.138											

Tabella A.8: Lob ADB - Numero cumulato dei sinistri pagati.

Appendice B

MCMC: distribuzioni a posteriori

Linea di business: Responsabilità civile autoveicoli terrestri (RCA).

Standard deviation dei fattori di disturbo

Di seguito si riporta la densità bivariata delle distribuzioni a posteriori ottenute tramite algoritmo Random Walk Metropolis-Hastings.

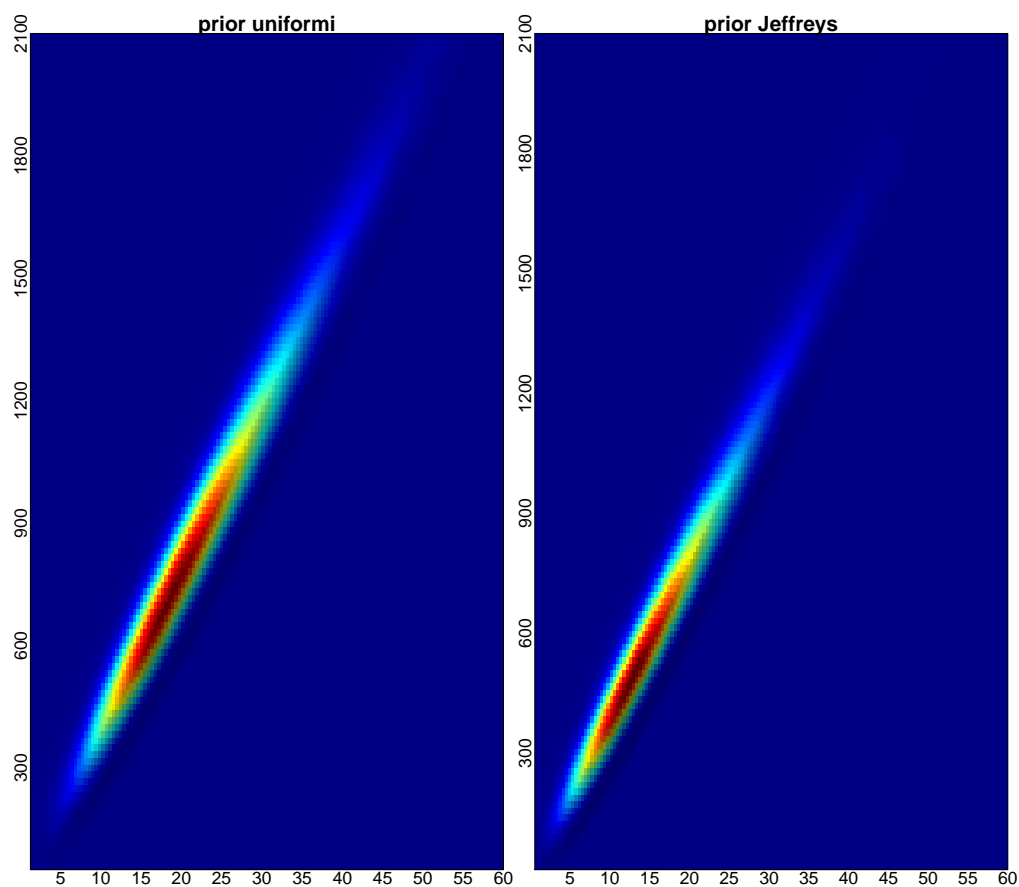


Figura B.1: Distribuzioni a posteriori (prior uniformi e di Jeffreys), associate alla v.a. $\tilde{\sigma}_{\tilde{q}}$; l'asse delle ascisse e delle ordinate si riferiscono rispettivamente alle v.a. \tilde{A} e \tilde{B} .

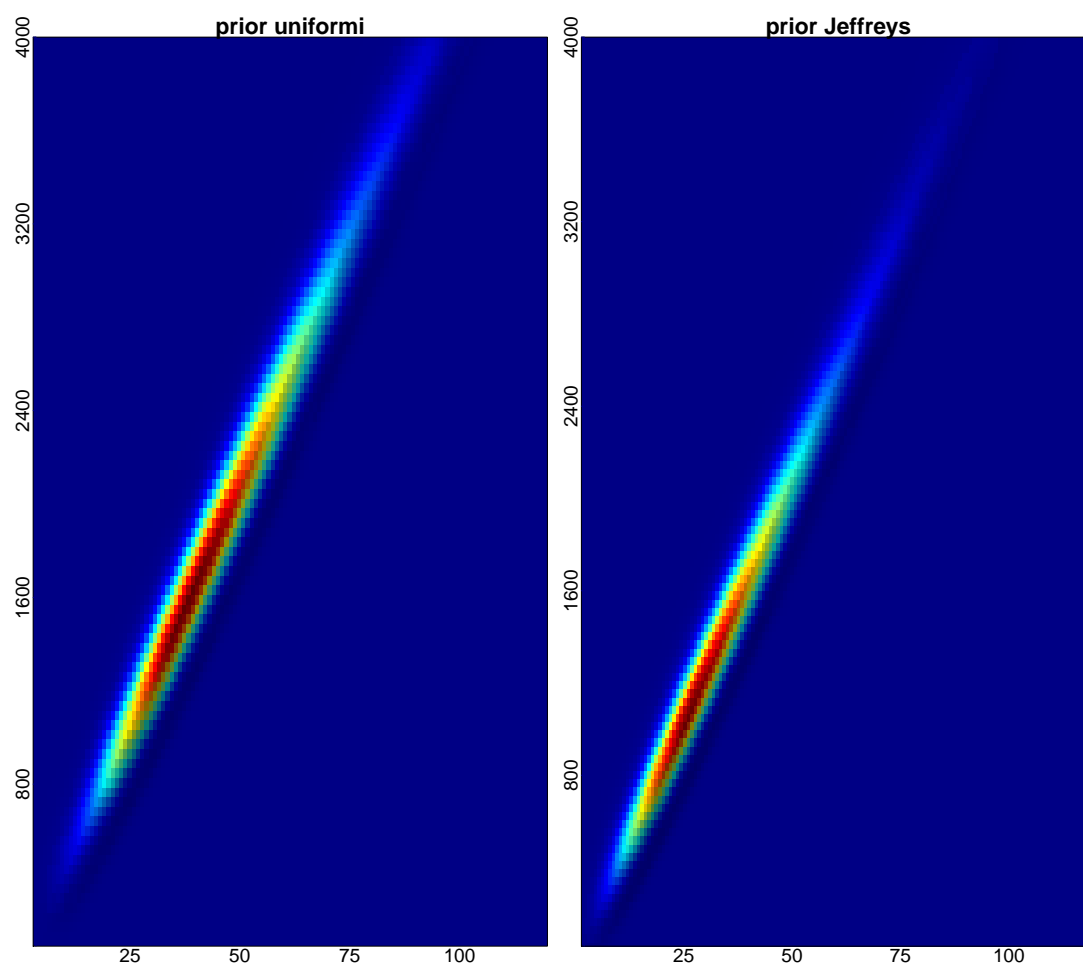


Figura B.2: Distribuzioni a posteriori (prior uniformi e di Jeffreys), associate alla v.a. $\tilde{\sigma}_p$; l'asse delle ascisse e delle ordinate si riferiscono rispettivamente alle v.a. \tilde{A} e \tilde{B} .

La tabella riporta i valori attesi delle distribuzioni a posteriori considerando le due tipologie di distribuzioni a priori adottate.

Lob: RCA

Prior	Uniforme	Jeffreys
$E(\tilde{A}_q \tilde{\sigma}_q)$	23,492	17,653
$E(\tilde{B}_q \tilde{\sigma}_q)$	922,235	689,669
$E(\tilde{A}_p \tilde{\sigma}_p)$	48,929	36,764
$E(\tilde{B}_p \tilde{\sigma}_p)$	2.043,989	1.532,360

Tabella B.1: Lob RCA - Valori attesi delle distribuzioni a posteriori.

Correlazione tra i fattori di disturbo

I grafici seguenti mostrano la densità bivariata delle distribuzioni a posteriori ottenute tramite algoritmo Random Walk Metropolis-Hastings, considerando sia le distribuzioni a priori uniformi che le distribuzioni di Jeffreys.

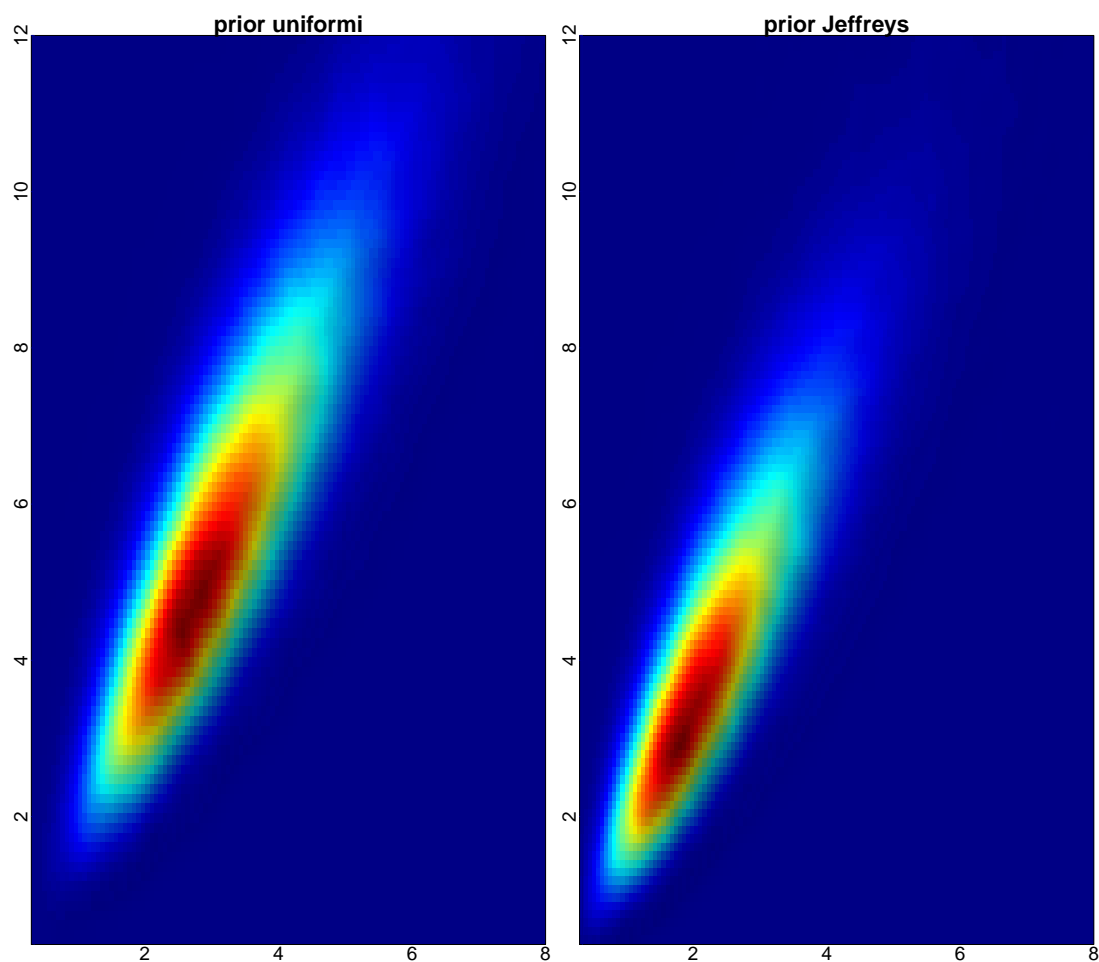


Figura B.3: Distribuzioni a posteriori (prior uniformi e di Jeffreys), associate alla v.a. $\tilde{\rho}$; l'asse delle ascisse si riferisce alla v.a. \tilde{C} , mentre l'asse delle ordinate alla variabile \tilde{D} .

La tabella riporta i valori attesi delle distribuzioni a posteriori considerando le due tipologie di distribuzioni a priori adottate.

Lob: RCA

Prior	Uniforme	Jeffreys
$E(\tilde{C} \tilde{\rho})$	3,482	2,658
$E(\tilde{D} \tilde{\rho})$	6,226	4,692

Tabella B.2: Lob RCA - Valori attesi delle distribuzioni a posteriori.

Linea di business: Responsabilità civile generale (RCG) Standard deviation dei fattori di disturbo

I grafici seguenti mostrano la densità bivariata delle distribuzioni a posteriori ottenute tramite algoritmo Random Walk Metropolis-Hastings, considerando sia le distribuzioni a priori uniformi che le distribuzioni di Jeffreys.

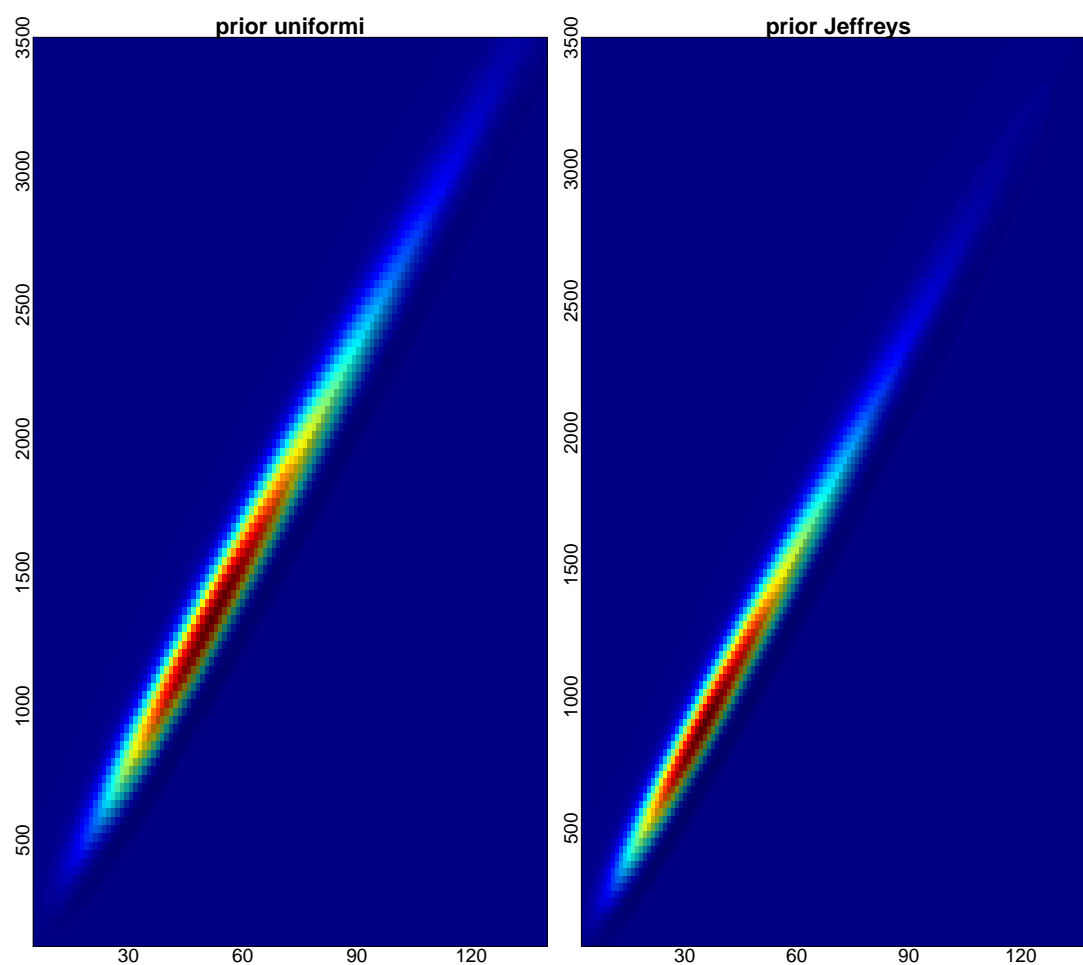


Figura B.4: Distribuzioni a posteriori (prior uniformi e di Jeffreys), associate alla v.a. $\tilde{\sigma}_{\tilde{q}}$; l'asse delle ascisse si riferisce alla v.a. \tilde{A} , mentre l'asse delle ordinate alla variabile \tilde{B} .

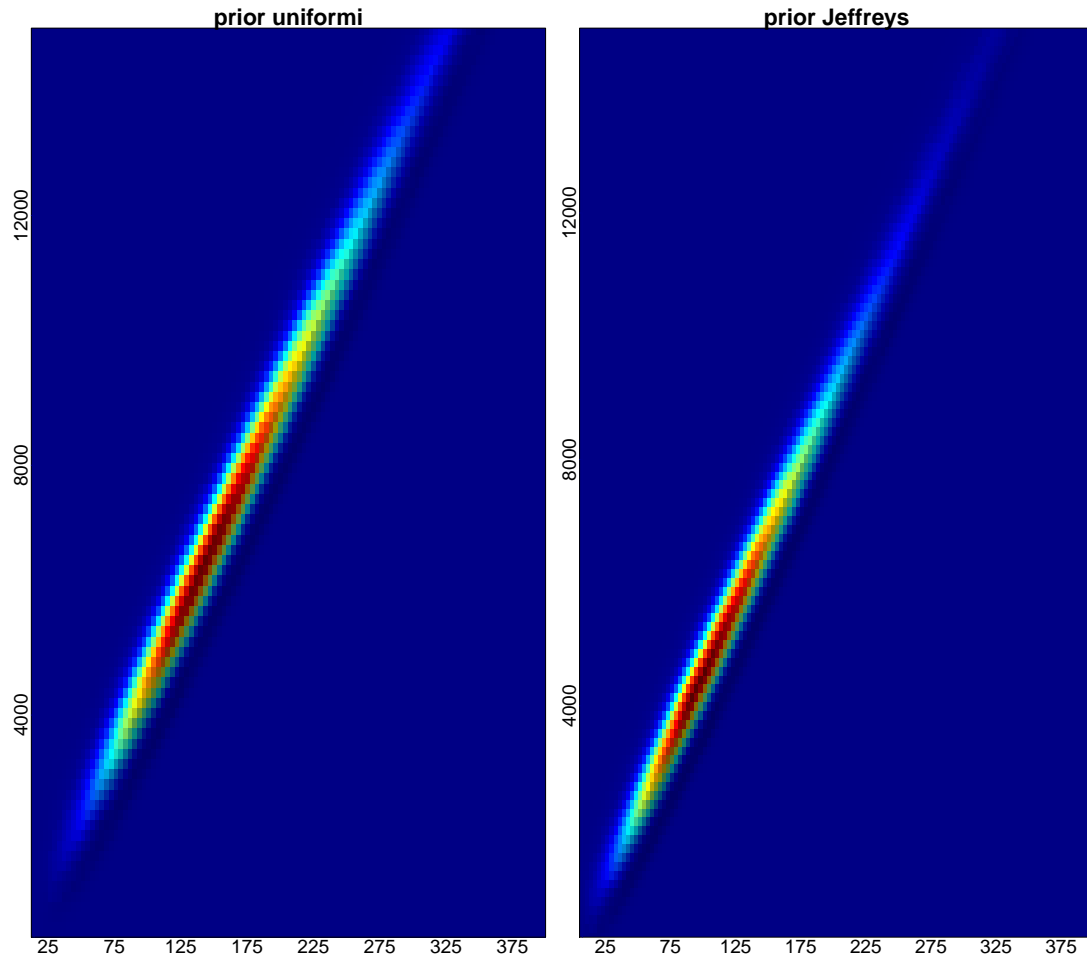


Figura B.5: Distribuzioni a posteriori (prior uniformi e di Jeffreys), associate alla v.a. $\tilde{\sigma}_{\tilde{p}}$; l'asse delle ascisse si riferisce alla v.a. \tilde{A} , mentre l'asse delle ordinate alla variabile \tilde{B} .

La tabella riporta i valori attesi delle distribuzioni a posteriori considerando le due tipologie di distribuzioni a priori adottate.

Lob: RCG

Prior	Uniforme	Jeffreys
$E(\tilde{A}_{\tilde{q}} \tilde{\sigma}_{\tilde{q}})$	63,564	47,769
$E(\tilde{B}_{\tilde{q}} \tilde{\sigma}_{\tilde{q}})$	1.647,150	1.235,768
$E(\tilde{A}_{\tilde{p}} \tilde{\sigma}_{\tilde{p}})$	177,226	132,998
$E(\tilde{B}_{\tilde{p}} \tilde{\sigma}_{\tilde{p}})$	8.042,860	6.032,038

Tabella B.3: Lob RCG - Valori attesi delle distribuzioni a posteriori.

Correlazione tra i fattori di disturbo

I grafici seguenti mostrano la densità bivariata delle distribuzioni a posteriori ottenute tramite algoritmo Random Walk Metropolis-Hastings, considerando sia le distribuzioni a priori uniformi che le distribuzioni di Jeffreys.

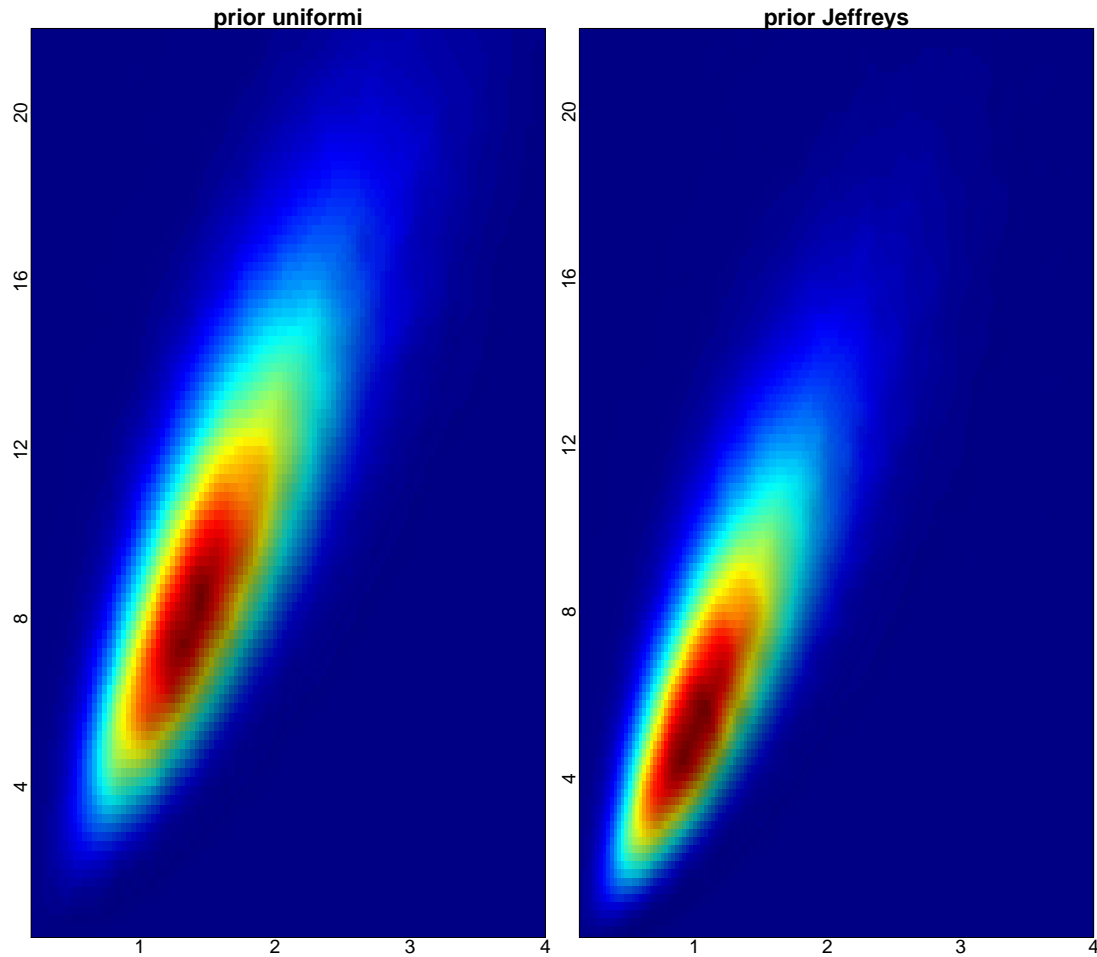


Figura B.6: Distribuzioni a posteriori (prior uniformi e di Jeffreys), associate alla v.a. $\tilde{\rho}$; l'asse delle ascisse si riferisce alla v.a. \tilde{C} , mentre l'asse delle ordinate alla variabile \tilde{D} .

La tabella riporta i valori attesi delle distribuzioni a posteriori considerando le due tipologie di distribuzioni a priori adottate.

Lob: RCG

Prior	Uniforme	Jeffreys
$E(\tilde{C} \tilde{\rho})$	1,731	1,346
$E(\tilde{D} \tilde{\rho})$	10,832	7,997

Tabella B.4: Lob RCG - Valori attesi delle distribuzioni a posteriori.

Linea di business: Incendio ed eventi naturali (IEN) Standard deviation dei fattori di disturbo

I grafici seguenti mostrano la densità bivariata delle distribuzioni a posteriori ottenute tramite algoritmo Random Walk Metropolis-Hastings, considerando sia le distribuzioni a priori uniformi che le distribuzioni di Jeffreys.

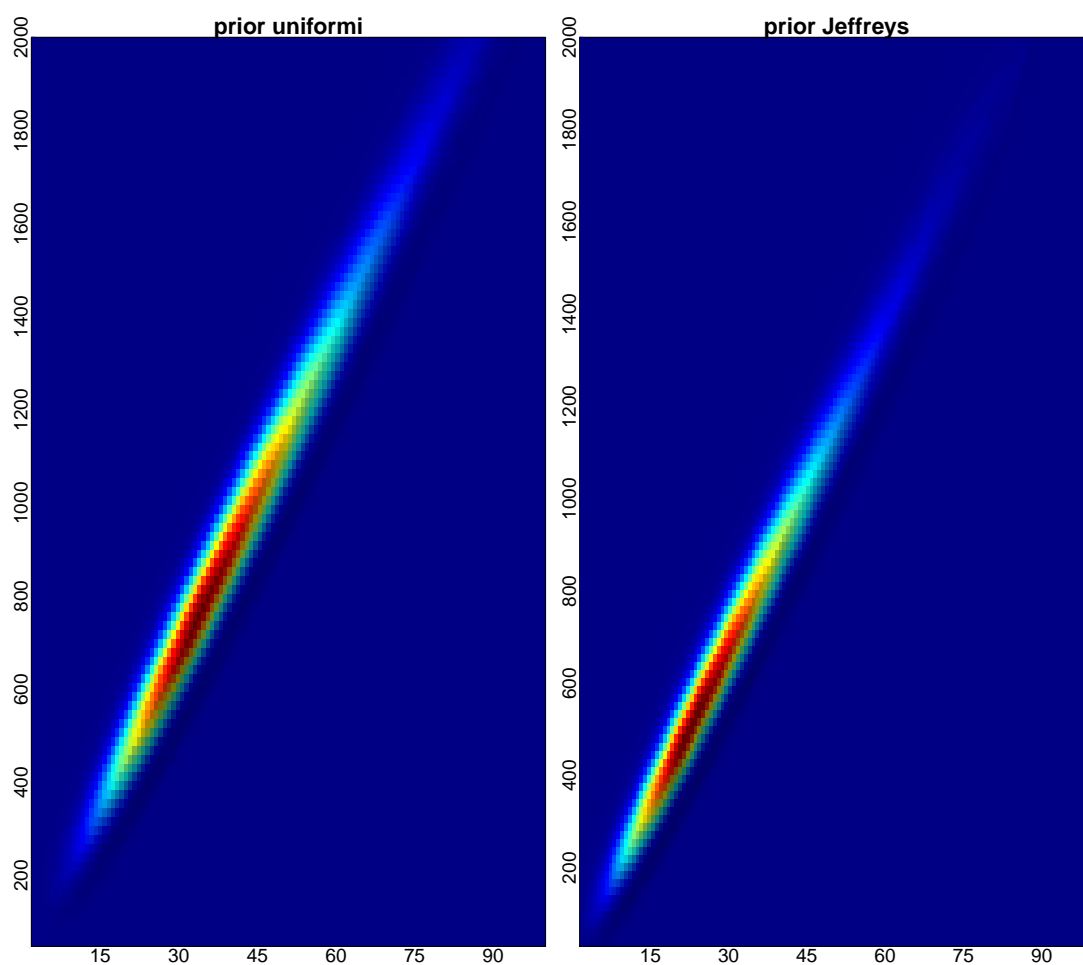


Figura B.7: Distribuzioni a posteriori (prior uniformi e di Jeffreys), associate alla v.a. $\tilde{\sigma}_{\tilde{q}}$; l'asse delle ascisse si riferisce alla v.a. \tilde{A} , mentre l'asse delle ordinate alla variabile \tilde{B} .

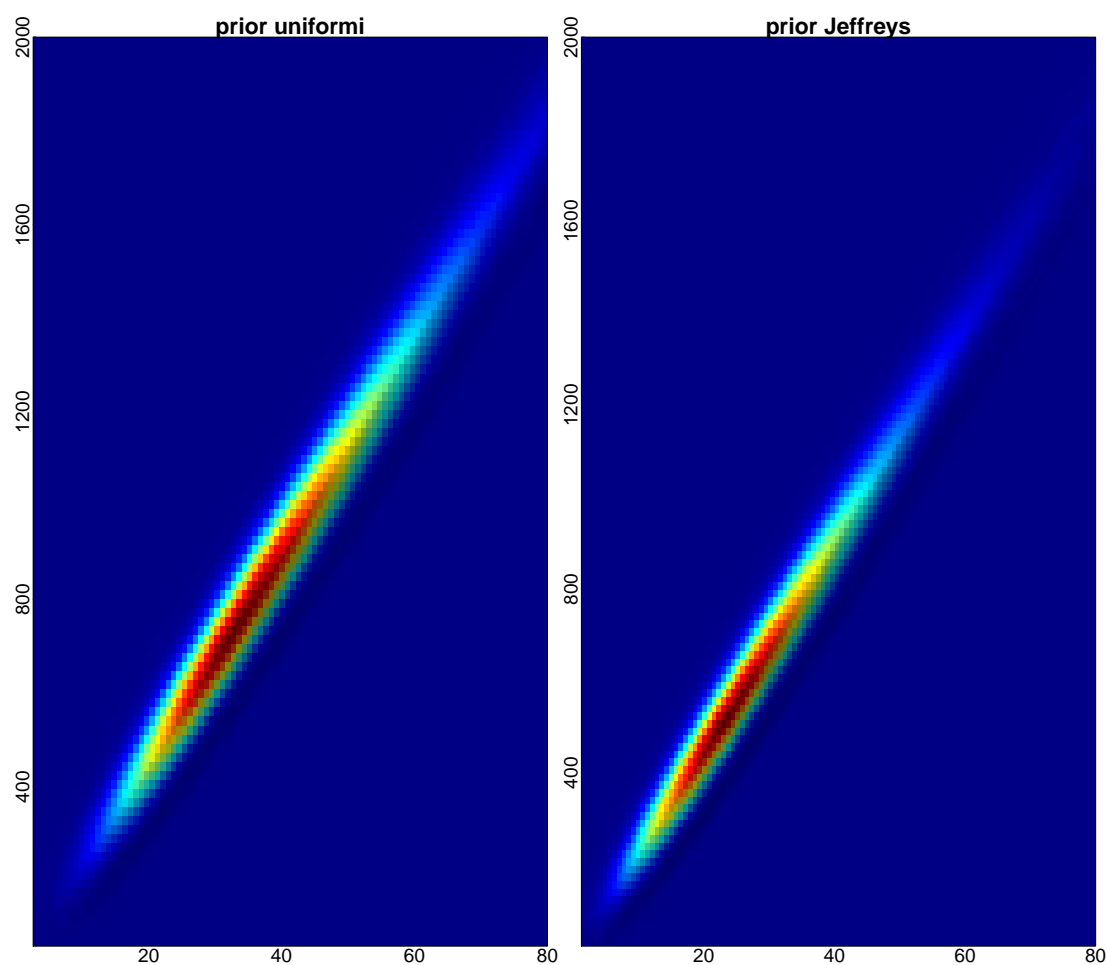


Figura B.8: Distribuzioni a posteriori (prior uniformi e di Jeffreys), associate alla v.a. $\tilde{\sigma}_{\tilde{p}}$; l'asse delle ascisse si riferisce alla v.a. \tilde{A} , mentre l'asse delle ordinate alla variabile \tilde{B} .

La tabella riporta i valori attesi delle distribuzioni a posteriori considerando le due tipologie di distribuzioni a priori adottate.

Lob: IEN

Prior	Uniforme	Jeffreys
$E(\tilde{A}_{\tilde{q}} \tilde{\sigma}_{\tilde{q}})$	43,186	32,390
$E(\tilde{B}_{\tilde{q}} \tilde{\sigma}_{\tilde{q}})$	981,926	734,532
$E(\tilde{A}_{\tilde{p}} \tilde{\sigma}_{\tilde{p}})$	42,059	31,620
$E(\tilde{B}_{\tilde{p}} \tilde{\sigma}_{\tilde{p}})$	949,486	711,998

Tabella B.5: Lob IEN - Valori attesi delle distribuzioni a posteriori.

Correlazione tra i fattori di disturbo

I grafici seguenti mostrano la densità bivariata delle distribuzioni a posteriori ottenute tramite algoritmo Random Walk Metropolis-Hastings, considerando sia le distribuzioni a priori uniformi che le distribuzioni di Jeffreys.

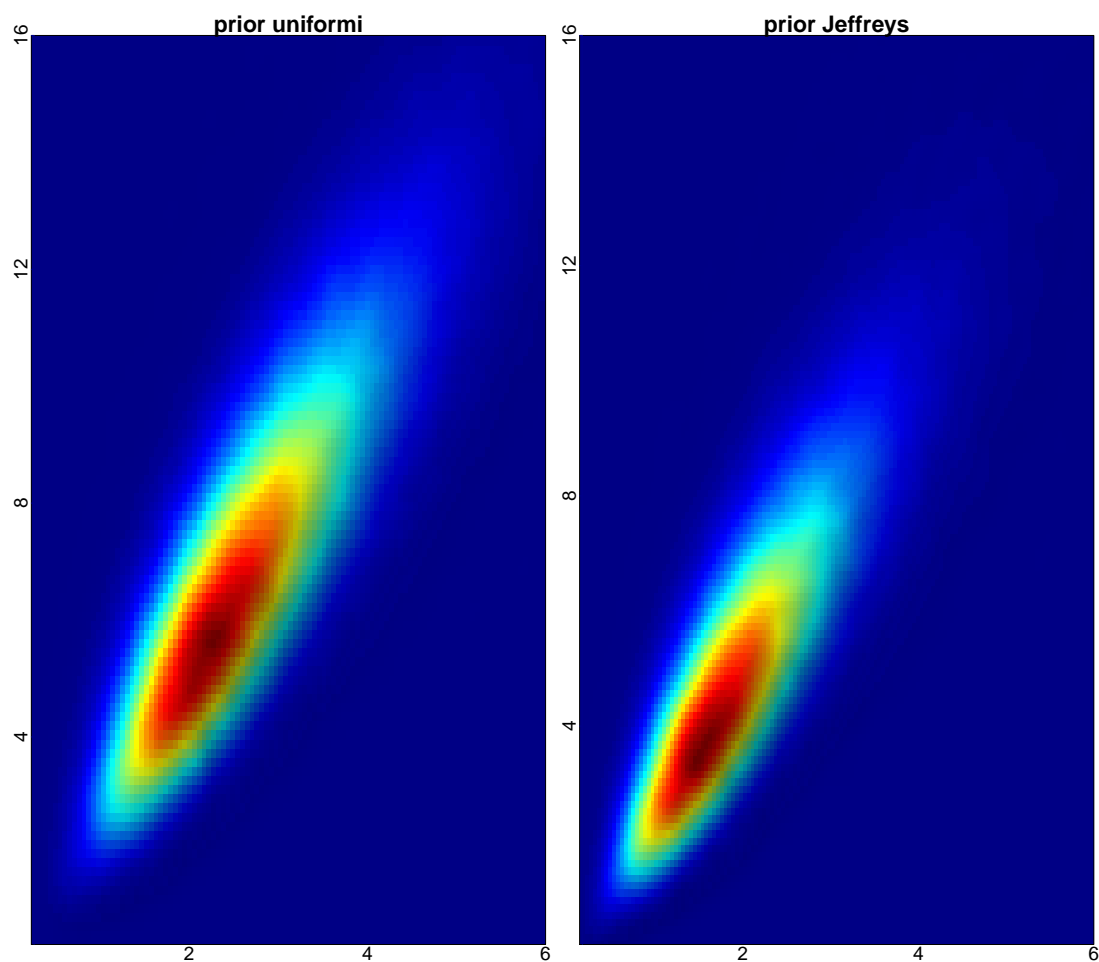


Figura B.9: Distribuzioni a posteriori (prior uniformi e di Jeffreys), associate alla v.a. $\tilde{\rho}$; l'asse delle ascisse si riferisce alla v.a. \tilde{C} , mentre l'asse delle ordinate alla variabile \tilde{D} .

La tabella riporta i valori attesi delle distribuzioni a posteriori considerando le due tipologie di distribuzioni a priori adottate.

Lob: IEN

Prior	Uniforme	Jeffreys
$E(\tilde{C} \tilde{\rho})$	2,856	2,187
$E(\tilde{D} \tilde{\rho})$	7,466	5,590

Tabella B.6: Lob IEN - Valori attesi delle distribuzioni a posteriori.

Linea di business: Altri danni ai beni (ADB) Standard deviation dei fattori di disturbo

I grafici seguenti mostrano la densità bivariata delle distribuzioni a posteriori ottenute tramite algoritmo Random Walk Metropolis-Hastings, considerando sia le distribuzioni a priori uniformi che le distribuzioni di Jeffreys.

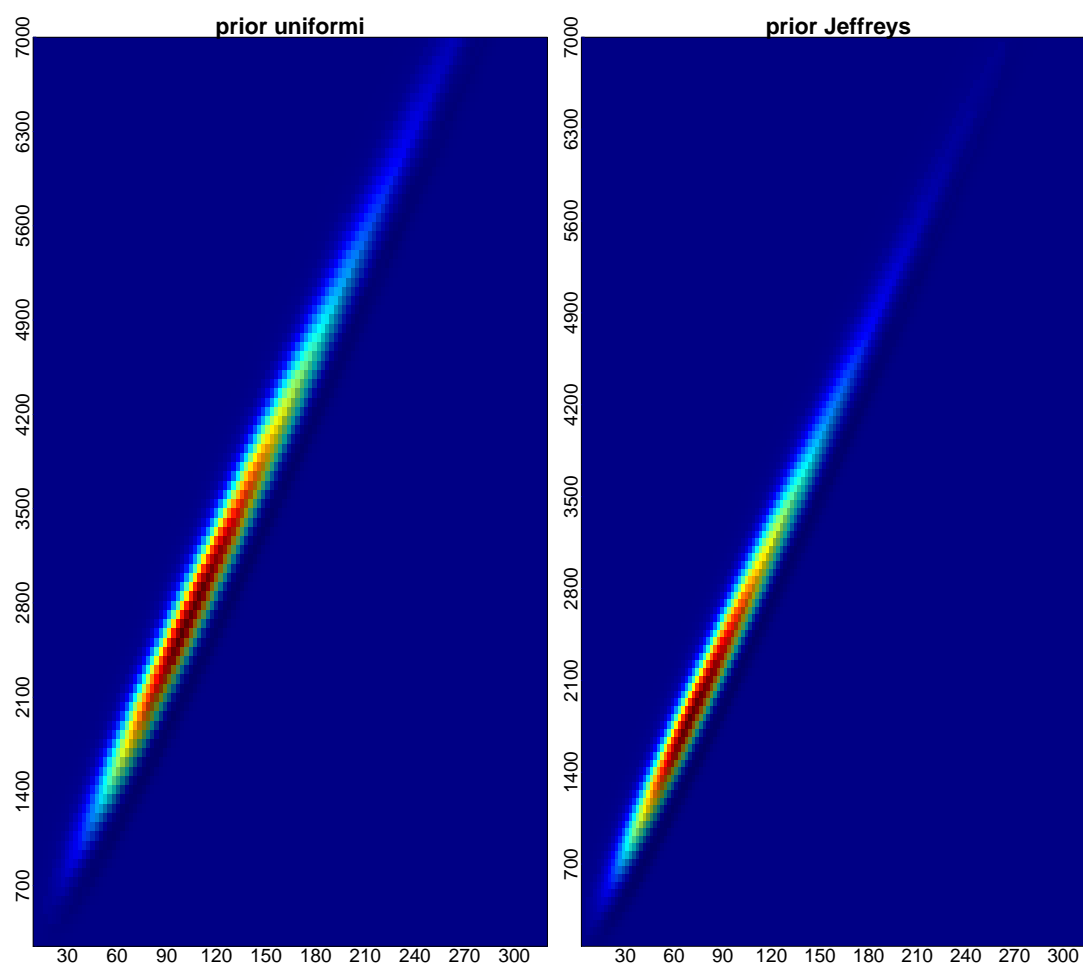


Figura B.10: Distribuzioni a posteriori (prior uniformi e di Jeffreys), associate alla v.a. $\tilde{\sigma}_{\tilde{q}}$; l'asse delle ascisse si riferisce alla v.a. \tilde{A} , mentre l'asse delle ordinate alla variabile \tilde{B} .

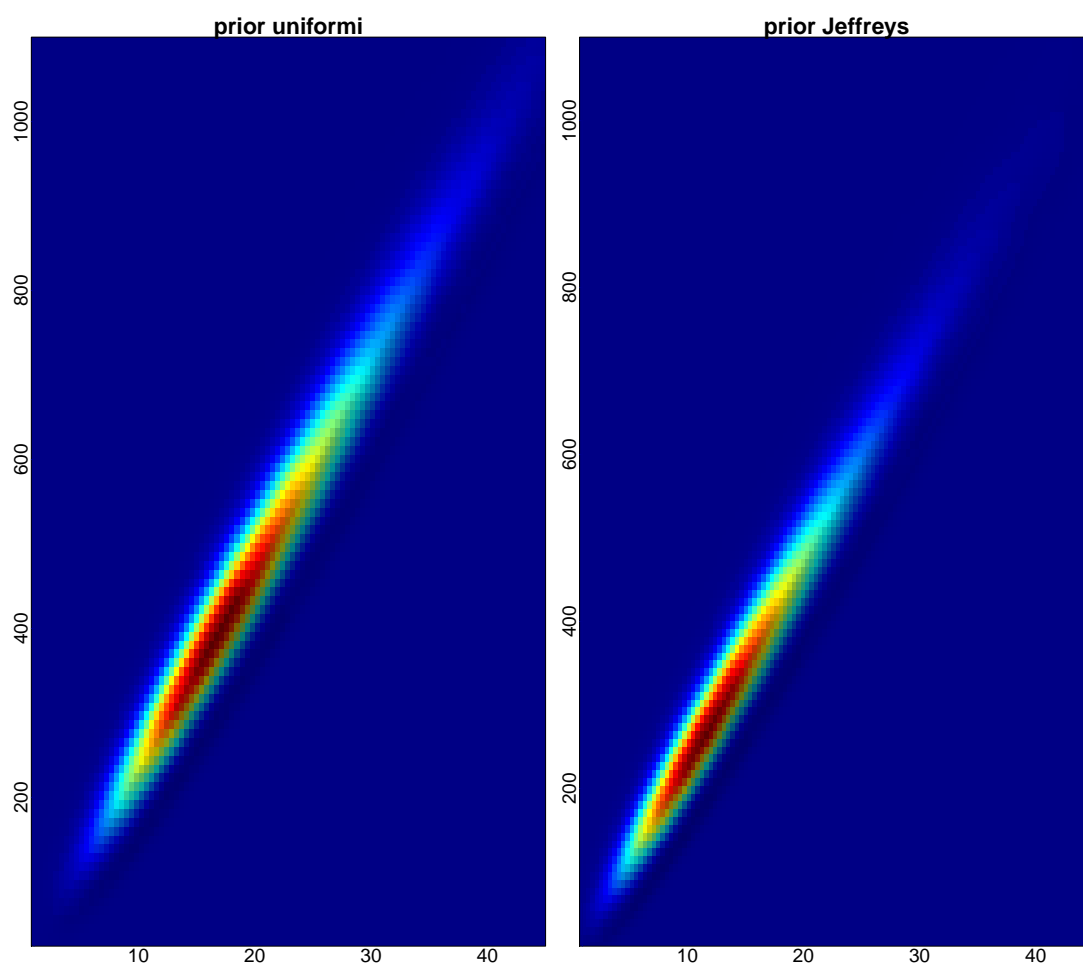


Figura B.11: Distribuzioni a posteriori (prior uniformi e di Jeffreys), associate alla v.a. $\tilde{\sigma}_{\tilde{p}}$; l'asse delle ascisse si riferisce alla v.a. \tilde{A} , mentre l'asse delle ordinate alla variabile \tilde{B} .

La tabella riporta i valori attesi delle distribuzioni a posteriori considerando le due tipologie di distribuzioni a priori adottate.

Lob: ADB

Prior	Uniforme	Jeffreys
$E(\tilde{A}_{\tilde{q}} \tilde{\sigma}_{\tilde{q}})$	129,495	97,064
$E(\tilde{B}_{\tilde{q}} \tilde{\sigma}_{\tilde{q}})$	3.392,259	2.540,532
$E(\tilde{A}_{\tilde{p}} \tilde{\sigma}_{\tilde{p}})$	21,165	15,949
$E(\tilde{B}_{\tilde{p}} \tilde{\sigma}_{\tilde{p}})$	502,461	376,716

Tabella B.7: Lob ADB - Valori attesi delle distribuzioni a posteriori.

Correlazione tra i fattori di disturbo

I grafici seguenti mostrano la densità bivariata delle distribuzioni a posteriori ottenute tramite algoritmo Random Walk Metropolis-Hastings, considerando sia le distribuzioni a priori uniformi che le distribuzioni di Jeffreys.

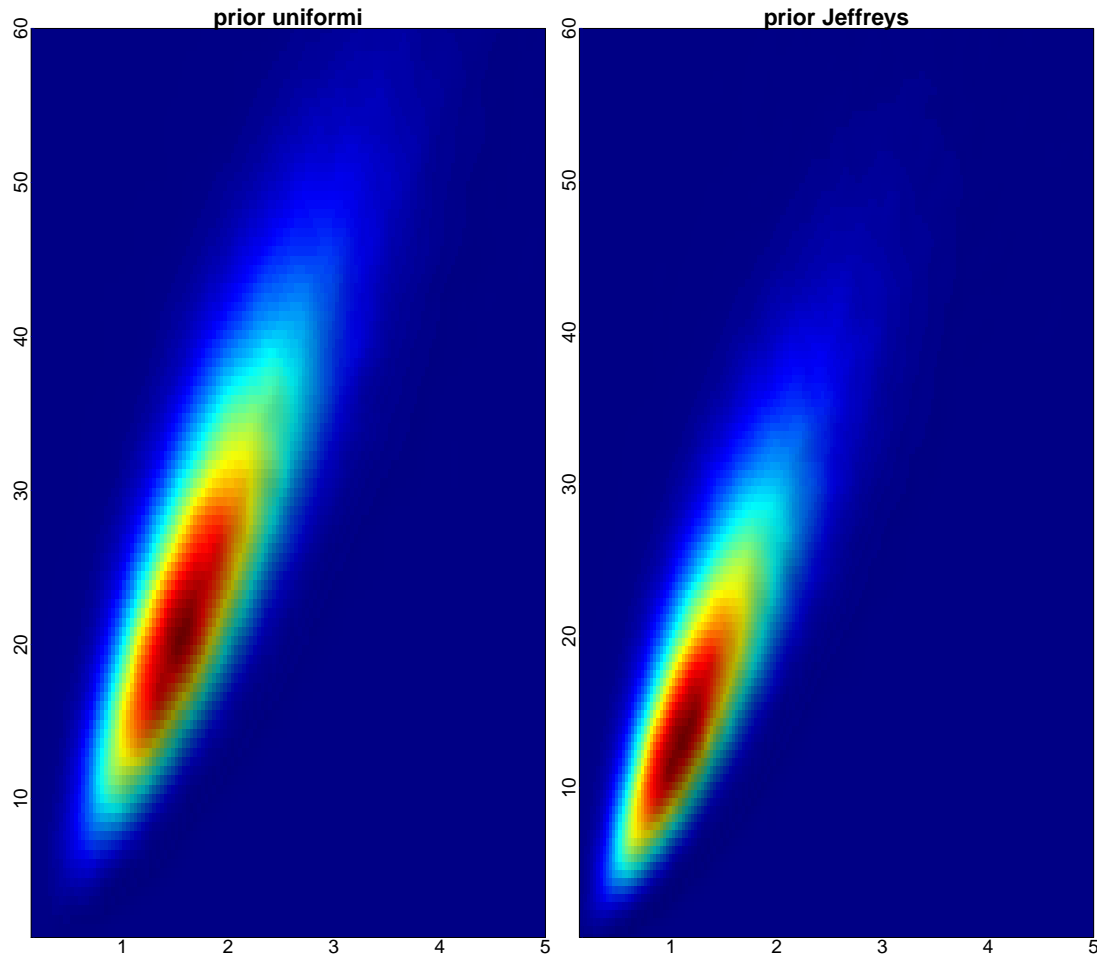


Figura B.12: Distribuzioni a posteriori (prior uniformi e di Jeffreys), associate alla v.a. $\tilde{\rho}$; l'asse delle ascisse si riferisce alla v.a. \tilde{C} , mentre l'asse delle ordinate alla variabile \tilde{D} .

La tabella riporta i valori attesi delle distribuzioni a posteriori considerando le due tipologie di distribuzioni a priori adottate.

Lob: ADB

Prior	Uniforme	Jeffreys
$E(\tilde{C} \tilde{\rho})$	1,947	1,509
$E(\tilde{D} \tilde{\rho})$	28,218	20,763

Tabella B.8: Lob ADB - Valori attesi delle distribuzioni a posteriori.

Appendice C

Gamma misturata

Sia \tilde{X} una variabile aleatoria Gamma di parametri $\alpha > 0$ e $\beta > 0$. La funzione di densità è:

$$f_{\tilde{X}}(x) = \frac{\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta x}, \quad x \in \mathbb{R}^+$$

dove $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$ è la funzione gamma.

Il momento semplice di ordine j è definito come,

$$E(\tilde{X}^j) = \frac{\alpha(\alpha+1)\dots(\alpha+j-1)}{\beta^j}$$

mentre il cumulante di ordine j risulta,

$$K_j(\tilde{X}) = \frac{(j-1)!\alpha}{\beta^j}$$

Di seguito si ricavano le caratteristiche del fattore di disturbo \tilde{q} (considerazioni analoghe valgono per la variabile \tilde{p}). La v.a. \tilde{q} associata al numero dei sinistri è una variabile Gamma di parametri uguali:

$$\tilde{q} \sim \text{Gamma} \left(\frac{1}{\tilde{\sigma}_q^2}; \frac{1}{\tilde{\sigma}_q^2} \right)$$

dove $\tilde{\sigma}_q^2$ è a sua volta una variabile casuale. In termini generali è possibile calcolare i momenti della variabile sistematica \tilde{q} senza conoscere la forma distributiva della v.a. $\tilde{\sigma}_q^2$.

Il valore atteso risulta essere,

$$E(\tilde{q}) = E_{\tilde{\sigma}_q^2} [E_{\tilde{q}}(\tilde{q} | \tilde{\sigma}_q^2)] = E_{\tilde{\sigma}_q^2} \left[\frac{\tilde{\sigma}_q^2}{\tilde{\sigma}_q^2} \right] = 1$$

quindi, l'aspettativa del fattore di disturbo rimane pari ad 1.

La varianza è ottenuta come segue:

$$\begin{aligned} Var(\tilde{q}) &= E_{\tilde{\sigma}_q} [Var_{\tilde{q}}(\tilde{q} | \tilde{\sigma}_q)] + Var_{\tilde{\sigma}_q} [E_{\tilde{q}}(\tilde{q} | \tilde{\sigma}_q)] \\ &= E_{\tilde{\sigma}_q} (\tilde{\sigma}_q^2) + Var_{\tilde{\sigma}_q} (1) = E_{\tilde{\sigma}_q} (\tilde{\sigma}_q^2) \end{aligned}$$

La media quadratica della v.a. $\tilde{\sigma}_q$, quindi, rappresenta la varianza della variabile \tilde{q} .

Il coefficiente di variazione è pari alla radice quadrata della varianza:

$$CV(\tilde{q}) = \sqrt{E_{\tilde{\sigma}_q}(\tilde{\sigma}_q^2)}$$

Infine, l'indice di asimmetria è ottenuto come segue:

$$\begin{aligned} \gamma(\tilde{q}) &= \frac{\mu_3(\tilde{q})}{[Var(\tilde{q})]^{\frac{3}{2}}} = \frac{E(\tilde{q}^3) - 3E(\tilde{q})Var(\tilde{q}) - E(\tilde{q})^3}{[Var(\tilde{q})]^{\frac{3}{2}}} \\ &= \frac{1 + 3E_{\tilde{\sigma}_q}(\tilde{\sigma}_q^2) + 2E_{\tilde{\sigma}_q}(\tilde{\sigma}_q^4) - 3E_{\tilde{\sigma}_q}(\tilde{\sigma}_q^2) - 1}{[Var(\tilde{q})]^{\frac{3}{2}}} = \frac{2E_{\tilde{\sigma}_q}(\tilde{\sigma}_q^4)}{[Var(\tilde{q})]^{\frac{3}{2}}} \end{aligned}$$

dove vale la relazione $E(\tilde{q}^3) = E_{\tilde{\sigma}_q} [E_{\tilde{q}}(\tilde{q}^3 | \tilde{\sigma}_q)] = E_{\tilde{\sigma}_q} (1 + 3\tilde{\sigma}_q^2 + 2\tilde{\sigma}_q^4)$. Si noti che il cumulante di ordine terzo della v.a. \tilde{q} , che individua il numeratore dell'asimmetria, dipende solo dal momento semplice di ordine quarto della variabile $\tilde{\sigma}_q$.

Sotto l'ipotesi che la v.a. $\tilde{\sigma}_q$ sia distribuita in accordo con una variabile Gamma di parametri α e β , $\tilde{\sigma}_q \sim Gamma(\alpha; \beta)$, le caratteristiche del fattore di disturbo possono essere riformulate come segue.

Varianza:

$$Var(\tilde{q}) = \frac{\alpha(\alpha + 1)}{\beta^2}$$

Coefficiente di variazione:

$$CV(\tilde{q}) = \frac{\sqrt{\alpha(\alpha + 1)}}{\beta}$$

Indice di asimmetria:

$$\gamma(\tilde{q}) = \frac{2(\alpha + 2)(\alpha + 3)}{\beta\sqrt{\alpha(\alpha + 1)}}$$

La figura seguente permette di apprezzare i valori del coefficiente di variazione e dell'indice di asimmetria del fattore di disturbo al variare dei parametri α e β della variabile misturante $\tilde{\sigma} \sim Gamma(\alpha; \beta)$.

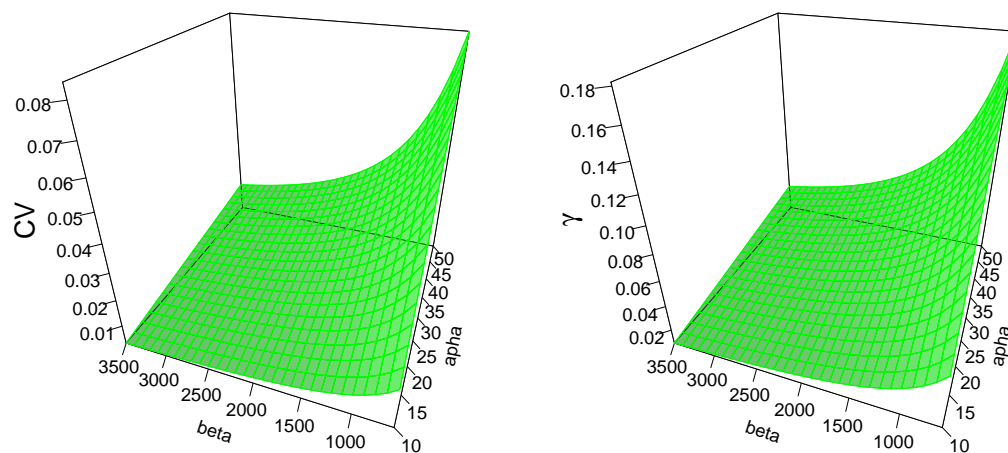


Figura C.1: Coefficiente di variazione e indice di asimmetria del fattore di disturbo al variare dei parametri α e β della variabile Gamma misturante.

Notazione

\tilde{X}	variabile aleatoria
$F_{\tilde{X}}(x), F(x)$	funzione di ripartizione
$F_{\tilde{X}}^{-1}(x), F^{-1}(x)$	funzione di ripartizione inversa (funzione quantilica)
$F^{(-1)}(x)$	funzione quasi-inversa di F
$f_{\tilde{X}}(x), f(x)$	funzione di densità
$E(\tilde{X}), E$	valore atteso
$var(\tilde{X}), var$	varianza
$std(\tilde{X}), std$	scarto quadratico medio
$cov(\tilde{X}; \tilde{Y}), cov$	covarianza
$CV(\tilde{X}), CV$	coefficiente di variazione
$q_{\alpha}(\tilde{X}), q_{\alpha}$	quantile di ordine α
$C_{\tilde{X}, \tilde{Y}}, C(u, v), C$	funzione di ripartizione della copula
$c_{\tilde{X}, \tilde{Y}}, c(u, v), c$	funzione di densità della copula
Π	copula prodotto
M	copula limite superiore di Fréchet
W	copula limite inferiore di Fréchet
φ	generatore delle copule archimedee
$\varphi^{[-1]}$	funzione pseudo-inversa di φ
$\tau(\tilde{X}, \tilde{Y}), \tau$	misura di concordanza di Kendall
$\rho_S(\tilde{X}, \tilde{Y}), \rho_S$	misura di concordanza di Spearman
$\rho(\tilde{X}, \tilde{Y}), \rho$	coefficiente di correlazione lineare di Pearson

$\lambda_U(\tilde{X}, \tilde{Y}), \lambda_U$	coefficiente di dipendenza di coda superiore
$\lambda_L(\tilde{X}, \tilde{Y}), \lambda_L$	coefficiente di dipendenza di coda inferiore
\mathbf{I}	matrice identità
\mathbb{I}	intervallo unitario $[0, 1]$
\mathbb{I}^2	quadrato unitario, $[0, 1] \times [0, 1]$
\mathbb{R}	insieme dei numeri reali
\mathbb{R}^2	piano dei numeri reali, $\mathbb{R} \times \mathbb{R}$
\mathbb{R}^+	insieme dei numeri reali positivi