

Sull'analisi del Dynamic Policyholder Behaviour nei riscatti attraverso le Copule bivariate

Fabio Baione *

Andrea Santoro †

Keywords: Dynamic PolicyHolder Behaviour, Copula Regression, Lapse, Surrender
JEL Classification: G22, G40, C25

Sommario

L'evoluzione della normativa e la sempre maggiore importanza attribuita al conseguimento di una stima accurata dei flussi di cassa di una compagnia di assicurazione, porta alla necessità di analizzare più dettagliatamente diversi elementi dell'attività della stessa. In particolare, in un framework market-consistent, la valutazione dei flussi di cassa è coerente con la valutazione proposta sul mercato finanziario e la corretta modellizzazione del fenomeno di interesse permette di prevederne le dinamiche future, reagendo repentinamente ad eventuali movimenti sfavorevoli dei flussi di attività e passività.

Questo articolo si prefigge come obiettivo l'impiego di una metodologia di stima basata sulle copule per poter descrivere il Dynamic PolicyHolder Behaviour (DPHB), ovvero la componente del comportamento dell'assicurato legata agli sviluppi economico-finanziari, con particolare riferimento al fenomeno dei riscatti, osservando come le oscillazioni di un benchmark di mercato possono incidere sul comportamento degli assicurati nell'esercizio dell'opzione di riscatto.

La presente analisi, quindi, propone un approccio alternativo di tipo regressivo sul tema dei riscatti dinamici mostrando, attraverso un'applicazione su dati reali, il possibile utilizzo della copula-regression, considerando che tale approccio non trova riscontro né nella letteratura né nella pratica attuariale.

*Dipartimento di Statistica, Sapienza Università di Roma, fabio.baione@uniroma1.it

†Sapienza Università di Roma

1 Introduzione

1.1 Premessa

Il susseguirsi di crisi finanziarie ha elevato l'importanza dell'analisi e gestione del rischio: in particolare tale tematica in ambito assicurativo, con l'introduzione di Solvency II [IVASS], ha ottenuto sempre più rilevanza, diventando centrale per la sana, prudente e corretta gestione dell'impresa. Da un punto di vista normativo, vi è stata l'esigenza di effettuare ipotesi il più possibile realistiche e la necessità di valutare le diverse opzioni presenti nei contratti assicurativi che, se esercitate, possono influenzare significativamente i flussi di cassa dell'assicuratore. Tra queste opzioni, quella di riscatto è collegata ad uno dei principali rischi derivanti dal comportamento dell'assicurato: il *lapse risk*.

La definizione di *lapse risk* non è univoca, essa assume diversa connotazione nella letteratura: nel già citato framework SolvencyII, il modulo del *lapse risk* include tutte le opzioni legali e/o contrattuali che possono portare alla terminazione, riduzione, limitazione o sospensione della copertura assicurativa, nonché opzioni che consentono la costituzione, il rinnovo, l'aumento, la proroga o la ripresa della copertura stessa.

Per tale ragione, si può affermare che i riscatti costituiscono solo una delle possibili cause del fenomeno indicato dal termine *lapse*; tuttavia, in virtù del fatto che – come spiegato in [Milhaud & Dutang 2018] – molti studi mostrano come per gli aspetti dinamici il tasso di riscatto sia preponderante nel descrivere il *lapse risk* all'interno delle compagnie vita, si è preferito concentrare l'attenzione su questa componente.

I riscatti, se non correttamente gestiti, possono costituire un possibile danno per l'attività assicurativa e i rischi da essi derivanti sono di grande valenza economica per diverse ragioni, infatti:

1. qualora dovessero esserci costi o spese iniziali sostenuti dalla compagnia e non ancora coperti con l'acquisizione dei premi, vi è il rischio che questi non vengano recuperati;
2. in caso di riscatti di massa, possono sorgere problemi di liquidità, legati alla richiesta di denaro da parte dell'assicurato e la necessità di vendere attività per onorare l'impegno;
3. vi è la possibilità di perdita dei profitti futuri, specialmente nel caso di contratti molto profittevoli;
4. può sorgere un problema di *adverse selection* perché gli assicurati che sono a conoscenza della loro situazione e della loro necessità di copertura assicurativa tendono a non riscattare la propria polizza. Tipicamente, individui che hanno necessità di una copertura assicurativa – i.e. maggiormente soggetti al rischio – riscattano con una frequenza ridotta in virtù di questa loro esposizione e, per tale motivo, si va a creare una anti-selezione nel momento in cui i tassi diventano alti a sufficienza da rendere conveniente il riscatto per soggetti non particolarmente esposti al rischio;
5. ultimo, ma non meno importante, può sorgere un danno reputazionale laddove si osservi un elevato numero di riscatti.

Per limitare tale fenomeno e recuperare i costi inizialmente sostenuti, le compagnie inseriscono tipicamente delle penali che disincentivano economicamente l'esercizio del riscatto.

Nell'ambito delle assicurazioni sulla vita umana, caratterizzati da durate di medio-lungo periodo è usuale distinguere il comportamento assunto dagli assicurati nei confronti del riscatto nelle seguenti componenti¹:

- irrazionale o strutturale: usualmente di natura endogena, si considera come una componente di "base" ed è legata a processi decisionali interni al soggetto ed è, quindi, considerata indipendente da fattori esogeni quale, ad esempio, l'andamento del mercato finanziario. In tal senso è definita come scenario-independent.
- razionale o tipica: usualmente di natura esogena, caratterizza il modello dinamico da quello statico, collega il surrender rate al comportamento razionale dell'assicurato che risulta scenario-specific. In questo modo si ha un legame diretto con l'ambiente finanziario e l'assicurato viene visto come un soggetto razionale in grado di riscattare quando risulta vantaggioso (i.e. il valore di riscatto è superiore a quello di continuazione).². In altre parole, nel comportamento razionale si assume che il soggetto assicurato:
 - si comporti come un investitore neutrale al rischio che decide di riscattare nel momento in cui il valore di esercizio dell'opzione supera il valore di continuazione del contratto³.
 - si comporti come un investitore avverso al rischio e, in tal caso, il suo obiettivo è massimizzare l'utilità attesa.

Risulta superfluo sottolineare che gli assicurati nella realtà non sono mai completamente razionali, perché se così fosse non sarebbero spiegati alcuni comportamenti come la prosecuzione di un contratto nel momento in cui il valore di riscatto supera quello di continuazione. Per tale motivo è preferibile, sotto il profilo operativo, utilizzare modelli sub-ottimali.

Il modello proposto in questa sede rientra nella categoria dei modelli sub-ottimali, che descrivono il comportamento del tasso di riscatto non solo in funzione di fattori economici esogeni, ma anche di elementi endogeni quali ad esempio le caratteristiche contrattuali e dell'assicurato. In tale contesto, è possibile distinguere due differenti approcci di calcolo

- one-stage model: in questa tipologia di modelli si va a stimare la frequenza di riscatti, sempre considerando la presenza di una componente statica e una dinamica, in un unico step.
- two-stage model: si stima prima la componente di base (idiosincronica) riguardante le caratteristiche della polizza e quelle individuali del contraente. Poi si tiene conto di un fattore correttivo – in letteratura (si veda ad esempio [Eling & Kochanski 2013]) identificato dai Dynamic Lapse Multipliers (DLM) - che considera elementi di carattere economico-finanziario

Il vantaggio di questo secondo approccio risiede nella possibilità di avere una maggiore consistenza economica nel rispetto dei vincoli presenti per la relazione funzionale tra il tasso di riscatto e spread di mercato – anche se da un punto di vista puramente statistico, questi vincoli possono risultare penalizzanti.

¹Si veda ad esempio [Cerchiara et al. 2008]

²Da un punto di vista normativo, SolvencyII [IVASS] e i principi CFO [CFO] richiedono esplicitamente di analizzare la dipendenza tra il lapse risk e altri rischi tipici dell'impresa, come ad esempio il market risk.

³Seguendo la stessa logica presente nelle opzioni di tipo Americano.

1.2 Una breve review della letteratura sul policyholder behaviour per i riscatti

Considerata la rilevanza dell'argomento, la modellizzazione del surrender rate è ampiamente trattata in letteratura, sia in casi specifici che in condizioni più generali. Gli articoli di seguito riportati sono stati selezionati sia per fornire un esempio della pluralità di casi in cui l'analisi del tasso di riscatto può inserirsi, sia per mostrare come questo risulta condizionato dagli andamenti economico-finanziari in contesti specifici.

In [Kim 2005] si osserva come, oltre all'andamento dei tassi di interesse, vi siano altri elementi esogeni rilevanti ai fini della stima dei tassi di riscatto come i tassi di disoccupazione, tassi di crescita economica, effetti stagionali e l'antidurata (ovvero l'età della polizza rispetto alla data di emissione). Tuttavia, l'autore precisa come non vi sia un unico modello corretto per descrivere l'andamento dei tassi di riscatto, ma che questo dipenda anche dalla tipologia di contratto assicurativo che si considera e dalle variabili a disposizione.

In [Milhaud, Loisel & Maume-Deschamps 2011] si cercano delle caratteristiche statiche discriminanti per la tematica dei riscatti. Gli autori hanno evidenziato come le persone più anziane e gli assicurati che pagano un premio periodico tendano a riscattare maggiormente, inoltre, per motivazioni evidentemente differenti, gli individui molto poveri e molto ricchi non sono interessati a riscattare⁴. E' stato inoltre osservato come il maggior numero di riscatti avviene temporalmente quando cessano i vincoli economici, come le penali. In conclusione, lo studio ha dimostrato come gli assicurati che presentavano l'opzione del beneficio senza scopo di lucro riscattavano in misura inferiore rispetto a coloro che presentavano l'opzione del beneficio del profitto (PB), ciò perché se si è interessati al profitto è ragionevole supporre che si passi al prodotto che fornisce un profitto maggiore, oppure che si riscatti quando il valore supera il rendimento⁵.

In [Pinquet, Guillén & Ayuso 2011] viene osservato il rischio di riscatti all'interno di un portafoglio in run-off (i.e. chiuso al nuovo business) di contratti assicurativi individuali Long Term Care⁶ di una compagnia spagnola, con copertura non pienamente finanziata (not fully funded). Una particolarità evidenziata per la maggior parte dei contratti di questa tipologia, è l'elevato valore delle spese iniziali sostenute dall'assicurato (i.e. valore elevato del premio) e, per tale motivo, risulta necessaria una disamina del comportamento (lapse behavior) in quanto vi è il rischio concreto che riscattando l'assicurato possa subire perdite. Gli autori giungono alla conclusione che una conoscenza insufficiente dei prodotti finanziari possa essere una delle principali cause dei riscatti e perciò vi è la necessità di un miglioramento della conoscenza del rischio e dei prodotti per coprirlo.

In [MacKay et al. 2017] gli autori si concentrano sul lapse risk sotto un'altra luce, infatti in questo articolo si osserva come l'assicuratore - o meglio, la compagnia assicuratrice - può ridurre questo rischio mediante un opportuno design del prodotto assicurativo venduto, in particolare sulle rendite variabili, ovvero prodotti assicurativi a contenuto finanziario. L'obiettivo degli autori è, infatti, spiegare come costruire un prodotto che sia vendibile sul mercato e appetibile per gli assicurati, ma allo stesso tempo che disincentivi l'esercizio dell'opzione di riscatto anche nel worst case: in tal modo, proprio perché si assume che l'assicurato rimanga nel portafoglio, non sarebbe richiesto all'assicuratore di modellare il riscatto in fase di pricing e hedging del

⁴I primi per la perdita che avrebbero pagando le penali, i secondi perché non interessati.

⁵Si potrebbe ad esempio fare un parallelo con il valore di continuazione e quello di esercizio.

⁶Come noto, questa tipologia contrattuale copre rischi rari e la fine del contratto è stabilita unilateralmente, per tutelare l'assicurato.

contratto. Questo ha un duplice vantaggio: da una parte si riduce il rischio di scegliere il modello sbagliato per il lapse, dall'altra si semplifica la fase di risk management del contratto.

In [Nolte & Schneider 2017] si studia la propensione al riscatto di un collettivo di assicurati con contratti vita di tipo partecipativo (i.e. con partecipazione agli utili) andando ad osservare il legame che si ha tra il riscatto e la cultura finanziaria, la consulenza finanziaria e il processo decisionale euristico a seguito di uno shock sul reddito⁷. L'ipotesi del riscatto, ovviamente, si concretizza nel momento in cui l'assicurato che ha subito lo shock non ha fondi a sufficienza per coprire la perdita subita e trova nel valore di riscatto della polizza una soluzione al problema. Gli autori hanno osservato come, anche se alcuni investitori scelgono di riscattare il contratto come ultima possibilità, ve ne sono molti che anche in presenza di altre attività liquide preferiscono comunque coprire perdite inattese mediante il riscatto contrattuale e ciò va a confutare l'idea che delle necessità economiche possano spiegare il lapse risk nella sua interezza. Inoltre, un maggiore affidamento alle decisioni euristiche aumenta la probabilità di riscatto, mentre la cultura finanziaria ha invece un effetto riduttivo. Discorso a parte viene effettuato per la consulenza finanziaria, che riduce la probabilità di riscatto, ma solo in tempi di difficoltà finanziarie, i.e. in situazioni di shock.

In [Baione et al. 2018] viene considerato il problema di stima dei lapse rates dinamici in funzione delle condizioni del mercato finanziario, utilizzando una funzione originale a doppia sigmoide per rappresentare il legame tra il conjectural rate e uno spread di mercato e compatibili con ipotesi comportamentali degli assicurati di tipo economico. Una estensione è proposta dagli stessi autori ([Baione et al. 2021]) dove l'analisi viene ampliata considerando da un lato un confronto tra diverse funzioni impiegate nella pratica attuariale dall'altro approfondendo il modello a doppia sigmoide. Su quest'ultimo punto è trattato il tema delle proprietà matematiche della funzione a doppia sigmoide nonché stima dei parametri mostrando tramite un'applicazione che il modello proposto sia da considerarsi un valido strumento per descrivere il fenomeno riscatti dinamici degli assicurati.

In [Milhaud & Dutang 2018] si cerca di fornire una previsione dei tempi di riscatto dell'assicurato, utilizzando un approccio a rischi concorrenti, ovvero adottando una analisi della probabilità di permanenza/sopravvivenza dei rischi nel portafoglio relativa ai riscatti, utilizzando diverse fonti di rischio tra loro indipendenti e che si escludono a vicenda. Viene inoltre illustrato come la stima dell'intensità dei riscatti può subire delle distorsioni in seguito alla stima errata delle intensità delle altre fonti di rischio (seppur meno probabili).

1.3 Obiettivi del lavoro

Obiettivo del presente lavoro è di introdurre un approccio metodologico originale ed alternativo a quelli presenti in letteratura per la modellizzazione della parte dinamica del fenomeno dei riscatti presente all'interno dei modelli di tipo sub-ottimali a due fasi. In tale contesto, l'approccio proposto si basa sull'adozione delle copule bivariate per definire la relazione tra il fenomeno oggetto di analisi, il riscatto, e la variabile economico-finanziaria di riferimento. Tale approccio consente di ottenere vantaggi dovuti alla facoltà di:

- effettuare ipotesi sulle distribuzioni marginali delle due variabili di riferimento;
- definire una struttura di dipendenza specifica attraverso l'identificazione della funzione copula più appropriata;

⁷Questo infatti, come spiegato dagli autori, comporta una necessità economica e porta a riconsiderare le decisioni finanziarie.

- impiegare le proprietà della distribuzione congiunta per ottenere le distribuzioni marginali.

Quest'ultimo elemento è di particolare rilievo in quanto, come dimostrato nell'elaborato, è possibile impiegare modelli di copula regression per studiare la dipendenza sia in media che sui quantili delle distribuzioni marginali. Il lavoro è organizzato come segue. Nella Sezione 2 viene introdotto il contesto metodologico con particolare riferimento ai modelli Two-Stage, alle copule bivariate e alle relative metodologie di regressioni in media e quantiliche. Nella Sezione 3 viene discussa un'applicazione numerica effettuata su un portafoglio assicurativo vita in cui si impiega la metodologia di stima della copula su diverse famiglie di prodotti assicurativi. La Sezione 4 riporta le conclusioni e i futuri spunti di estensione della ricerca.

2 Contesto metodologico

2.1 I Two-Stage Models

Come indicato nella Sezione 1.1 i modelli sub-ottimali, possono essere caratterizzati da un processo metodologico a due fasi (Two-Stages) in cui la modellizzazione dinamica del tasso di riscatto assume la seguente formulazione:

$$r_t = r_t^{Base} \cdot \psi(\Delta_{mkt,t}; \boldsymbol{\beta}) \quad (1)$$

dove, il processo di calcolo delle diverse grandezze presenti nell'Eq. (1) è realizzato tramite i seguenti passi operativi:

- Stage 1: viene calcolato, anche tramite metodi deterministici, un tasso di riscatto base r_t^{Base} di natura strutturale, riguardante ad aspetti irrazionali e funzione delle caratteristiche del contratto e/o dell'assicurato.
- Stage 2: viene determinato il fattore di aggiustamento scenario-specifico $\psi(\Delta_{mkt,t}; \boldsymbol{\beta})$, o tasso congiunturale, funzione di un set di parametri $\boldsymbol{\beta}$ e, soprattutto, di una variabile di mercato $\Delta_{mkt,t}$; in particolare, quest'ultimo è calcolato come differenza tra il valore del rendimento di un benchmark di mercato finanziario ed il rendimento realizzato dalla compagnia di assicurazione e riconosciuto alla polizza ai fini della rivalutazione delle prestazioni.

Risulta evidente che l'utilizzo dell'Eq. (1) sottintende l'ipotesi di indipendenza tra la componente finanziaria – definita da ψ – e quella tecnica del tasso di riscatto – r_t^{Base} , qualsiasi sia la metodologia di stima per definire tali componenti.⁸

A seconda del valore assunto $\Delta_{mkt,t}$, ovviamente il valore del fattore correttivo sarà più o meno grande e, in particolare, poiché r_t assume valori compresi tra 0 e 1, si devono avere dei limiti per $\psi(\Delta_{mkt,t}; \boldsymbol{\beta})$, ovvero

$$r_t \in [0, 1] \quad \implies \quad \psi(\Delta_{mkt,t}; \boldsymbol{\beta}) \in \left[0, \frac{1}{r_t^{Base}} \right]$$

Per alleggerire la notazione delle formule, si adotterà $\psi(\Delta_{mkt,t}; \boldsymbol{\beta}) = \psi_t$.

Nel seguito si propone un'approccio per la definizione del tasso congiunturale basato sulle copule di cui si introducono le principali caratteristiche utili alla formalizzazione del problema.

⁸Questa è una ipotesi forte, che risulta necessaria in queste sede, ma che può essere rimossa qualora si disponga di un dataset più ampio con un maggior numero di variabili.

2.2 Le Copule statistiche

Teorema 2.1. Teorema di Sklar (caso bidimensionale)

Sia F una funzione di distribuzione congiunta con marginali G_1 e G_2 , allora esiste una copula C tale che

$$F(x_1, x_2) = C(G_1(x_1), G_2(x_2)) \quad \forall x_1, x_2 \in \bar{\mathcal{R}} \quad (2)$$

dove con $\bar{\mathcal{R}}$ si è indicato il sistema esteso dei numeri reali, ovvero il sistema che si ottiene aggiungendo a \mathcal{R} i termini $+\infty$ e $-\infty$ trattandoli come numeri reali.

Analogamente, è possibile definire F a partire dalle altre funzioni: se C è una funzione copula e si conoscono le distribuzioni marginali G_1 e G_2 , allora F definita nella Eq. (2) è la loro funzione di distribuzione congiunta.

In particolare, essendo in questo contesto G_1 e G_2 continue, la funzione C è unica.

Definizione 2.1. Dato un vettore di variabili aleatorie bidimensionale (X_1, X_2) nello spazio di probabilità (Ω, \mathcal{F}, P) , la copula C di (X_1, X_2) è una funzione $C : [0, 1]^2 \rightarrow [0, 1]$ che corrisponde alla distribuzione congiunta delle trasformazioni uniformi U_1 e U_2

$$\begin{aligned} C(u_1, u_2) &= P[U_1 < u_1, U_2 < u_2] \\ &= P[G_1(X_1) < u_1, G_2(X_2) < u_2] \end{aligned}$$

dove u_1 e u_2 sono, rispettivamente, $u_1 = G_1(x_1)$ e $u_2 = G_2(x_2)$. Se G_1 (risp. G_2) è una funzione di ripartizione continua, allora la variabile aleatoria $G_1(x_1)$ (risp. $G_2(x_2)$) segue una distribuzione uniforme in $[0, 1]$

$$G_1(X_1; x_1) \sim Unif(0, 1) \quad G_2(X_2; x_2) \sim Unif(0, 1)$$

L'analisi di dati multivariati mediante l'applicazione delle copule statistiche consente di studiare la distribuzione congiunta di più variabili definendo separatamente le distribuzioni marginali di ciascuna variabile e la struttura di dipendenza presente tra queste.

Poiché i dati a disposizione non sono distribuiti secondo uniformi in $[0, 1]$, è opportuno utilizzare la trasformazione integrale di probabilità per standardizzare le variabili aleatorie passando da variabili appartenenti ad una qualsiasi distribuzione, a variabili che seguono una distribuzione uniforme⁹.

Ovviamente, essendo le scelte di G_1 e G_2 basate sulla stima di verosimiglianza il risultato che ne consegue sarà una approssimazione: usando una *estimated probability integral transform* si costruiscono degli pseudo-copula data.

$$\begin{cases} \hat{G}_1(X_1) = U_1 & \sim Unif(0, 1) \\ \hat{G}_2(X_2) = U_2 & \sim Unif(0, 1) \end{cases} \quad (3)$$

Omettendo per comodità la notazione di stima (il "cappello"), nel caso in esame risulta evidente che $X_1 = G_1^{-1}(U_1)$ e $X_2 = G_2^{-1}(U_2)$.

⁹La trasformazione risulta perfettamente corretta se la distribuzione utilizzata è la vera distribuzione della variabile aleatoria, tuttavia nei casi pratici non si conosce a priori la distribuzione dei dati e perciò spesso si ottengono risultati approssimativamente corretti, per campioni sufficientemente grandi, utilizzando la distribuzione che meglio fitta i dati a disposizione.

Definizione 2.2. La funzione di densità della copula è definita come

$$c(u_1, u_2) = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2}$$

Per variabili aleatorie continue, come nel caso in esame, è possibile collegare la funzione di densità della copula c con la densità della distribuzione congiunta f

$$c(u_1, u_2) = f(G_1^{-1}(u_1), G_2^{-1}(u_2)) \cdot \left| \begin{bmatrix} \frac{\partial x_1}{\partial u_1} & \frac{\partial x_1}{\partial u_2} \\ \frac{\partial x_2}{\partial u_1} & \frac{\partial x_2}{\partial u_2} \end{bmatrix} \right| \quad (4)$$

da cui

$$c(u_1, u_2) = \frac{f(G_1^{-1}(u_1), G_2^{-1}(u_2))}{g_1(G_1^{-1}(u_1))g_2(G_2^{-1}(u_2))} = \frac{f(x_1, x_2)}{g_1(x_1)g_2(x_2)} \quad (5)$$

e di conseguenza è possibile, invertendo i termini, ottenere la densità congiunta a partire dalle altre grandezze come

$$f(x_1, x_2) = c(u_1, u_2)g_1(x_1)g_2(x_2) \quad (6)$$

Considerato che, da un punto di vista statistico, le copule consentono di descrivere e studiare misure di dipendenza senza scala tra le variabili catturandone qualsiasi struttura sottostante, è necessario impiegare una specifica misura di dipendenza. Tra queste, particolare rilevanza assume l'indice Tau di Kendall, per il quale la rappresentazione per variabili aleatorie assume la forma:

$$\begin{aligned} \tau_K(X_1, X_2) &= 4 \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(x_1, x_2) dF(x_1, x_2) \right] - 1 \\ &= 4 \left[\int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) \right] - 1 \end{aligned} \quad (7)$$

Risulta $\tau_K \in [-1, 1]$, dove il valore -1 è raggiunto in caso di perfetta dipendenza negativa e il valore 1 in caso di perfetta dipendenza positiva. Questo indice trova grande impiego nei metodi di stima del parametro nelle famiglie di copule uniparametriche, con il metodo "Inversione del Tau di Kendall".

2.3 Copula Regression

Sebbene l'analisi della distribuzione congiunta di due o più variabili possa costituire un ambito di notevole interesse, nella pratica può risultare più interessante spingersi oltre e utilizzare l'approccio mediante copule per stimare come una variabile si comporta in media al modificarsi delle altre: in questo contesto si parla di copula-regression. Infatti, nello svolgere una analisi sulla forma funzionale che lega una variabile dipendente a delle variabili risposta, è possibile analizzare tramite le copule sia la dinamica congiunta sia il comportamento marginale di una variabile condizionatamente alle altre.

Per definire la funzione di regressione copula in ambito bivariato è possibile riprendere la definizione data in [Sungur 2005] operando una semplice modifica della notazione per evitare di generare confusione.

Definizione 2.3. Sia (U_1, U_2) una variabile aleatoria bivariata con distribuzioni marginali uniformi in $[0,1]$ e copula C . La funzione copula regression di U_1 rispetto ad U_2 è definita come

$$r_C(u_2) = \mathbf{E}_C[U_1|U_2 = u_2] = 1 - \int_0^1 C_{u_2}(u_1)du_1 \quad (8)$$

Inoltre, date X_1 e X_2 con distribuzioni continue G_1 e G_2 , distribuzione congiunta F e copula C si ha che la distribuzione condizionata di U_1 , dato U_2 è

$$\begin{aligned} C_{u_2}(u_1) &= h_{1|2}(u_1|u_2) = P[U_1 \leq u_1|U_2 = u_2] \\ &= \frac{\partial C(u_1, u_2)}{\partial u_2} \end{aligned} \quad (9)$$

Osservazione 2.1. Come è possibile osservare nella Eq. (9), nel riferirsi alle funzioni copula condizionate – i.e. le derivate parziali di una copula – nella letteratura (ad esempio [Czado 2019]) si parla spesso di h -function.

Ovviamente, a partire da questi risultati è poi possibile risalire alle distribuzioni originarie delle variabili di interesse, ottenendo

(i) La funzione di regressione di X_1 rispetto a X_2

$$\begin{aligned} \hat{X}_1 &= \mathbf{E}[X_1|X_2 = x_2] \\ &= G_1^{-1} \left(1 - \int_0^1 C_{u_2}(u_1)du_1 \Big|_{u_2=G_2(x_2)} \right) \\ &= G_1^{-1}(r_C(G_2(x_2))) \end{aligned} \quad (10)$$

(ii) La distribuzione condizionata di X_1 , dato X_2

$$\begin{aligned} P[X_1 \leq x_1|X_2 = x_2] &= P[U_1 \leq G_1(x_1)|U_2 = G_2(x_2)] \\ &= \frac{\partial C(u_1, u_2)}{\partial u_2} \Big|_{u_1=G_1(x_1), u_2=G_2(x_2)} \\ &= C_{u_2}(u_1) \Big|_{u_1=G_1(x_1), u_2=G_2(x_2)} \end{aligned} \quad (11)$$

2.4 Quantile Copula Regression

Un ulteriore vantaggio di utilizzare le copule per un approccio di tipo regressione risiede nella possibilità di ottenere non solo una stima della media condizionata ma anche dei quantili condizionati. Infatti, per poter ottenere una regressione dei quantili tramite le copule è possibile impiegare la seguente espressione proposta in [Kraus & Czado 2017]:

$$\begin{aligned} q_\alpha(x_2) &= G_{X_1|X_2}^{-1}(\alpha|x_2) = \\ &= \hat{G}_1^{-1}(\hat{C}_{U_1|U_2}^{-1}(\alpha|\hat{u}_2)) \end{aligned} \quad (12)$$

dove $\alpha \in (0, 1)$ corrisponde al livello di probabilità che si intende considerare per la stima del relativo quantile q_α .

Particolarmente interessante è la forma che assume l'intervallo di confidenza definito con questo approccio: le traiettorie degli estremi superiori e inferiori dell'intervallo, infatti, non sono semplicemente una traslazione dovuta ad uno shift additivo della curva dei valori centrali, ma

hanno una propria forma dovuta alla struttura probabilistica e di dipendenza presente all'interno del dataset, come si avrà modo di apprezzare meglio nella successiva sezione.

La rilevanza di tale strumento è evidente se si pensa, ad esempio, a sue applicazioni per fenomeni con variabili aventi distribuzione asimmetrica o a dati che presentano una struttura di dipendenza maggiore sulle code.

3 Un'applicazione delle copule bivariate per la stima del Dynamic Policyholder Behaviour

Nel seguito si riporta un'applicazione su un dataset reale riferito ai riscatti osservati su un portafoglio di polizze di una compagnia di assicurazioni sulla vita italiana.

L'obiettivo è di modellare il Dynamic PolicyHolder Behaviour (DPHB) impiegando le copule bivariate e il modello di regressione introdotti nella precedente sezione; in particolare, è interessante studiare come il modello considerato descriva il DPHB senza considerare un determinato comportamento economico, come avviene ad esempio nel modello doppia-sigmoide proposto in [Baione et al. 2018] e [Baione et al. 2021]¹⁰.

Allo scopo di impiegare le copule bivariate nel contesto dei Two-Stage models introdotti nella sezione 2.1 è, quindi, sufficiente porre $X_1 = \psi_t$ e $X_2 = \Delta_{mkt,t}$, con funzioni di ripartizione G_1 e G_2 – le cui densità sono $g_1(\cdot)$ e $g_2(\cdot)$ – e distribuzione congiunta F

$$G_1(x_1) = P[X_1 \leq x_1] \quad G_2(x) = P[X_2 \leq x_2]$$

$$F(x_1, x_2) = P[X_1 \leq x_1, X_2 \leq x_2]$$

dove si assume che la funzione F abbia le derivate di primo e secondo ordine continue.

3.1 Descrizione del dataset

Il portafoglio è composto da prodotti assicurativi rivalutabili collegati ad una specifica gestione separata e caratterizzata dalle seguenti forme tecniche:

1. Capitale Differito
2. Capitalizzazione
3. Mista
4. Rendita Differita
5. Termine Fisso
6. Vita Intera

Per ciascuna forma tecnica, il dataset contiene la serie storica dei tassi di riscatto mensili r_t osservati nel periodo compreso tra il 01/01/2010 e il 01/12/2020. Inoltre, sono disponibili con la medesima cadenza i valori di spread $\Delta_{mkt,t}$, calcolato come differenza tra il rendimento di benchmark, il rendimento di un titolo di mercato BTP 10Y, e il rendimento accreditato

¹⁰Si tenga presente che a seconda del dataset un modello sarà preferito ad un altro, non vi è un unico modello sempre valido: si assume perciò che tutte le osservazioni che verranno effettuate lo saranno limitatamente ai dati osservati.

(retrocesso) alla polizza.

Nel seguito si riporta a titolo esemplificativo l'andamento dello spread per le polizze miste avendo osservato che nel calcolo dello spread vi sono lievi differenze tra le forme tecniche analizzate a causa di piccole differenze tra il tasso medio minimo garantito o alle aliquote medie di retrocessione.

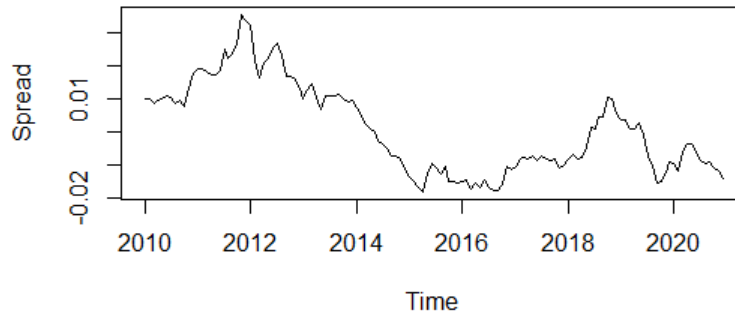


Figura 1: Serie storica degli spread per la forma tecnica Mista.

Non avendo a disposizione ulteriori informazioni sulle caratteristiche degli assicurati e dei contratti, per ciascuna forma contrattuale ($j = 1, \dots, 6$), la componente deterministica r_j^{Base} viene calcolata come media dei tassi di riscatto osservati sull'intero periodo di osservazione: $r_j^{Base} = \sum_{t=1}^n r_{j,t}/n, j = 1, \dots, 6$. Invertendo l'Eq. (1), sono stati così calcolati gli andamenti dei tassi congiunturali rappresentati in Figura 2.

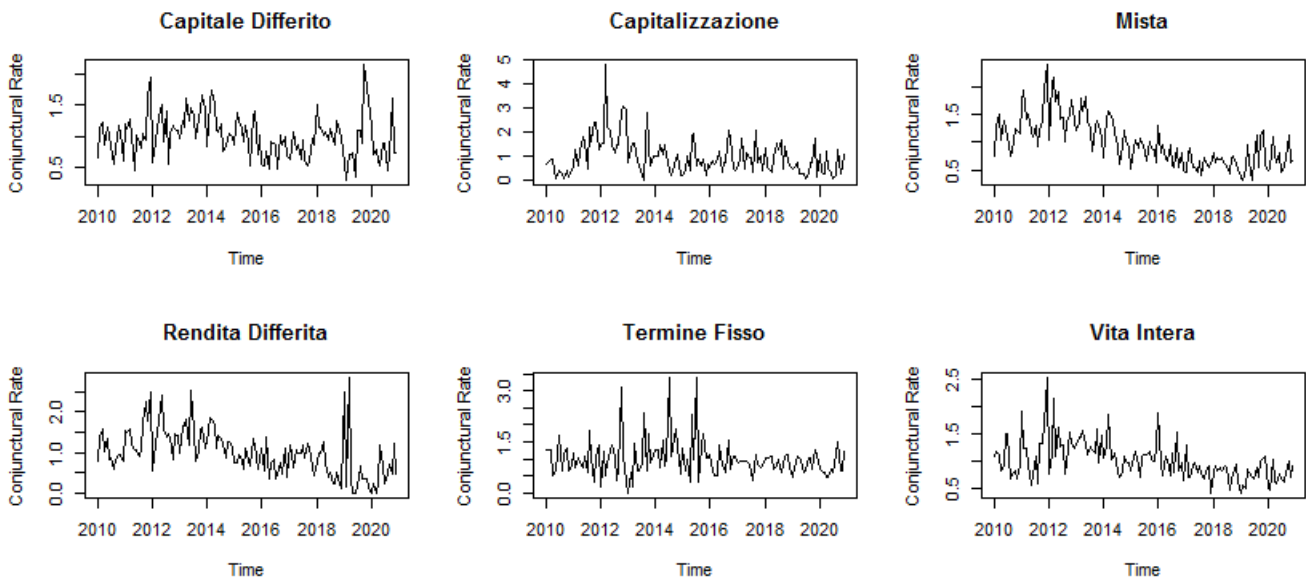


Figura 2: Serie storica dei conjunctural rates per le forme tecniche

Com'è possibile osservare in Figura 2, appare evidente che non tutte le forme contrattuali esaminate mostrano, in termine di riscatti, lo stesso comportamento. Considerato invece che, come affermato, gli spread non mostrano andamenti dissimili rispetto alla forma tecnica esaminata, si può supporre che le variazioni economico-finanziarie incidano in modo differente sulle diverse forme tecniche analizzate. Tale analisi grafica è confermata dagli indici statistici riportati in Tabella 1.

<i>Forma Tecnica</i>	<i>Min</i>	<i>Max</i>	<i>Mean</i>	<i>Dev. std.</i>
Capitale Differito	0.2866	2.1333	1.000	0.3359
Capitalizzazione	0.0000	4.8105	1.000	0.7463
Mista	0.2964	2.4061	1.000	0.4322
Termine Fisso	0.0000	3.3915	1.000	0.5260
Rendita Differita	0.0000	2.8472	1.000	0.5674
Vita Intera	0.3988	2.5351	1.000	0.3521

Tabella 1: Confronto tra le statistiche dei conjunctural rates delle forme tecniche

Dalla Tabella1 emerge un valore pari a 1 per ciascuna media e questa è evidentemente una conseguenza delle ipotesi effettuate in fase di creazione della variabile. Per quanto riguarda le altre statistiche, la deviazione standard è molto diversa tra le forme contrattuali e ciò è anche legato ai diversi range nei quali si spostano le traiettorie dei tassi congiunturali. Queste differenze rendono particolarmente interessante approfondire il tema della dipendenza tra le diverse forme contrattuali e la dinamica del benchmark.

3.2 La stima dei parametri delle copule bivariate

La prima fase di lavoro riguarda la stima dei parametri, sia delle distribuzioni G_1 e G_2 che della copula $C(., .)$. Nel presente studio è stato utilizzato il metodo della massima verosimiglianza ovvero

$$\hat{\boldsymbol{\theta}}_{MLE} = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmax}} \ell(\boldsymbol{\theta}; x_1, x_2)$$

con ℓ funzione di log-verosimiglianza. Gli stimatori così ottenuti, ovviamente, presentano le proprietà di consistenza e distribuzione asintotica normale.

Per effettuare questo tipo di analisi si è preferito procedere separando la fase di selezione delle distribuzioni e della stima dei relativi parametri, dalla fase di selezione e stima delle copule: operando in tal senso si definiscono dapprima le distribuzioni sottostanti e poi si stima la struttura di dipendenza.

In alternativa, sarebbe stato possibile lavorare iterativamente applicando la log-verosimiglianza al modello bivariato presente nella Eq.(6), ottenendo

$$\begin{aligned} \ell(\boldsymbol{\theta}; x_1, x_2) = & \sum_{j=1}^n \ln[g_1(x_{1,j})] + \sum_{j=1}^n \ln[g_2(x_{2,j})] + \\ & + \sum_{j=1}^n \ln[c(G_1(x_{1,j}), G_2(x_{2,j})); \boldsymbol{\theta}] \end{aligned} \quad (13)$$

dove anche le distribuzioni e la copula sarebbero cambiate con l'obiettivo di ottenere la verosimiglianza maggiore.

La stima delle distribuzioni delle variabili di interesse è avvenuta utilizzando il pacchetto MASS di R, ottenendo così una stima dei parametri e dei valori assunti dai principali criteri di confronto, che hanno consentito di individuare le forme distributive con verosimiglianza maggiore: secondo tale criterio, la distribuzione più adatta a descrivere gli Spread è la Gaussiana, mentre per i diversi tassi congiunturali sono state utilizzate la Gamma (per la forma tecnica capitale differito), la LogNormale (per le forme tecniche mista e vita intera) e la Weibull (per le forme tecniche capitalizzazione, rendita differita e termine fisso) (Figura3).

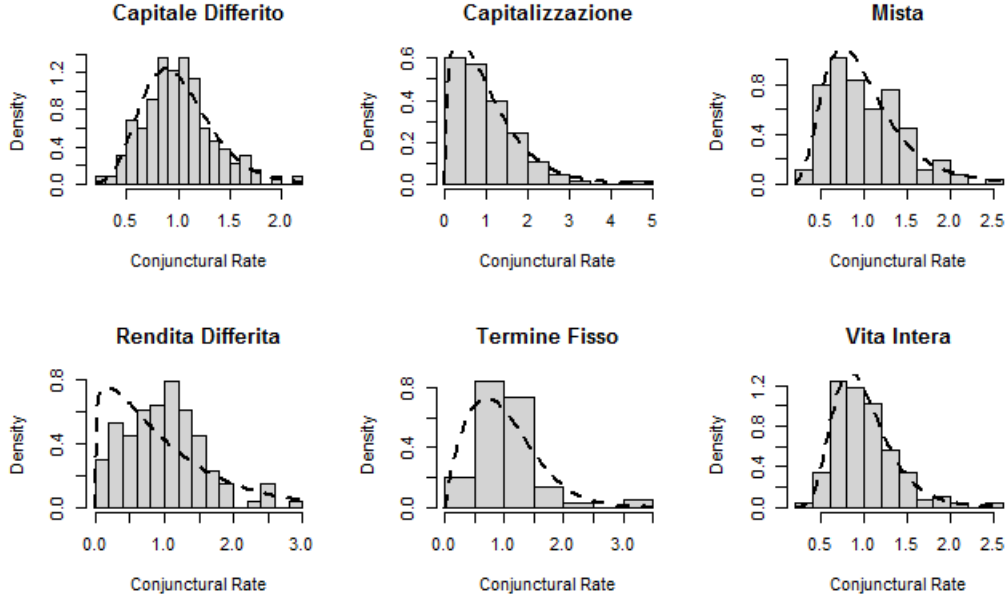


Figura 3: Distribuzioni empiriche (Istogrammi) e teoriche (linea tratteggiata) dei conjunctural rates

Sebbene nella Figura 3 non vi sia un fitting perfetto delle curve teoriche rispetto a quelle empiriche, l'andamento descritto dalle due distribuzioni è attendibile e ciò risulta sufficiente; infatti, come dimostrato in [Kim et al. 2007] mediante un approccio simulativo, l'errore commesso scegliendo distribuzioni marginali errate costituisce un problema solo nel caso in cui la scelta di queste sia profondamente sbagliata.

Una volta stimate le distribuzioni dei dati, risulta interessante svolgere un'analisi visiva della struttura di dipendenza tra le variabili di interesse. Come noto, gli scatter plots dei valori osservati sono spesso influenzati dalla scala impiegata e per tale ragione può risultare utile studiare gli pseudo-copula data \mathbf{U} – dove sono rimossi gli effetti marginali – e i valori normalizzati, ottenuti dai precedenti come

$$\begin{cases} Z_1 = \Phi^{-1}(U_1) & \sim \mathcal{N}(0, 1) \\ Z_2 = \Phi^{-1}(U_2) & \sim \mathcal{N}(0, 1) \end{cases} \quad (14)$$

così da fornire alle marginali una distribuzione gaussiana standard.

La stima della struttura di dipendenza presente all'interno dei dati rappresentati in Figura 4 è stata effettuata tramite il pacchetto VineCopula di R, che all'interno di una vasta selezione di copule bivariate ha permesso di individuare quella più idonea, fornendone poi la stima di massima verosimiglianza dei parametri.

Nella Figura 5, poi, è possibile osservare tramite i contour plot la differenza tra le distribuzioni kernel delle copule e le rispettive strutture teoriche individuate. Sebbene vi siano delle lievi difformità, nel complesso la struttura di dipendenza evidenziata dalla copula teorica è simile a quella presente all'interno dei dati pseudo-copula.

Le stime dei parametri per la calibrazione delle copule di ciascuna forma tecnica j analizzata sono riportate nella Tabella 2. Sono inoltre indicate le famiglie di Copule risultate più adatte – approfondite in Appendice A – con i rispettivi valori assunti dal Tau di Kendall osservati $\hat{\tau}_K$ e stimati τ_K , ed i valori empirici nonché i p -value del test di indipendenza tra le variabili e la log-verosimiglianza.

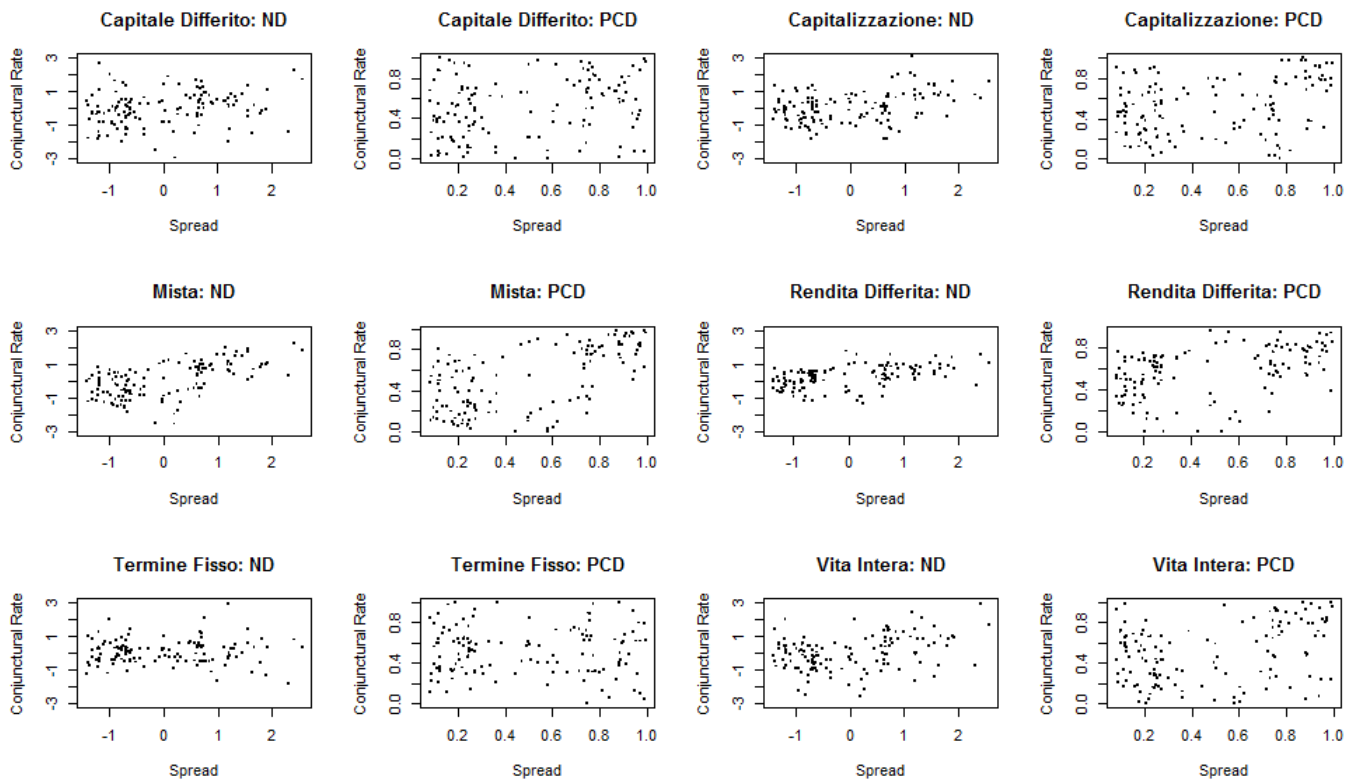


Figura 4: Scatterplots dei normalized data (ND) e pseudo-copula data (PCD)

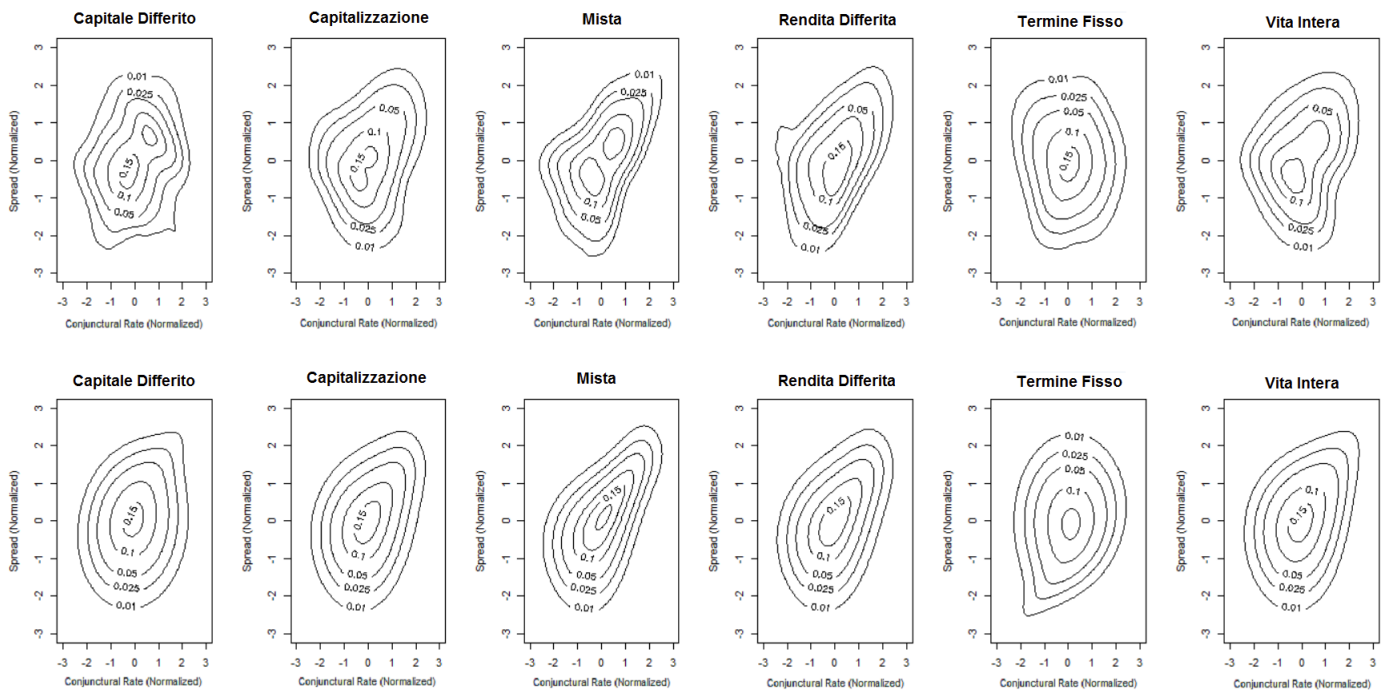


Figura 5: Confronto tra le stime kernel (riga superiore) e i contour plot teorici (riga inferiore) delle copule

Dalla Tabella 2 emerge che, per la maggior parte delle tipologie contrattuali, le copule che meglio descrivono le strutture di dipendenza sono biparametriche – la BB8 e la Tawn 1 con e

	<i>Copula</i>	θ_1	θ_2	τ_K	$\hat{\tau}_K$	<i>p-value</i>	ℓ
<i>Capitale Differito</i>	Tawn type 1	1.31	0.36	0.13	0.16	0.0073**	4.54
<i>Capitalizzazione Mista</i>	BB8	1.73	0.93	0.22	0.19	0.0014**	10.47
<i>Rendita Differita</i>	BB8	2.54	0.94	0.39	0.39	3.07e-11***	34.45
	BB8	2.16	0.89	0.29	0.34	9.27e-09***	13.36
<i>Termine Fisso</i>	Rotated Tawn type 1 180 degrees	1.59	0.18	0.11	0.02	0.7556	1.18
<i>Vita Intera</i>	Survival Clayton	0.52	0	0.21	0.18	0.0029**	11.60

Tabella 2: Principali output della fase di stima dei parametri delle copule per forma tecnica

senza rotazione per quest’ultima – e solo per i contratti a Vita Intera è risultata sufficiente una copula uniparametrica – la Survival Clayton.

Confrontando poi la stima teorica τ_K con quella empirica $\hat{\tau}_K$ si ha una coerenza nei risultati ottenuti, ad eccezione delle Termine Fisso che presentano una evidente differenza. Inoltre, tale forma tecnica – anche in virtù del valore assunto dal Tau di Kendall empirico, prossimo allo 0 – differisce dalle altre anche per il valore assunto dal *p-value*: esso è infatti troppo elevato e, perciò, non è possibile rifiutare l’ipotesi nulla di indipendenza. Una spiegazione di tale fenomeno può essere trovata ad esempio nella natura del contratto che prevede, in caso di decesso dell’assicurato nel corso della durata contrattuale, il pagamento ai beneficiari designati in polizza dal contraente del capitale assicurato solo alla scadenza del contratto e nel caso di riscatto un importo inferiore ai premi versati; è più probabile, quindi, che i contraenti siano meno sensibili alle variazioni allo scenario economico-finanziario in quanto la prestazione assicurata è prevista alla scadenza contrattuale.

3.3 Analisi di regressione con le copule

Una volta definita la struttura copula, si passa alla costruzione del modello di regressione impiegando l’Eq. (10). In questo modo è possibile ottenere la stima dell’andamento di ciascun tassi congiunturali, al variare dello spread e di cui si riporta la rappresentazione grafica nella Figura 6.

Lo sviluppo analitico delle *h*-function (cfr. Eq. (9)) è riportato in Appendice A. Per quanto concerne la soluzione dell’integrale presente nell’Eq. (10), non essendo possibile ottenerla in forma chiusa, è stata calcolata impiegando il metodo Monte Carlo su un set di 5000 simulazioni.

L’elemento che accomuna tutte le forme tecniche analizzate è l’andamento non decrescente, risultato che sottolinea l’evidente importanza degli aspetti dinamici legati al mercato all’interno del processo nel quale l’assicurato decide o meno di riscattare il contratto, e tende ad appiattirsi superata una certa soglia del valore dello spread.

Un aspetto interessante da sottolineare è l’aver un valore del tasso congiunturale stimato inferiore ad 1 anche per valori nulli dello spread: ciò può essere conseguenza, ad esempio, del fatto che anche piccole oscillazioni dei rendimenti non sempre inducono ad un aumento delle frequenze di riscatto. Tale comportamento inoltre, seppur di poco, varia da una forma tecnica

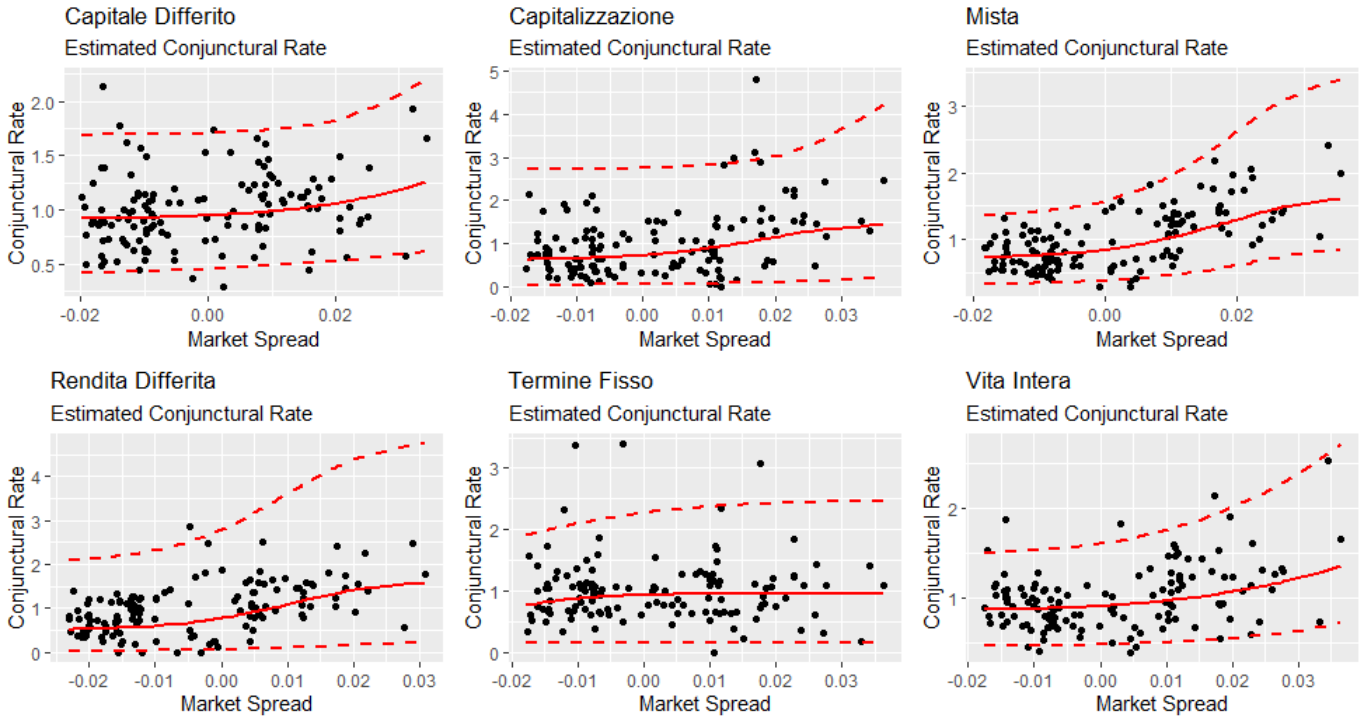


Figura 6: Stime dei conjunctural rates

all'altra e ciò potrebbe derivare dai diversi costi che vengono sostenuti per uscire dal contratto sottoscritto.

Ulteriore spunto di analisi è fornito dagli intervalli di confidenza ottenuti attraverso la copula quantile regression (cfr. Eq. (12)). Questi, come già spiegato, non sono regolari – come avviene nella maggior parte delle analisi inferenziali – ma, essendo legati alla copula sottostante e alla distribuzione della variabile stimata, consentono di ottenere un andamento specifico in funzione del livello di confidenza impiegato.

Nel caso in esame si può osservare come l'intervallo di confidenza al 95% riesca a catturare correttamente tale percentuale di osservazioni per ciascuna forma tecnica, presentando allo stesso tempo i vantaggi presenti in un approccio parametrico.

Ovviamente, tenendo conto delle ipotesi effettuate, il passaggio da dinamica del conjunctural rate a dinamica del surrender rate è ottenuto tramite la semplice moltiplicazione della costante r_j^{Base} e di conseguenza il grafico del tasso di riscatto stimato cambierebbe esclusivamente per l'uso di una nuova scala: valgono allora considerazioni analoghe a quelle svolte precedentemente.

4 Conclusioni e future estensioni

In virtù dei risultati ottenuti, si è evinto che un approccio copula-regression può essere utilizzato efficacemente in contesti assicurativi dinamici per definire la relazione presente tra il surrender rate e gli andamenti economico-finanziari ed è possibile affermare che tale strumento presenta in generale caratteristiche più che valide per inserirlo nell'ampio ventaglio di scelte laddove si voglia svolgere una analisi di regressione.

Esso fornisce una linea di continuità tra i modelli di regressione e l'approccio copula-based determinando, come già spiegato, il grande vantaggio di poter separare la definizione della distribuzione delle variabili dalla definizione della struttura di dipendenza che caratterizza i dati,

inserendosi così in modo più robusto in un contesto probabilistico e permettendo di svolgere anche stime intervallari oltre che puntuali – come invece accade nei modelli deterministici.

Ricordando che con la scelta di un modello si formula una ipotesi economica circa l'andamento di determinati fenomeni e come questi influiscono in un sistema più ampio, il caso analizzato può essere visto come una semplificazione della realtà – nella quale i benchmark di riferimento e le variabili da utilizzare sono in numero maggiore alle due considerate. In contesti più generali, estensioni possono essere effettuate utilizzando un maggior numero di variabili e seguendo un approccio Pair-Copula Construction (PCC) o Vine Copulas per definire la funzione Copula che deve poi essere utilizzata nell'effettuare la regressione.

A Le Copule utilizzate

Nella presente appendice vengono approfonditi brevemente quelli che sono gli aspetti matematici, meramente computazionali, delle copule utilizzate nello studio.

A.1 BB8

La famiglia Extended Joe BB8 è una famiglia di copule simmetriche e presenta la seguente funzione C è definita come:

$$\begin{aligned}
 C(u_1, u_2; \theta_1, \theta_2) &= \theta_2^{-1} [1 - \{1 - [1 - (1 - \theta_2)^{\theta_1}]^{-1} \\
 &\quad \cdot [1 - (1 - \theta_2 u_1)^{\theta_1}] [1 - (1 - \theta_2 u_2)^{\theta_1}]\}^{\frac{1}{\theta_1}}] \\
 &= \frac{1 - \theta_1 \sqrt[{\theta_1}]{1 - \frac{[1 - (1 - \theta_2 u_1)^{\theta_1}][1 - (1 - \theta_2 u_2)^{\theta_1}]}{1 - (1 - \theta_2)^{\theta_1}}}}{\theta_2}
 \end{aligned} \tag{15}$$

e h-function, ottenuta derivando la funzione copula rispetto a u_2

$$\begin{aligned}
 h_{1|2}(u_1|u_2) &= C_{u_2}(u_1) = \\
 &= \frac{(1 - \theta_2 u_2)^{\theta_1 - 1} [1 - (1 - \theta_2 u_1)^{\theta_1}]}{1 - (1 - \theta_2)^{\theta_1}} \left[1 - \frac{[1 - (1 - \theta_2 u_2)^{\theta_1}] [1 - (1 - \theta_2 u_1)^{\theta_1}]}{1 - (1 - \theta_2)^{\theta_1}} \right]^{\frac{1}{\theta_1} - 1}
 \end{aligned} \tag{16}$$

Dimostrazione A.1. Per non appesantire i calcoli, limitatamente alla presente dimostrazione si adotta la seguente notazione

$$\begin{cases}
 y = [1 - \theta_2 u_2]^{\theta_1} \\
 d = \frac{[1 - (1 - \theta_2 u_1)^{\theta_1}]}{1 - (1 - \theta_2)^{\theta_1}} \\
 z = \frac{[1 - (1 - \theta_2 u_1)^{\theta_1}] [1 - (1 - \theta_2 u_2)^{\theta_1}]}{1 - (1 - \theta_2)^{\theta_1}} = d(1 - y)
 \end{cases}$$

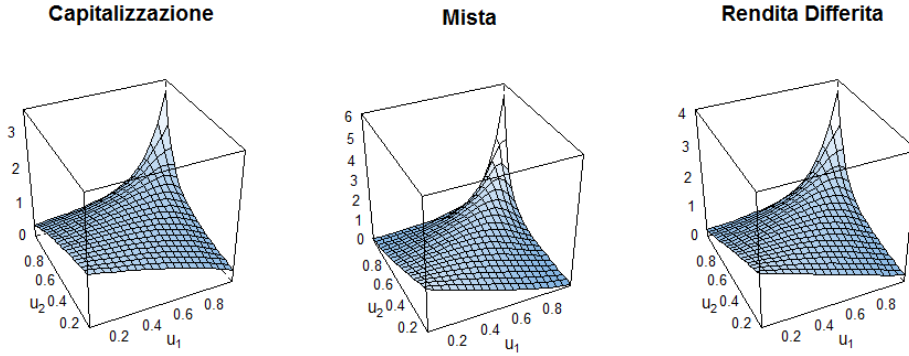
$$\begin{aligned}
 h_{1|2}(u_1|u_2) &= \frac{\partial}{\partial u_2} C(u_1, u_2) \\
 &= -\frac{1}{\theta_2} \frac{\partial}{\partial u_2} \theta_1 \sqrt[{\theta_1}]{1 - \frac{[1 - (1 - \theta_2 u_1)^{\theta_1}] [1 - (1 - \theta_2 u_2)^{\theta_1}]}{1 - (1 - \theta_2)^{\theta_1}}} \\
 &= -\frac{1}{\theta_2} \frac{\partial}{\partial u_2} \theta_1 \sqrt[{\theta_1}]{1 - z} = -\frac{1}{\theta_2} \left[\frac{1}{\theta_1} (1 - z)^{\frac{1}{\theta_1} - 1} \frac{\partial}{\partial u_2} [1 - z] \right] \\
 &= \frac{1}{\theta_2} \left[\frac{1}{\theta_1} (1 - z)^{\frac{1}{\theta_1} - 1} \frac{\partial}{\partial u_2} [d - dy] \right] \\
 &= -\frac{d}{\theta_2} \left[\frac{1}{\theta_1} (1 - z)^{\frac{1}{\theta_1} - 1} \frac{\partial}{\partial u_2} [1 - \theta_2 u_2]^{\theta_1} \right] \\
 &= -\frac{d}{\theta_2} \left[\frac{1}{\theta_1} (1 - z)^{\frac{1}{\theta_1} - 1} \theta_1 [1 - \theta_2 u_2]^{\theta_1 - 1} (-\theta_2) \right] \\
 &= d \left[(1 - z)^{\frac{1}{\theta_1} - 1} [1 - \theta_2 u_2]^{\theta_1 - 1} \right]
 \end{aligned}$$

Sostituendo i valori si ottiene la Eq. (16)

Derivando la funzione copula rispetto a u_1 e u_2 , con procedimento analogo al caso precedente, si ottiene la densità

$$\begin{aligned}
c(u_1, u_2; \theta_1, \theta_2) &= \frac{\theta_2 \theta_1 (1 - \theta_2 u_2)^{\theta_1 - 1} (1 - \theta_2 u_1)^{\theta_1 - 1}}{1 - (1 - \theta_2)^{\theta_1}} \left[1 - \frac{[1 - (1 - \theta_2 u_2)^{\theta_1}][1 - (1 - \theta_2 u_1)^{\theta_1}]}{1 - (1 - \theta_2)^{\theta_1}} \right]^{\frac{1}{\theta_1} - 1} \\
&\quad - \frac{1}{[1 - (1 - \theta_2)^{\theta_1}]^2} \theta_2 \left(\frac{1}{\theta_1} - 1 \right) \theta_1 (1 - \theta_2 u_2)^{\theta_1 - 1} [1 - (1 - \theta_2 u_2)^{\theta_1}] (1 - \theta_2 u_1)^{\theta_1 - 1} [1 - (1 - \theta_2 u_1)^{\theta_1}] \\
&\quad \cdot \left[1 - \frac{[1 - (1 - \theta_2 u_2)^{\theta_1}][1 - (1 - \theta_2 u_1)^{\theta_1}]}{1 - (1 - \theta_2)^{\theta_1}} \right]^{\frac{1}{\theta_1} - 2}
\end{aligned} \tag{17}$$

che, usando i parametri precedentemente stimati, permette di ottenere le seguenti rappresentazioni per le superfici di densità copula delle tipologie contrattuali con funzione copula BB8



A.2 Tawn Type 1 e Tawn Type 1 Rotated 180°

La famiglia Tawn è suddivisa in due tipologie sulla base dei parametri δ_1 e δ_2 , i quali possono essere interpretati come parametri di asimmetria, infatti fissando uno dei due uguale ad 1 questi consentono di definire le due forme di Copula Tawn biparametriche (Tawn type 1 e Tawn type 2), che si differenziano per la direzione dell'asimmetria – $\delta_1 = 1$ per una asimmetria a sinistra e $\delta_2 = 1$ per una asimmetria a destra.

La funzione Copula è definita come

$$\begin{aligned}
C(u_1, u_2; \theta, \delta_1, \delta_2) &= (1 - \delta_1) \frac{\log(u_2)}{\log(u_1 u_2)} - (1 - \delta_2) \left(1 - \frac{\log(u_2)}{\log(u_1 u_2)} \right) \\
&\quad + \left[\left(\delta_1 \left(1 - \frac{\log(u_2)}{\log(u_1 u_2)} \right) \right)^\theta + \left(\delta_2 \frac{\log(u_2)}{\log(u_1 u_2)} \right)^\theta \right]^{\frac{1}{\theta}}
\end{aligned} \tag{18}$$

con $\delta_1, \delta_2 \in [0, 1]$ e $\theta \in [1, +\infty)$.

La cui h-function è

$$\begin{aligned}
h_{1|2}(u_1|u_2) &= C_{u_2}(u_1) = \\
&= \frac{1}{\theta} \left\{ \theta \delta_1 \left[\frac{\log(u_2)}{u_2 \log^2(u_2 u_1)} - \frac{1}{u_2 \log(u_2 u_1)} \right] \left[\delta_1 \left(1 - \frac{\log(u_2)}{\log(u_2 u_1)} \right) \right]^{\theta-1} + \theta \left[\frac{\delta_2}{u_2 \log(u_2 u_1)} \right. \right. \\
&\quad \left. \left. - \frac{\delta_2 \log(u_2)}{u_2 \log^2(u_2 u_1)} \right] \left[\frac{\delta_2 \log(u_2)}{\log(u_2 u_1)} \right]^{\theta-1} \right\} \left\{ \left[\delta_1 \left(1 - \frac{\log(u_2)}{\log(u_2 u_1)} \right) \right]^\theta + \left[\frac{\delta_2 \log(u_2)}{\log(u_2 u_1)} \right]^\theta \right\}^{\frac{1}{\theta}-1} \\
&\quad - \frac{(1-\delta_1)\log(u_2)}{u_2 \log^2(u_2 u_1)} + \frac{1-\delta_1}{u_2 \log(u_2 u_1)} - (1-\delta_2) \left[\frac{\log(u_2)}{u_2 \log^2(u_2 u_1)} - \frac{1}{u_2 \log(u_2 u_1)} \right]
\end{aligned} \tag{19}$$

Dimostrazione A.2. Come nella precedente dimostrazione, per non appesantire la notazione si pone

$$\begin{cases} z = \frac{\log(u_2)}{\log(u_1 u_2)} \\ z' = \frac{\partial}{\partial u_2} z = \frac{\frac{1}{u_2} [\log(u_1 u_2) - \log(u_2)]}{[\log(u_1 u_2)]^2} \end{cases}$$

$$\begin{aligned}
h_{1|2}(u_1|u_2) &= \frac{\partial}{\partial u_2} C(u_1, u_2) \\
&= (1-\delta_1) \frac{\partial}{\partial u_2} \frac{\log(u_2)}{\log(u_1 u_2)} - (1-\delta_2) \frac{\partial}{\partial u_2} \left(1 - \frac{\log(u_2)}{\log(u_1 u_2)} \right) \\
&\quad + \frac{\partial}{\partial u_2} \left[\left(\delta_1 \left(1 - \frac{\log(u_2)}{\log(u_1 u_2)} \right) \right)^\theta + \left(\delta_2 \frac{\log(u_2)}{\log(u_1 u_2)} \right)^\theta \right]^{\frac{1}{\theta}} \\
&= (1-\delta_1) \left[\frac{\frac{1}{u_2} \log(u_1 u_2) - \log(u_2) \frac{u_1}{u_1 u_2}}{[\log(u_1 u_2)]^2} \right] - (1-\delta_2) \left[-\frac{\frac{1}{u_2} \log(u_1 u_2) - \log(u_2) \frac{u_1}{u_1 u_2}}{[\log(u_1 u_2)]^2} \right] \\
&\quad + \frac{\partial}{\partial u_2} \{ [\delta_1 (1-z)]^\theta + [\delta_2 z]^\theta \}^{\frac{1}{\theta}} \\
&= (1-\delta_1) z' - (1-\delta_2) (-z') + \frac{1}{\theta} \{ [\delta_1 (1-z)]^\theta + [\delta_2 z]^\theta \}^{\frac{1}{\theta}-1} \theta \{ [\delta_1 (1-z)]^{\theta-1} \delta_1 (-z') + [\delta_2 z]^{\theta-1} \delta_2 z' \}
\end{aligned}$$

Semplificando e sostituendo i valori si ottiene la Eq. (19)

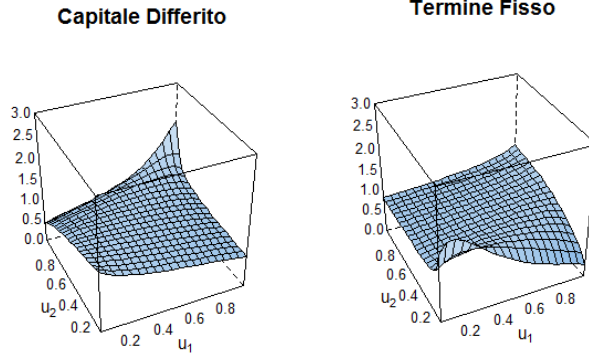
Per la densità della copula si ha

$$\begin{aligned}
c(u_1, u_2; \theta, \delta_1, \delta_2) &= \frac{d^2}{du_1 du_2} C(u_1, u_2; \theta, \delta_1, \delta_2) \\
&= \frac{1}{\theta} \left\{ \frac{(\theta - 1)\theta\delta_1^2 \log(u_2) \left[\frac{\log(u_2)}{u_2 \log^2(u_2 u_1)} - \frac{1}{u_2 \log(u_2 u_1)} \right]}{y \log^2(u_2 u_1)} \right. \\
&\quad \cdot \left[\delta_1 \left(1 - \frac{\log(u_2)}{\log(u_2 u_1)} \right) \right]^{\theta-2} - \frac{2\theta\delta_1 \log(u_2) \left[\delta_1 \left(1 - \frac{\log(u_2)}{\log(u_2 u_1)} \right) \right]^{\theta-1}}{u_2 u_1 \log^3(u_2 u_1)} \\
&\quad + \frac{\theta\delta_1 \left[\delta_1 \left(1 - \frac{\log(u_2)}{\log(u_2 u_1)} \right) \right]^{\theta-1}}{u_2 u_1 \log^2(u_2 u_1)} + \frac{2\theta\delta_2 \log(u_2) \left[\frac{\delta_2 \log(u_2)}{\log(u_2 u_1)} \right]^{\theta-1}}{u_2 u_1 \log^3(u_2 u_1)} \\
&\quad - \frac{(\theta - 1)\theta\delta_2 \log(u_2) \left[\frac{\delta_2}{u_2 \log(u_2 u_1)} - \frac{\delta_2 \log(u_2)}{u_2 \log^2(u_2 u_1)} \right] \left[\frac{\delta_2 \log(u_2)}{\log(u_2 u_1)} \right]^{\theta-2}}{u_1 \log^2(u_2 u_1)} \\
&\quad - \frac{\theta\delta_2 \left[\frac{\delta_2 \log(u_2)}{\log(u_2 u_1)} \right]^{\theta-1}}{u_2 u_1 \log^2(u_2 u_1)} \left\{ \left[\delta_1 \left(1 - \frac{\log(u_2)}{\log(u_2 u_1)} \right) \right]^\theta \right. \\
&\quad \left. + \left[\frac{\delta_2 \log(u_2)}{\log(u_2 u_1)} \right]^\theta \right\}^{\frac{1}{\theta}-1} + \frac{1}{\theta} \left(\frac{1}{\theta} - 1 \right) \left\{ \theta\delta_1 \left[\frac{\log(u_2)}{u_2 \log^2(u_2 u_1)} \right. \right. \\
&\quad \left. \left. - \frac{1}{u_2 \log(u_2 u_1)} \right] \left[\delta_1 \left(1 - \frac{\log(u_2)}{\log(u_2 u_1)} \right) \right]^{\theta-1} + \theta \left[\frac{\delta_2}{u_2 \log(u_2 u_1)} - \frac{\delta_2 \log(u_2)}{u_2 \log^2(u_2 u_1)} \right] \left[\frac{\delta_2 \log(u_2)}{\log(u_2 u_1)} \right]^{\theta-1} \right\} \\
&\quad \cdot \left\{ \frac{\theta\delta_1 \log(u_2) \left[\delta_1 \left(1 - \frac{\log(u_2)}{\log(u_2 u_1)} \right) \right]^{\theta-1}}{u_1 \log^2(u_2 u_1)} - \frac{\theta\delta_2 \log(u_2) \left[\frac{\delta_2 \log(u_2)}{\log(u_2 u_1)} \right]^{\theta-1}}{u_1 \log^2(u_2 u_1)} \right\} \\
&\quad \cdot \left\{ \left[\delta_1 \left(1 - \frac{\log(u_2)}{\log(u_2 u_1)} \right) \right]^\theta + \left[\frac{\delta_2 \log(u_2)}{\log(u_2 u_1)} \right]^\theta \right\}^{\frac{1}{\theta}-2} + \frac{2(1 - \delta_1) \log(u_2)}{u_2 u_1 \log^3(u_2 u_1)} \\
&\quad \left. - \frac{1 - \delta_1}{u_2 u_1 \log^2(u_2 u_1)} + \frac{2(1 - \delta_2) \log(u_2)}{u_2 u_1 \log^3(u_2 u_1)} - \frac{1 - \delta_2}{u_2 u_1 \log^2(u_2 u_1)} \right\}
\end{aligned} \tag{20}$$

Data la funzione di densità copula, la rotazione consente di estendere il possibile range di valori per la misura di dipendenza in quelle copule che presentano una struttura di dipendenza asimmetrica. In particolare, la rotazione di 180° si ottiene come segue:

$$c_{Tawn}^{180}(u_1, u_2; \theta, \delta_1, \delta_2) = c_{Tawn}(1 - u_1, 1 - u_2; \theta, \delta_1, \delta_2) \tag{21}$$

Usando i parametri precedentemente stimati, si ottengono le seguenti rappresentazioni per le superfici di densità copula delle tipologie contrattuali con funzione copula Tawn Type 1 (rotated e non rotated)



A.3 Survival Clayton

Un modo alternativo per definire le copule ruotate è chiamarle "Survival", ma da un punto di vista pratico si segue la stessa procedura definita in precedenza per la Tawn 180°.

La Copula Clayton presenta la seguente funzione Copula:

$$\begin{aligned}
 C(u_1, u_2; \theta) &= \max[u_1^{-\theta} + u_2^{-\theta} - 1, 0]^{-\frac{1}{\theta}} \\
 &= \max\left[\frac{1}{u_1^\theta} + \frac{1}{u_2^\theta} - 1, 0\right]^{-\frac{1}{\theta}} \\
 &= \sqrt[\theta]{\frac{1}{\max\left[\frac{1}{u_1^\theta} + \frac{1}{u_2^\theta} - 1, 0\right]}}
 \end{aligned} \tag{22}$$

con $\theta \in [-1, \infty) \setminus \{0\}$

Poiché nel caso in esame si ha $\theta = 0.52$ e in virtù del fatto che $u_1, u_2 \in [0, 1]$

$$\max\left[\frac{1}{u_1^\theta} + \frac{1}{u_2^\theta} - 1, 0\right] = \frac{1}{u_1^\theta} + \frac{1}{u_2^\theta} - 1$$

Per tale motivo, nella fattispecie, si ha

$$C(u_1, u_2; \theta) = [u_1^{-\theta} + u_2^{-\theta} - 1]^{-\frac{1}{\theta}} \tag{23}$$

e h-function

$$\begin{aligned}
 h_{1|2}(u_1|u_2) &= C_{u_2}(u_1) = \frac{\partial}{\partial u_2} C(u_1, u_2; \theta) = \\
 &= u_2^{-(\theta+1)} [u_1^{-\theta} + u_2^{-\theta} - 1]^{-\frac{1}{\theta}-1}
 \end{aligned} \tag{24}$$

Dimostrazione A.3.

$$\begin{aligned}
 h_{1|2}(u_1|u_2) &= \frac{\partial}{\partial u_2} C(u_1, u_2) \\
 &= -\frac{1}{\theta} [u_1^{-\theta} + u_2^{-\theta} - 1]^{-\frac{1}{\theta}-1} [-\theta u_2^{-\theta-1}]
 \end{aligned}$$

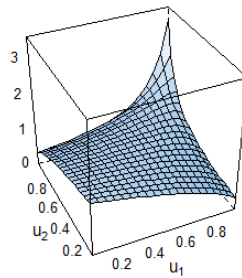
Da cui segue la Eq. (24)

Infine, la copula density

$$\begin{aligned}
c(u_1, u_2; \theta) &= \frac{d^2}{du_1 du_2} C(u_1, u_2; \theta) \\
&= \left(1 + \frac{1}{\theta}\right) \theta (u_1 u_2)^{-(\theta+1)} [u_1^{-\theta} + u_2^{-\theta} - 1]^{-\frac{1}{\theta}-2} \\
&= (\theta + 1) (u_1 u_2)^{-(\theta+1)} [u_1^{-\theta} + u_2^{-\theta} - 1]^{-\left(\frac{1}{\theta}+2\right)}
\end{aligned} \tag{25}$$

Adottando i parametri ottenuti in fase di stima, la rappresentazione della densità della Survival Clayton è

Vita Intera



Riferimenti bibliografici

- [Baione et al. 2018] Baione, F., Biancalana, D., Angelis, P. D., & Granito, I. (2018). Dynamic policyholder behavior and surrender option evaluation for life insurance. In *Mathematical and Statistical Methods for Actuarial Sciences and Finance* (pp. 81-85). Springer, Cham.
- [Baione et al. 2021] Baione, F., Biancalana, D., & De Angelis, P. (2021). An application of Sigmoid and Double-Sigmoid functions for dynamic policyholder behaviour. *Decisions in Economics and Finance*, 44(1), 5-22.
- [Cerchiara et al. 2008] Cerchiara, R. R., Edwards, M., & Gambini, A. (2008, October). Generalized linear models in life insurance: decrements and risk factor analysis under Solvency II. In *18th international AFIR colloquium*.
- [CFO] http://www.cfoforum.eu/downloads/CFO-Forum_EEV_Principles_and_Guidance_April_2016.pdf
- [Czado 2019] Czado, C. (2019). Analyzing dependent data with vine copulas. *Lecture Notes in Statistics*, Springer.
- [Eling & Kochanski 2013] Eling, M., & Kochanski, M. (2013). Research on lapse in life insurance: what has been done and what needs to be done?. *The Journal of Risk Finance*.
- [IVASS] https://www.ivass.it/normativa/internazionale/internazionale-ue/direttive/2009-138-ce/direttiva_solvency_II_aggiornata.pdf
- [Kim 2005] Kim, C. (2005). Modeling surrender and lapse rates with economic variables. *North American Actuarial Journal*, 9(4), 56-70.
- [Kim et al. 2007] Kim, G., Silvapulle, M. J., & Silvapulle, P. (2007). Comparison of semiparametric and parametric methods for estimating copulas. *Computational Statistics & Data Analysis*, 51(6), 2836-2850.
- [Kraus & Czado 2017] Kraus, D., & Czado, C. (2017). D-vine copula based quantile regression. *Computational Statistics & Data Analysis*, 110, 1-18.
- [MacKay et al. 2017] MacKay, A., Augustyniak, M., Bernard, C., & Hardy, M. R. (2017). Risk management of policyholder behavior in equity-linked life insurance. *Journal of Risk and Insurance*, 84(2), 661-690.
- [Milhaud, Loisel & Maume-Deschamps 2011] Milhaud, X., Loisel, S., & Maume-Deschamps, V. (2011). Surrender triggers in life insurance: what main features affect the surrender behavior in a classical economic context?. *Bulletin Français d'Actuariat*, 11(22), 5-48.
- [Milhaud & Dutang 2018] Milhaud, X., & Dutang, C. (2018). Lapse tables for lapse risk management in insurance: a competing risk approach. *European Actuarial Journal*, 8(1), 97-126.
- [Nolte & Schneider 2017] Nolte, S., & Schneider, J. C. (2017). Don't lapse into temptation: A behavioral explanation for policy surrender. *Journal of Banking & Finance*, 79, 12-27.
- [Pinquet, Guillén & Ayuso 2011] Pinquet, J., Guillén, M., & Ayuso, M. (2011). Commitment and lapse behavior in long-term insurance: a case study. *Journal of Risk and Insurance*, 78(4), 983-1002.

[Sungur 2005] Sungur, E. A. (2005). Some observations on copula regression functions. *Communications in Statistics—Theory and Methods*, 34(9-10), 1967-1978.