



Università degli Studi di Trieste

DIPARTIMENTO DI SCIENZE ECONOMICHE, AZIENDALI, MATEMATICHE E STATISTICHE

“BRUNO DE FINETTI”

CORSO DI LAUREA MAGISTRALE IN SCIENZE STATISTICHE E ATTUARIALI

TESI DI LAUREA MAGISTRALE

Modelli proxy per la determinazione dei requisiti di capitale secondo Solvency II

Laureando:

Johannes Jakob Schoenenwald

Relatore:

Prof. Ermanno Pitacco

Correlatori:

Prof. Francesco Pauli

Dott. Alberto Frascarolo

Indice

Introduzione	1
1 Modelli per la determinazione del requisito di capitale	5
1.1 Solvency II e la standard formula	6
1.1.1 La direttiva Solvency II	6
1.1.2 Il Solvency Capital Requirement	11
1.1.3 La Standard Formula	13
1.1.4 I limiti della Standard Formula	19
1.1.5 Verso un modello interno	25
1.2 Approccio stocastico al calcolo dell'SCR	27
1.2.1 Simulazioni stocastiche annidate	27
1.2.2 L'impostazione formale del modello stocastico	31
1.2.3 Ottimizzazione del modello stocastico	35
1.3 Modelli proxy	46
1.3.1 Portafogli replicanti	47
1.3.2 Curve fitting	51
1.3.3 Least-Squares Monte Carlo	56
1.4 I modelli a confronto	63
2 Il modello Least-Squares Monte Carlo	67
2.1 Costruzione di un modello LSMC	68
2.1.1 L'impostazione del metodo	68
2.1.2 Le basi teoriche dell'approccio	73
2.1.3 Analisi di convergenza dell'algoritmo	76
2.1.4 Determinazione della struttura del polinomio	81

2.1.5	La validazione delle funzioni proxy	87
2.2	Estensioni della forma strutturale del modello	90
2.2.1	Polinomi ortogonali	90
2.2.2	Modelli additivi e spline	92
2.2.3	Altre proposte in letteratura	95
2.3	Implementazione del modello LSMC	99
2.3.1	I dati e il contesto lavorativo	99
2.3.2	Applicazione a portafogli di attività e passività	105
2.3.3	Applicazione ai PVFP e calcolo dell'SCR	139
2.4	Conclusioni	146
3	Un modello proxy basato su metodi di machine learning	149
3.1	L'apprendimento con modelli ad albero	150
3.1.1	Alberi di regressione	150
3.1.2	Foreste casuali	155
3.1.3	Tecniche di boosting	157
3.2	Applicazioni al contesto assicurativo	163
3.2.1	Metodi basati su alberi di regressione	163
3.2.2	Metodi basati su foreste casuali	167
3.2.3	Metodi basati sul gradient boosting	171
	Considerazioni finali	179
	A Sulla valutazione neutrale al rischio	183
	B Funzioni R per l'implementazione efficiente del modello LSMC	189
	Bibliografia	199

Introduzione

L'entrata in vigore il 1 gennaio 2016 della direttiva 2009/138/EC (Solvency II) ha accentuato l'attenzione per un sistema efficiente di risk management all'interno delle imprese di assicurazione e riassicurazione operanti nell'Unione Europea e ha evidenziato la necessità di un calcolo più adeguato dei requisiti di capitale di cui le imprese devono disporre al fine di far fronte ai propri impegni.

Più in dettaglio, il capitale da allocare a copertura dei rischi assunti (Solvency Capital Requirement o brevemente SCR) viene visto come l'importo necessario a coprire con un livello di confidenza di almeno 99.5% ogni perdita imprevista nell'orizzonte temporale di un anno, basandosi su una valutazione market-consistent di attività e passività.

Data la complessità insita in un calcolo corretto e affidabile dell'SCR secondo tale definizione, per molte imprese l'implementazione di modelli adeguati è stata una sfida importante. La Commissione Europea, con il supporto dell'organismo di vigilanza sul mercato assicurativo europeo (EIOPA) e basandosi su diversi studi di impatto quantitativo (QIS), ha inserito nella direttiva la possibilità per le imprese di avvalersi di una formula semplificata di calcolo dell'SCR (standard formula). Tuttavia, a causa della potenziale inadeguatezza della standard formula nel descrivere le specificità dei rischi assunti da una particolare compagnia, il regolatore stesso ha reso obbligatorio lo sviluppo di sistemi interni di accertamento dei rischi a proprio carico (own risk and solvency assessment) e spinge affinché le imprese siano in grado di valutare il proprio SCR tramite un modello interno.

Nel contesto delle tecniche più recenti proposte dalla letteratura per lo sviluppo di un modello interno di calcolo del requisito di capitale di solvibilità in un'impresa di assicurazione, riveste particolare importanza il metodo Least-Squares Monte

Carlo (LSMC), che, basandosi sull'idea sottostante all'omonimo modello proposto da Longstaff e Schwartz nel 2001 per la valutazione delle opzioni americane, permette di calcolare i requisiti di capitale seguendo un approccio stocastico e viene attualmente considerato come best practice nel contesto assicurativo.

Uno degli aspetti caratteristici del modello LSMC è l'utilizzo di una funzione (proxy function) utilizzata per descrivere in maniera semplificata il valore di attivi e passivi sulla base dei fattori di rischio che li influenzano. Per quanto l'impianto teorico del modello ammetta qualsiasi forma funzionale, nella quasi totalità delle implementazioni pratiche viene utilizzata una funzione polinomiale. Tale scelta, che da un lato garantisce una sostanziale semplicità nell'implementazione del modello e nell'interpretazione dei risultati, rispetto a tecniche statistiche più avanzate è però potenzialmente carente da un punto di vista delle capacità di adattamento e di previsione. Risulta quindi opportuna la ricerca di soluzioni che permettano di ottenere un fit migliore e ridurre l'errore commesso a causa dell'approssimazione introdotta dalla proxy.

In particolare, negli ultimi anni stanno emergendo diverse ricerche sull'applicazione di tecniche di machine learning nelle scienze attuariali e finanziarie. Si ritiene infatti che rispetto ai modelli tradizionali le tecniche di machine learning possano condurre, spesso, a risultati più precisi, in particolare in relazione al problema di previsione. Tuttavia, all'aumentare della complessità del metodo utilizzato, diminuisce anche la semplicità nell'interpretazione dei risultati, fino ad arrivare a tecniche c.d. "black box" nelle quali il funzionamento interno del sistema nel passaggio da input a output risulta difficilmente interpretabile. Un metodo puramente black box risulta non propriamente adatto al contesto lavorativo in cui ci poniamo, in quanto solo difficilmente tecniche dal funzionamento poco trasparente verranno approvate dagli organismi di vigilanza.

In questa tesi si cercherà di analizzare alcune possibili estensioni del modello LSMC, con l'obiettivo di abbandonare in maniera graduale l'idea di una struttura basata su funzioni polinomiali al fine di sostituirla con metodi che potenzialmente presentano maggiore accuratezza previsiva, pur tenendo in considerazione che in un mondo regolamentato quale quello assicurativo la possibilità di interpretare e spiegare i risultati ottenuti è fondamentale. In particolare, proporremo un mo-

dello stocastico di stampo Least-Squares Monte Carlo che fa utilizzo di strutture basate sugli alberi di regressione per ottenere il valore delle attività e passività assicurative a partire dai valori dei fattori di rischio. Oltre a fornire un quadro teorico relativo ai modelli considerati, ci si concentrerà sulla loro implementazione in un contesto pratico e sul confronto empirico dei risultati con quelli ottenibili mediante il modello LSMC nella sua forma più tradizionale. In appendice saranno delineate le basi di finanza matematica necessarie ad impostare formalmente il modello stocastico analizzato e saranno forniti degli esempi di funzioni in linguaggio R che permettono di implementare in pratica il modello LSMC trattato. Il percorso seguito si articola in tre capitoli, organizzati nel modo seguente:

Il primo capitolo mira a fornire il contesto nel quale si colloca il lavoro di tesi e a descrivere i principali modelli per la determinazione del requisito di capitale di solvibilità: in particolare, si andrà ad analizzare la standard formula proposta nella direttiva Solvency II e i suoi limiti, il modello teorico di simulazioni stocastiche annidate da utilizzare per lo sviluppo di un modello interno, infine i modelli proxy utilizzati nella pratica assicurativa.

Il secondo capitolo studia con maggiore dettaglio il modello Least-Squares Monte Carlo, ponendo l'attenzione in particolare sugli aspetti più critici alla sua base. Vengono ripresi i risultati teorici in letteratura relativamente all'impostazione formale e alla convergenza del modello e illustrate metodologie per lo sviluppo della funzione proxy. Successivamente sono presentate alcune estensioni analitiche del modello proposte in letteratura: si punta a mantenere ancora il legame funzionale in forma chiusa tra il valore del portafoglio e il valore assunto dai fattori di rischio, sviluppando tuttavia la forma funzionale al fine di ottenere un migliore adattamento ai dati e/o una maggiore capacità previsiva. Infine, il modello analizzato sarà implementato su portafogli di attività e passività di una compagnia di assicurazione vita.

Il terzo capitolo, in ultimo, mira a fornire un breve quadro relativo alle tecniche di machine learning incentrate sui modelli ad albero e a proporre un'applicazione al metodo stocastico per il calcolo del capitale, come alternativa alla regressione Least-Squares tipica del modello LSMC. I risultati saranno poi confrontati con quelli ottenuti con il metodo LSMC tradizionale nel capitolo precedente.

Capitolo 1

Modelli per la determinazione del requisito di capitale

Nel corso di questo capitolo si cerca di fornire una rassegna dei possibili metodi proposti dalla letteratura per la determinazione dei requisiti di capitale di solvibilità richiesti alle imprese operanti nel settore assicurativo dalla direttiva comunitaria Solvency II.

Dapprima si richiama brevemente il quadro normativo fornito dalla direttiva, in particolare analizzando le norme sottostanti al calcolo dei requisiti di capitale e la formula semplificata proposta. Si passa poi allo studio delle possibili strutture alla base di un modello interno, rimarcando la necessità di ottenere una distribuzione di probabilità della perdita, l'impossibilità computazionale di adottare pienamente un approccio stocastico annidato, e infine i modelli proxy comunemente usati, che, basandosi su un calcolo semplificato (tramite una proxy, appunto) del valore del portafoglio in ognuno degli scenari stocastici, mirano a ridurre lo sforzo computazionale delle simulazioni stocastiche annidate, permettendo comunque di arrivare ad una distribuzione di probabilità per la perdita.

Si cercherà di analizzare vantaggi e svantaggi dei diversi metodi proposti, con particolare riguardo alla loro implementazione pratica.

1.1 Solvency II e la standard formula

Con l'avvento del regime normativo Solvency II si è assistito all'inizio di una nuova era per il risk management assicurativo in Europa. La direttiva infatti ha rivoluzionato la vigilanza sul settore assicurativo, i cui processi sono stati rivisti e omogeneizzati a livello europeo, e ha imposto un approccio orientato al rischio (sistema risk-based) per il quale le imprese operanti nel settore devono essere in grado di gestire efficacemente tutti i rischi a cui sono esposte, tenendo in considerazione in particolare anche le correlazioni tra le diverse tipologie di rischio¹.

Di seguito saranno richiamati gli aspetti cruciali del progetto Solvency II, con particolare riguardo alle norme di calcolo del capitale di solvibilità.

1.1.1 La direttiva Solvency II

Il progetto Solvency II, entrato definitivamente in vigore con il recepimento della direttiva comunitaria 2009/138/EC il 1 gennaio 2016, è in realtà frutto di un lungo processo di riforma delle normative sulla vigilanza sul settore assicurativo, con origini risalenti agli anni '90. La prima fase - usualmente denominata Solvency I - si è conclusa nel 2002 con l'emanazione delle Direttive 2002/13/EC, 2002/83/EC di fatto aggiornando la normativa europea precedente relativa agli anni '70 (Direttive 73/239/CE e 79/267/EEC) senza mutarne però la logica di fondo: il margine di solvibilità, idea già introdotta negli anni '70, veniva ancora calcolato come percentuale delle riserve matematiche nei rami vita e dei premi lordi annui, o della media annua delle competenze sinistri nei rami danni, e il sistema in generale si fondava su limitazioni nell'esposizione ai rischi (sistema rule-based) anziché studiare la natura e l'impatto dei rischi stessi.

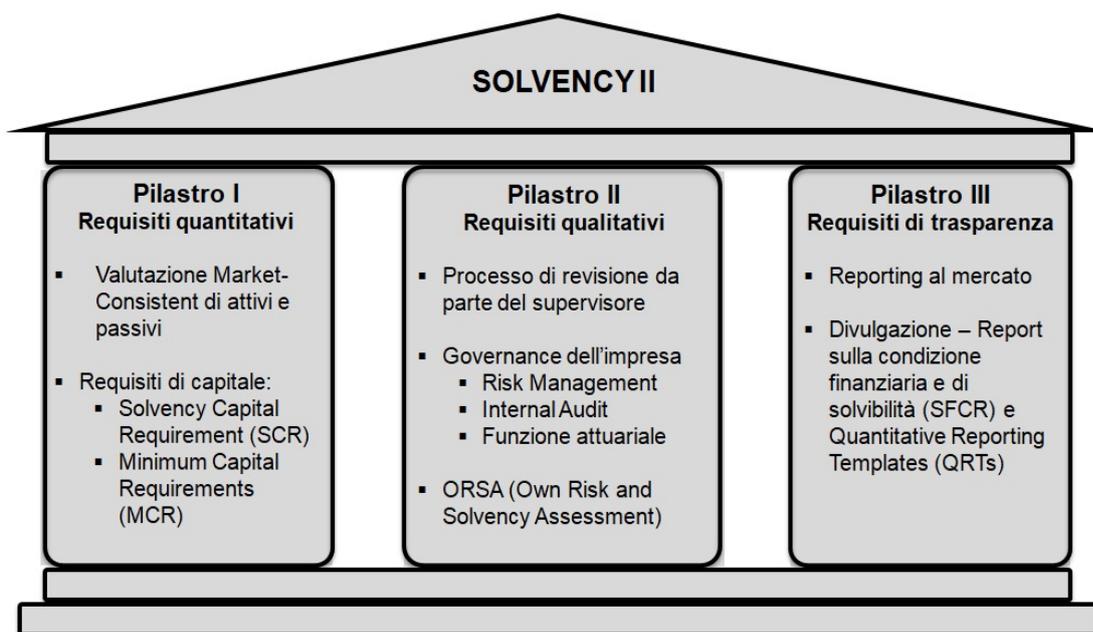
Alla riforma Solvency I è seguita una seconda fase più radicale - Solvency II - con l'obiettivo di riformare non soltanto le metodologie di calcolo del requisito patrimoniale e delle riserve tecniche, ma l'intero sistema di vigilanza, al fine di

¹Il rischio da un punto di vista matematico è definito come la variabile aleatoria raffigurante, in questo caso, la variazione di capitale associata ad un evento futuro e incerto (causa di rischio). Per semplicità di linguaggio, in questo lavoro con il termine rischio ci riferiremo, a seconda del contesto, sia alla variabile aleatoria, sia alla causa di rischio.

spingere le imprese operanti nel settore assicurativo a costruire un sistema efficiente di controllo interno, in grado di tenere conto delle specificità di tutti i rischi assunti dalla singola impresa in un'ottica risk-based: rispetto alla normativa precedente, la fonte di rischiosità per l'impresa di assicurazione non è derivante solo dalle passività, ossia dagli impegni scaturiti dalle polizze, ma anche dai rischi di mercato ai quali sono soggetti gli attivi e dalle interrelazioni tra le diverse tipologie di rischio. In questo nuovo schema, il metodo di calcolo del margine di solvibilità nella vecchia normativa presenta significative limitazioni.

Un altro obiettivo di Solvency II è dato dalla necessità di armonizzare e uniformare la legislazione europea in merito alla vigilanza sulle assicurazioni, rendendola coerente non solo tra i vari paesi dell'unione, ma anche con le corrispondenti normative bancarie e con gli standard contabili internazionali: anche in precedenza le direttive comunitarie fornivano uno schema di riferimento che i paesi membri dovevano seguire, ma sulle concrete modalità di calcolo del margine di solvibilità era lasciata notevole libertà ai diversi paesi, generando significative differenze nell'ammontare dei requisiti di solvibilità richiesti, il che, in particolare considerando la crescente internazionalizzazione del settore assicurativo e la liberalizzazione del mercato, risultava difficilmente sostenibile e non avvantaggiava la comparabilità degli ammontari tra imprese operanti in paesi diversi.

Solvency II, come l'antecedente normativa Basilea II per il settore bancario, segue una struttura a tre pilastri, rappresentanti le tre categorie di requisiti richiesti alle imprese assicurative:



La struttura a tre pilastri di Solvency 2.

- Il primo pilastro rappresenta i requisiti quantitativi richiesti all'impresa, ossia, in particolare, esso specifica la struttura e le norme di calcolo del Solvency Capital Requirement e del Minimum Capital Requirement, ma anche le metodologie da utilizzare per la valutazione degli attivi e delle riserve tecniche.
- Il secondo pilastro rappresenta i requisiti qualitativi: rispetto alla normativa precedente, i requisiti quantitativi non sono più visti come sufficienti a garantire la solvibilità dell'impresa con alto grado di fiducia, e risulta quindi necessario un sistema interno di vigilanza risk-based. Rientrano nel secondo pilastro i requisiti relativi al sistema di governance dell'impresa e al modo in cui la funzione di risk management deve integrarsi con la struttura organizzativa dell'impresa. Viene richiesto, in particolare, che l'impresa sia dotata di adeguati processi per identificare e quantificare - in modo coerente e basandosi su dati possibilmente interni - i rischi ai quali è esposta e che le decisioni economiche effettuate dall'azienda tengano debitamente conto dell'impatto di tali rischi.

- Il terzo pilastro, infine, rappresenta i requisiti di disclosure, ossia di trasparenza, per i quali devono essere rese pubbliche descrizioni dettagliate su esposizione e concentrazione rispetto a ogni categoria di rischio e sulle tecniche di mitigazione adottate, oltre che informazioni relative all'adeguatezza del capitale (incluso l'SCR) e alle metodologie di valutazione di attivi e passivi.

Analizzeremo più in dettaglio soltanto il primo dei tre pilastri, in quanto la restante parte di questa tesi si concentrerà su modelli per la determinazione dell'SCR. Va detto, tuttavia, che i pilastri sono fortemente interconnessi tra di loro, e i modelli stocastici per il calcolo della perdita che noi vedremo al fine della costruzione di un modello interno per l'SCR (primo pilastro) possono essere anche alla base della misurazione interna del capitale nell'*own risk and solvency assessment* (secondo pilastro)².

Per quanto riguarda in dettaglio il primo pilastro, esso specifica la struttura che deve assumere lo stato patrimoniale di un'impresa di assicurazione. Di seguito sono riportate le principali modifiche rispetto alla normativa precedente:

- Le riserve tecniche non sono più calcolate secondo il principio della persona prudente, ossia aumentate implicitamente o esplicitamente per tenere conto del rischio, ma sono composte da due componenti distinte, la *best estimate* - rappresentante il valore atteso dei flussi futuri, scontati al tasso *risk-free* - e il *risk margin*. Quest'ultima componente rappresenta una stima del compenso richiesto dagli operatori di mercato per farsi carico del rischio associato alla passività. Per calcolare il *risk margin*, Solvency II ha adottato un approccio basato sul costo di mantenimento del capitale (CoC): la riserva tecnica per un portafoglio è interpretata come l'importo che l'assicuratore dovrebbe pagare se dovesse trasferire i relativi impegni ad un altro assicuratore, e il *risk margin* è calcolato determinando il costo della costituzione dell'ammontare di fondi

²In realtà, come vedremo, l'SCR segue un'ottica monoannuale, mentre l'ORSA, secondo le linee guida fornite dall'EIOPA, misura il capitale anche con proiezioni a medio e a lungo termine. Si potrebbero, in linea teorica, estendere i modelli al caso pluriennale, tuttavia la letteratura esistente in merito è molto scarsa. Alcuni suggerimenti su come estendere un modello di stampo LSMC a questo contesto sono forniti da Hørig (2013) e da Morrison et al. (2013).

propri pari al requisito patrimoniale di solvibilità necessario a far fronte alle obbligazioni del portafoglio considerato in un'ottica di run-off³.

- Al criterio del valore contabile utilizzato storicamente nel bilancio è stato sostituito il criterio del fair value nella valutazione di attivi e passivi. Ove possibile, solitamente si ritiene di far coincidere il fair value di un bene con il suo valore di mercato, mentre se non è presente un mercato attivo per il bene (come spesso accade, in particolare, per le best estimates delle passività assicurative) né se ne può ricavare il valore con opportuni aggiustamenti a partire da mercati attivi di beni simili a quello oggetto di valutazione, si dovrà utilizzare un modello adeguato come proxy del fair value; il modello dovrà comunque, per quanto possibile, essere calibrato a partire da dati di mercato.
- Il capitale proprio è suddiviso in capitale di solvibilità (SCR), del quale tratteremo in maniera più approfondita nel seguito, e Free Surplus. Inoltre, la direttiva identifica un livello minimo di capitale di solvibilità (Minimum Capital Requirement, o brevemente MCR), sotto al quale l'impresa non deve mai scendere, pena la revoca dell'autorizzazione qualora la situazione non sia sanata nel breve termine. L'MCR viene calcolato tramite una formula semplificata e deve collocarsi tra il 25% e il 45% dell'SCR, oltre che essere superiore a dei valori minimi fissati a seconda della tipologia di attività svolta. Per i dettagli relativi al calcolo del Minimum Capital Requirement, di cui non tratteremo ulteriormente in questa sede, si rimanda all'articolo 129 della direttiva Solvency II.

³Formalmente si ipotizza di trasferire la passività ad una Reference Undertaking priva di obbligazioni e di fondi propri prima del trasferimento, e che essa costituisca fondi propri pari agli SCR richiesti per farsi carico della passività fino alla sua estinzione, infine si calcola il margine $RM_{CoC} = c \sum_{t \geq 0} SCR_{RU}(t)(1 + r_{t+1})^{-(t+1)}$ ove $SCR_{RU}(t)$ è l'SCR che la Reference Undertaking deve allocare al tempo t per coprire il rischio nell'anno $[t, t+1]$, r_t il tasso privo di rischio con scadenza t , c il tasso di costo del capitale (usualmente posto pari al 6%), ossia un tasso di lungo periodo rappresentante l'extra-rendimento oltre il tasso risk-free e interpretabile come il costo-opportunità derivante dal dover investire il capitale di solvibilità richiesto in investimenti privi di rischio anziché in attività più redditizie.

1.1.2 Il Solvency Capital Requirement

La direttiva Solvency II definisce l'SCR nell'art. 101 come l'importo necessario a coprire con un livello di confidenza di almeno 99.5% ogni perdita imprevista nell'orizzonte temporale di un anno, valutando attivi e passivi in maniera consistente con il mercato e tenendo conto di tutti i rischi quantificabili a cui l'impresa è soggetta. Matematicamente, ciò corrisponde all'utilizzo di una misura di rischio di tipo Value-at-Risk (VaR) al livello 99.5% sulla variazione di fondi propri nel corso di un anno⁴.

Per formalizzare questi concetti, seguiamo l'approccio di Bauer et al. (2012). Definiamo il capitale disponibile (Available Capital, brevemente AC) come le risorse finanziarie a disposizione dell'impresa, utilizzabili ad esempio per coprire perdite impreviste. Consideriamo l'AC come differenza tra il valore di mercato degli attivi e il valore di mercato dei passivi. Nelle imprese di assicurazione spesso è facile calcolare il valore di mercato degli attivi, ma ciò non è il caso per i passivi. Si è giunti così alla formulazione del Market-Consistent Embedded Value (MCEV), un metodo di valutazione del capitale specifico per il settore assicurativo.

I dettagli relativi alle ipotesi sottostanti a questo metodo sono riassunti dai Market Consistent Embedded Value Principles pubblicati dal European Insurance CFO Forum nel 2009. Il MCEV è ivi definito come la somma del patrimonio netto rettificato (Adjusted Net Asset Value o ANAV) - ricavato a partire dal capitale netto definito a bilancio (Net Asset Value o NAV) con opportuni aggiustamenti - e del valore attuale dei profitti futuri (Present Value of Future Profits o PVFP), ossia il valore attuale atteso del flusso di cassa per gli azionisti dovuto al portafoglio in-force (escludendo quindi le entrate dal new business futuro), al netto del costo del capitale (CoC). Per maggiori dettagli su come calcolare queste quantità si rimanda ancora ai principi pubblicati dal CFO Forum. In particolare, la maggiore sfida risiede nel calcolo del PVFP, che, dipendendo fortemente dagli andamenti

⁴In realtà, il VaR non è l'unica misura di rischio che è stata presa in considerazione: nelle fasi iniziali di Solvency II, l'EIOPA (allora CEIOPS) propendeva per l'utilizzo del Tail VaR come misura di rischio più appropriata, per poi scartare l'idea a favore del più semplice VaR, al contrario della corrispondente normativa svizzera (Swiss Solvency Test), che ha optato per il Tail VaR al 99% anziché per il VaR al 99.5%. Per ulteriori dettagli sulle misure di rischio e le loro proprietà si rimanda a McNeil et al. (2005).

futuri del mercato finanziario richiede l'utilizzo di adeguati modelli stocastici per una valutazione approssimata.

Risulta pertanto ad una fissata epoca t :

$$MCEV_t = ANAV_t + PVFP_t - CoC_t$$

Formuliamo le seguenti ipotesi:

1. In $t = 0$, il capitale disponibile calcolato secondo i principi dettati da Solvency II coincida con il MCEV⁵, ossia valga $AC_0 = MCEV_0$.
2. Nell'anno $[0, 1]$ sia stato conseguito un utile X_1 (per tenere conto di una perdita è sufficiente considerare un utile negativo) e valga $AC_1 = MCEV_1 + X_1$.
3. La funzione di perdita sull'intervallo $[0, 1]$ valutata in $t = 0$ sia calcolata come:

$$L(0, 1) = AC_0 - \frac{AC_1}{1 + r(0, 1)}$$

ove $r(0, 1)$ è il tasso risk-free sull'intervallo $[0, 1]$.

Intuitivamente, posto $\alpha = 0.995$, potremmo definire l'SCR secondo la direttiva come la minima quantità x tale per cui vale:

$$P(AC_1 \geq 0 \mid AC_0 = x) \geq \alpha \tag{1.1}$$

Ciò infatti garantirebbe che se il capitale disponibile in 0 fosse x , allora la probabilità di essere solvente in 1 sarebbe maggiore del 99.5 %. Nella pratica, si preferisce tuttavia definire l'SCR nel seguente modo:

$$SCR_0 = \arg \min_x \{P(L(0, 1) > x) \leq 1 - \alpha\} = VaR_\alpha(L(0, 1)) \tag{1.2}$$

⁵Vi sono, in realtà, delle lievi differenze tra le due quantità, dovute alle differenze tra il risk margin di Solvency II e il CoC del MCEV e all'ammissibilità per il computo dell'AC di alcune componenti di capitale quali, ad esempio, i prestiti subordinati. Tali componenti possono tuttavia essere considerate come trascurabili ai fini di questa esposizione.

La definizione (1.2) è approssimativamente equivalente alla precedente ed è anche quella adottata dalla direttiva nell'art. 101. Per i dettagli formali relativi alle possibili definizioni matematiche dell'SCR si rimanda a Christiansen & Niemayer (2012). Ivi si analizzano in dettaglio le proprietà relative alle diverse definizioni formali dell'SCR e si dimostra, in particolare, che le due definizioni proposte sopra coincidono solo in casi restrittivi, ma anche che sotto opportune ipotesi si può verificare la convergenza dell'SCR calcolato secondo la definizione (1.2) a quello calcolato secondo la definizione (1.1).

Definiamo poi il Free Surplus/Excess Capital in 0 come $EC_0 = AC_0 - SCR_0$. Per la definizione (1.2) vale:

$$P\left(\frac{AC_1}{1+r(0,1)} \geq EC_0\right) \geq \alpha$$

Ossia si ha che la probabilità (nello stato di informazione relativo a $t = 0$) che il capitale disponibile in 1 attualizzato sia superiore all'Excess Capital è almeno pari ad α . Riorganizzando l'equazione sopra si ottiene inoltre:

$$P(AC_1 - EC_0(1+r(0,1)) \geq 0) \geq \alpha$$

Ne segue che nelle nostre ipotesi si assume implicitamente che l'Excess Capital sia investito al tasso risk-free.

Definiamo infine il Solvency Ratio come $\frac{AC_0}{SCR_0}$, uno degli indicatori sintetici più utilizzati in pratica per misurare il grado di patrimonializzazione di un'impresa di assicurazione e la sua capacità di far fronte a eventi imprevisti.

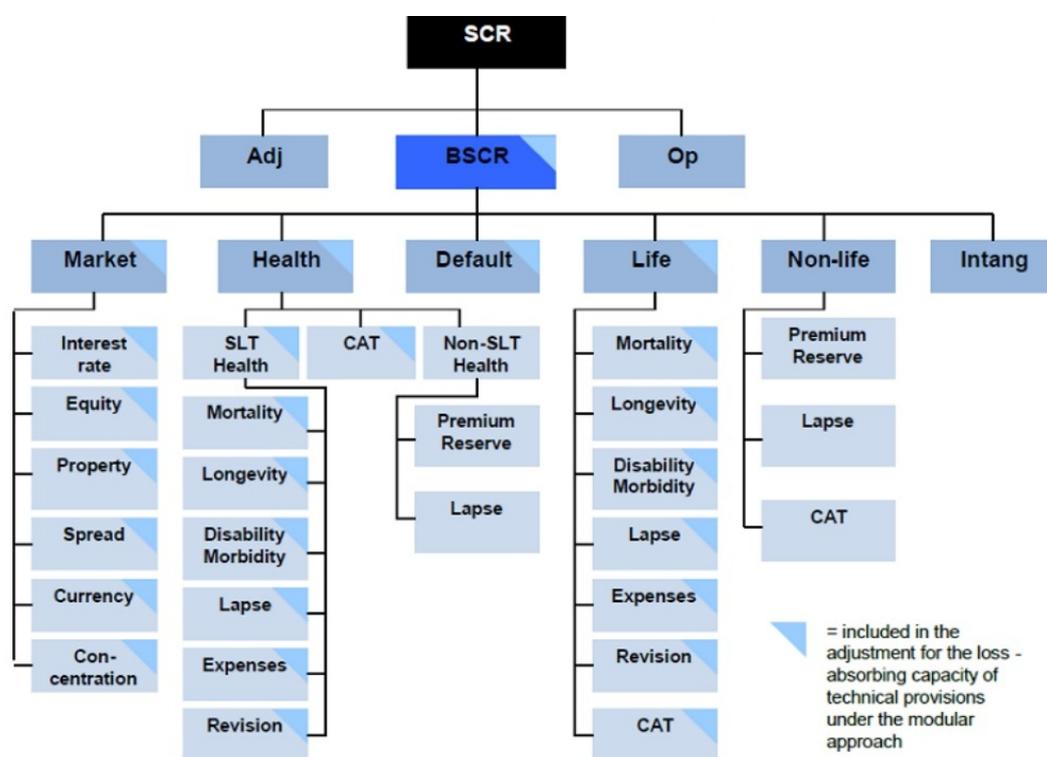
1.1.3 La Standard Formula

La standard formula proposta per il calcolo dell'SCR mira a fornire una procedura semplificata che sia relativamente facile da implementare anche per le imprese assicurative di piccole dimensioni e che in un qualche modo fornisca un'approssimazione dell'SCR per come lo abbiamo definito nella sezione precedente.

La struttura della standard formula ha subito varie modifiche nel percorso di introduzione di Solvency II. Noi andremo direttamente ad analizzare la sua forma definitiva a seguito del QIS 5 nel 2010.

La standard formula segue un approccio modulare e riflette l'impostazione risk-based assunta dalla direttiva: il rischio globale a cui è esposto l'assicuratore è diviso in più moduli di rischio, ognuno diviso a sua volta in sotto-moduli. L'impostazione da seguire per la sua applicazione si articola a grandi linee in due passi:

1. Calcolo del capitale di solvibilità richiesto per ogni causa di rischio vista singolarmente.
2. Aggregazione degli importi per singola causa tenendo conto di correlazioni predefinite tra i rischi.



La struttura modulare alla base della standard formula. Fonte: QIS 5 Technical Specifications

Più in dettaglio, l'SCR complessivo è visto come somma del Basic Solvency Capital Requirement (BSCR), del requisito di solvibilità per rischi operativi e di una componente negativa di aggiustamento che mira a tenere conto della capacità di assorbimento di perdite delle riserve tecniche e delle imposte differite.

$$SCR = BSCR + Adj + Op \quad (1.3)$$

ove il Basic Solvency Capital Requirement a sua volta ha una struttura modulare ed è calcolato tramite la formula:

$$BSCR = \sqrt{\sum_{i,j} SCR_i SCR_j \rho_{ij} + SCR_{int}} \quad (1.4)$$

ove:

- SCR_i, SCR_j rappresentano i capitali di solvibilità relativi alle 5 grandi categorie di rischio considerate: market risk, counterparty default risk, life underwriting risk, non-life underwriting risk, health underwriting risk.
- SCR_{int} rappresenta il requisito di capitale relativo ai beni intangibili.
- Le correlazioni ρ_{ij} sono specificate dall'EIOPA e riassunte nella seguente tabella:

i \ j	Market	Default	Life	Health	Non-life
Market	1				
Default	0.25	1			
Life	0.25	0.25	1		
Health	0.25	0.25	0.25	1	
Non-life	0.25	0.5	0	0	1

La matrice di correlazione tra le categorie di rischio. Fonte: QIS 5 Technical Specifications

I singoli requisiti di capitale relativi alle 5 grandi categorie sono nuovamente calcolati con un approccio modulare, seguendo la stessa struttura della formula (1.4) a partire dai requisiti di capitale per ognuno dei sotto-moduli. Le matrici di correlazione sono ancora specificate dall'EIOPA.

Per calcolare i requisiti di capitale relativi a ciascun sotto-modulo, ci si basa su un approccio deterministico, generalmente diverso per ciascun sotto-modulo: in linea generale, a partire da ogni fattore di rischio visto singolarmente, si considera uno scenario stressato per quel fattore di rischio (specificato dall'EIOPA) e si studia l'andamento di una opportuna grandezza di volume, come la variazione dei fondi propri, in quello scenario. Maggiori dettagli sulla struttura della standard formula sono contenuti nella sezione 2 del documento *QIS 5 Technical Specifications* (2010), disponibile sul sito dell'EIOPA.

Per comprendere a fondo quale sia il legame tra la standard formula e la definizione dell'SCR che abbiamo indicato nella sezione precedente, dobbiamo introdurre il metodo di varianza-covarianza per il calcolo del VaR, storicamente chiamato metodo delta-normal e ampiamente utilizzato in finanza per calcolare in via approssimata il VaR di un portafoglio tramite una formula analitica.

Dato un portafoglio, consideriamo il vettore delle variabili di stato $\mathbf{x}_t = (x_{1,t}, \dots, x_{n,t})$ e il vettore dei rendimenti su un orizzonte temporale Δt , $\mathbf{r}_t = (r_{1,t}, \dots, r_{n,t})$ ove $r_{i,t} = \frac{x_{i,t+\Delta t} - x_{i,t}}{\Delta t}$. Sia poi $P(t, \mathbf{x}_t)$ il valore del portafoglio in t .

Le ipotesi alla base del modello sono:

1. Il vettore dei rendimenti abbia distribuzione normale multivariata⁶ con valore atteso pari al vettore nullo e matrice di varianze e covarianze costante rispetto a t , ossia ipotizziamo $\mathbf{r}_t \sim \mathcal{N}^{(n)}(\mathbf{0}, \Sigma_r)$.
2. La funzione $P(t, \mathbf{x}_t)$ sia una funzione lineare, e quindi, in particolare, derivabile rispetto a ciascun argomento e con derivate di ordine superiore al primo nulle.

Sviluppando $P(t, \mathbf{x}_t)$ in serie di Taylor intorno a (t_0, \mathbf{x}_0) , essendo tutte le derivate di ordine superiore al primo nulle si ha:

$$P(t, \mathbf{x}_t) = P(t_0, \mathbf{x}_0) + \frac{\partial}{\partial t} P(t, \mathbf{x}_0)(t - t_0) + \mathbf{g}(t_0)^T(\mathbf{x}_t - \mathbf{x}_0)$$

⁶In realtà è sufficiente che la distribuzione sia ellittica, ma per semplicità di esposizione consideriamo il caso classico di normalità.

ove abbiamo indicato con $\mathbf{g}(t_0)$ il vettore delle derivate parziali di $P(t_0, \mathbf{x}_t)$ rispetto agli elementi di \mathbf{x}_t . Posto per semplicità $\Delta t = t - t_0$ e $\Delta \mathbf{x} = \mathbf{x}_t - \mathbf{x}_0$, $P'_t = \frac{\partial}{\partial t} P(t, \mathbf{x}_0)$, $\mathbf{g} = \mathbf{g}(t_0)$, la variazione del valore del portafoglio in Δt sarà:

$$\Delta P(t, \mathbf{x}_t) = P'_t \Delta t + \mathbf{g}^T \Delta \mathbf{x}$$

Essendo $\Delta \mathbf{x} = \text{diag}(\mathbf{x}_0) \mathbf{r}_t$, per l'ipotesi di normalità vale $\Delta \mathbf{x} \sim \mathcal{N}^{(n)}(\mathbf{0}, \Sigma)$ con $\Sigma = \text{diag}(\mathbf{x}_0) \Sigma_r \text{diag}(\mathbf{x}_0)$ e quindi sfruttando ancora le proprietà della normale multivariata risulta:

$$\Delta P(t, \mathbf{x}_t) \sim \mathcal{N}(P'_t \Delta t, \mathbf{g}^T \Sigma \mathbf{g}) \quad (1.5)$$

Ricordando poi che il Value-at-Risk di livello α per il numero aleatorio L raffigurante la perdita di portafoglio è definito come:

$$VaR_\alpha(L) = \inf\{l \in \mathbb{R} : \Pr(L > l) \leq 1 - \alpha\}$$

allora posto $L = -\Delta P(t, \mathbf{x}_t)$ si ha:

$$\Pr(-\Delta P(t, \mathbf{x}_t) > VaR_\alpha(L)) = 1 - \alpha \iff \Pr\left(Z > \frac{VaR_\alpha(L) + P'_t \Delta t}{\sqrt{\mathbf{g}^T \Sigma \mathbf{g}}}\right) = 1 - \alpha$$

ove abbiamo indicato con $Z = \frac{-\Delta P(t, \mathbf{x}_t) + P'_t \Delta t}{\sqrt{\mathbf{g}^T \Sigma \mathbf{g}}}$ una variabile aleatoria normale standard. Ne segue:

$$VaR_\alpha(L) = -P'_t \Delta t + \Phi^{-1}(\alpha) \sqrt{\mathbf{g}^T \Sigma \mathbf{g}} \quad (1.6)$$

ove come usuale abbiamo indicato con Φ la funzione di ripartizione della normale standard.

A partire da questo modello, per ricondurci alla standard formula interpretiamo \mathbf{x}_t come vettore dei fattori di rischio e $P(t, \mathbf{x}_t)$ come il valore del patrimonio netto in funzione dei fattori di rischio. Inoltre, coerentemente con Solvency II, poniamo $\Delta t = 1$, e, per rendere più scorrevole la notazione, $P(t, \mathbf{x}_t) = P(\mathbf{x})$. Ipotizzando ancora che i fattori di rischio catturino interamente la variazione di capitale sul periodo, ossia che valga $P'_t = 0$, ne consegue:

$$VaR_\alpha(L) = \Phi^{-1}(\alpha) \sqrt{\mathbf{g}^T \Sigma \mathbf{g}} \quad (1.7)$$

A partire dalla relazione $Cov(x, y) = \sigma_x \sigma_y \rho_{xy}$ possiamo riscrivere la matrice di varianze e covarianze come $\Sigma = V^T C V = V C V$ ove C è la matrice di correlazione e $V = diag(\boldsymbol{\sigma})$ con $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_n)$ e σ_i rappresenta la radice della varianza di Δx_i . Allora possiamo scrivere:

$$VaR_\alpha(L) = \Phi^{-1}(\alpha) \sqrt{\mathbf{g}^T V C V \mathbf{g}} = \sqrt{\mathbf{g}^T V \Phi^{-1}(\alpha) C \Phi^{-1}(\alpha) V \mathbf{g}}$$

Posto $\mathbf{s} = \Phi^{-1}(\alpha) V \mathbf{g}$ si ottiene ancora:

$$VaR_\alpha(L) = \sqrt{\mathbf{s}^T C \mathbf{s}} = \sqrt{\sum_{i,j} s_i s_j \rho_{ij}} \quad (1.8)$$

Per come abbiamo definito \mathbf{s} si ha $s_i = \Phi^{-1}(\alpha) \sigma_i \frac{\partial}{\partial x_i} P(\mathbf{x})$.

Considerata ora la variazione di valore del portafoglio dovuta al fattore di rischio x_i , che per l'ipotesi di linearità della funzione $P(\mathbf{x})$ è pari a $\Delta x_i \frac{\partial}{\partial x_i} P(\mathbf{x})$ e osservato che, per le proprietà della normale multivariata, nelle nostre ipotesi vale $\Delta x_i \sim \mathcal{N}(0, \sigma_i^2)$, allora posto $L_i = \Delta x_i \frac{\partial}{\partial x_i} P(\mathbf{x})$, procedendo analogamente a prima otteniamo $VaR_\alpha(L_i) = \Phi^{-1}(\alpha) \sigma_i \frac{\partial}{\partial x_i} P(\mathbf{x}) = s_i$.

Ne segue che possiamo interpretare gli s_i come Value-at-Risk a livello α della perdita relativa al singolo fattore di rischio x_i .

Riprendendo l'equazione (1.8) si ha quindi:

$$VaR_\alpha(L) = \sqrt{\sum_{i,j} VaR_\alpha(L_i) VaR_\alpha(L_j) \rho_{ij}} \quad (1.9)$$

Posto $\alpha = 0.995$ e ricordando la definizione di SCR otteniamo proprio la struttura della standard formula.

Abbiamo quindi visto il quadro teorico di riferimento sottostante alla standard formula. Ulteriori dettagli relativi alle assunzioni effettuate nel giungere alla standard formula, in particolare riguardo al calcolo degli SCR_i , sul quale non ci siamo addentrati, sono forniti da EIOPA (2014).

1.1.4 I limiti della Standard Formula

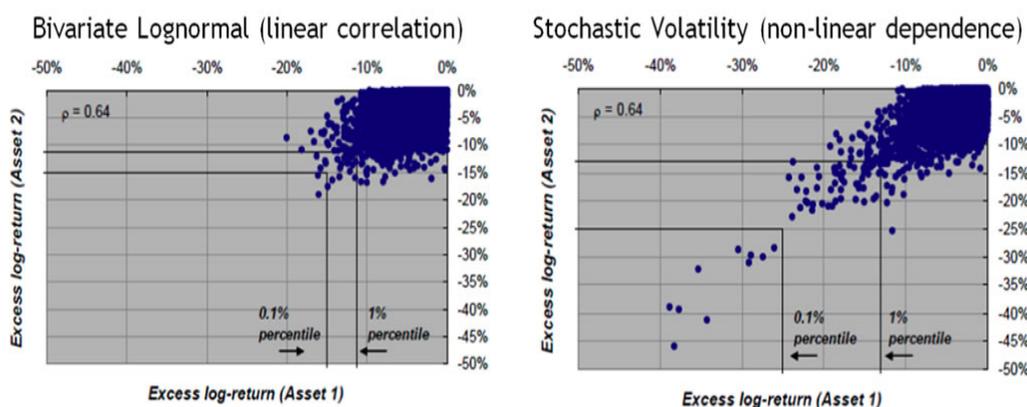
Nella sezione precedente abbiamo visto l'impianto teorico della standard formula e, in particolare, le due forti assunzioni alla sua base: la distribuzione congiuntamente normale dei fattori di rischio (che comporta la linearità della dipendenza tra i fattori di rischio, riassunta dal coefficiente di correlazione di Pearson) e la linearità della loro influenza sulla variazione di capitale. Entrambe queste assunzioni risultano spesso inadeguate nel contesto assicurativo, in particolare la linearità cade in presenza di opzioni e altre garanzie nei contratti (fenomeno molto diffuso nelle assicurazioni vita), e si è verificato che la standard formula tende a sotto-stimare l'impatto dei rischi. In parte, ciò è stato compensato tramite una scelta prudenziale delle correlazioni tra i fattori di rischio, spesso più elevate di quelle effettivamente osservate, per quanto ciò conduca ad un incremento implicito dell'SCR relativamente arbitrario, contrario all'ottica di trasparenza proposta dalla direttiva e incapace di cogliere adeguatamente la rischiosità specifica d'impresa. Peraltro, qualora un'impresa decidesse di adottare la standard formula, l'analisi dell'adeguatezza delle ipotesi alla base del modello le viene imposta dal regolatore nell'ambito dell'ORSA (secondo pilastro), e, qualora egli giudicasse che le ipotesi adottate non sono adeguate per quantificare la rischiosità specifica dell'impresa⁷, il regolatore può chiedere all'impresa l'allocazione di un capital add-on (art. 37 della direttiva), da aggiungere all'SCR come margine prudenziale di garanzia. È quindi nell'interesse dell'impresa stessa adottare un modello le cui ipotesi siano coerenti con la struttura del proprio portafoglio.

Riportiamo i risultati relativi ad alcuni studi riguardanti l'appropriatezza della standard formula:

- Koursaris (2010) studia l'andamento congiunto dei log-rendimenti di due titoli azionari, simulando una volta con volatilità stocastica, non assumendo ipotesi distributive e accettando dipendenze non-lineari; un'altra a partire dalla normale bivariata con stesse varianze e stessa correlazione tra i log-

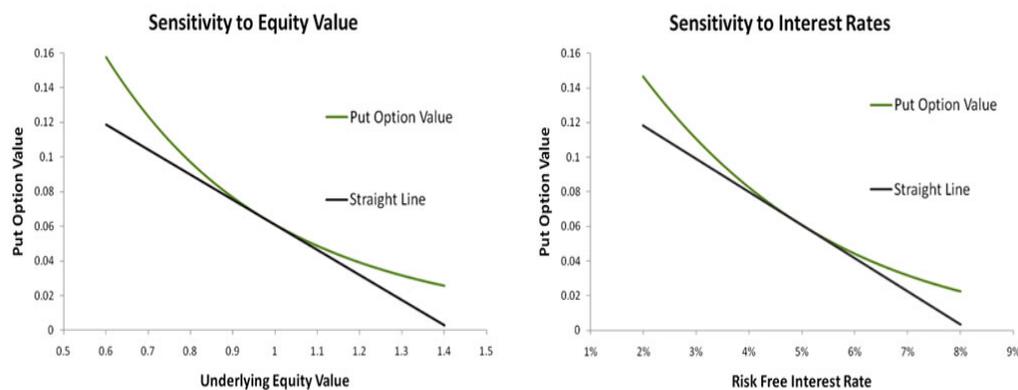
⁷Nell'art. 279 del regolamento delegato a supplemento della direttiva Solvency II, si considerano come inadeguate le ipotesi sottostanti al modello adottato se la deviazione dell'SCR rispetto a quello calcolabile con ipotesi più realistiche risulta superiore al 10% (o al 15% qualora vi siano forti evidenze a favore dell'adeguatezza del modello adottato).

rendimenti. Koursaris mostra che l'ipotesi di normalità sottostima significativamente le due code e, in particolare, il fenomeno della tail-dependence, ossia la dipendenza sulla coda (si verifica infatti che in certi casi anche variabili tra loro incorrelate possono presentare una forte dipendenza qualora si guardi esclusivamente alla coda della distribuzione). Ciò risulta particolarmente significativo in quanto per il calcolo del VaR è la coda della distribuzione a rivestire fondamentale importanza.



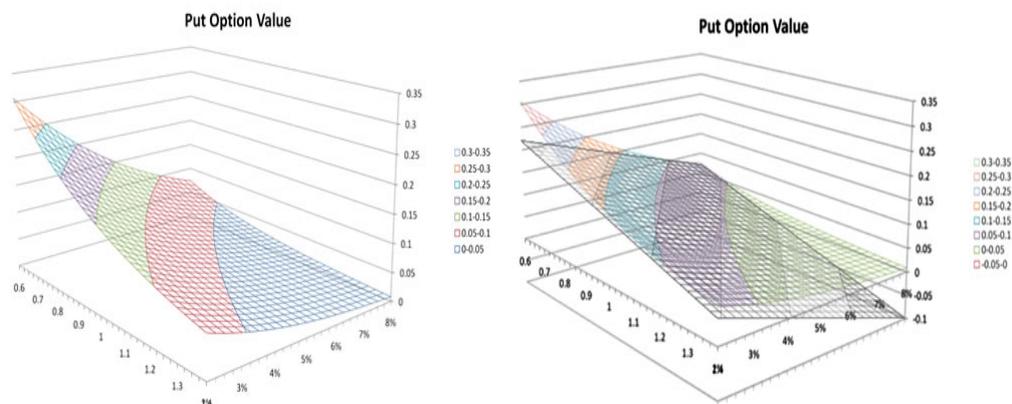
Distribuzioni congiunte dei log-rendimenti a confronto. Fonte: Koursaris (2010)

Inoltre, Koursaris studia l'andamento del valore di un'opzione put plain vanilla al variare del valore del sottostante e del tasso d'interesse risk-free.



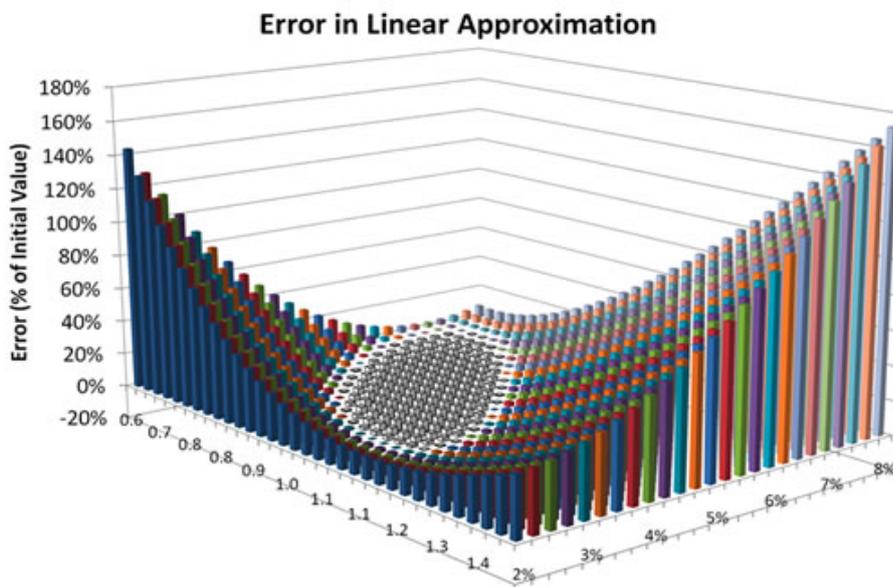
Andamento del valore dell'opzione rispetto ai due fattori di rischio considerati. Fonte: Koursaris (2010)

L'andamento, mostrato in figura, è non-lineare per entrambe le componenti. Per seguire l'approccio della standard formula bisogna quindi approssimare la funzione con una lineare:



Andamento congiunto del valore dell'opzione rispetto ai due fattori di rischio e approssimazione lineare. Fonte: Koursaris (2010)

L'approssimazione risulta buona solo per una piccola parte della superficie, e l'errore raggiunge valori superiori al valore iniziale dell'opzione. Visto che l'area importante per il calcolo dell'SCR dipenderà dalle correlazioni, non è assolutamente detto che essa ricada nella zona approssimata bene.



Errore dovuto all'approssimazione lineare in termini percentuali rispetto al valore iniziale dell'opzione. Fonte: Koursaris (2010)

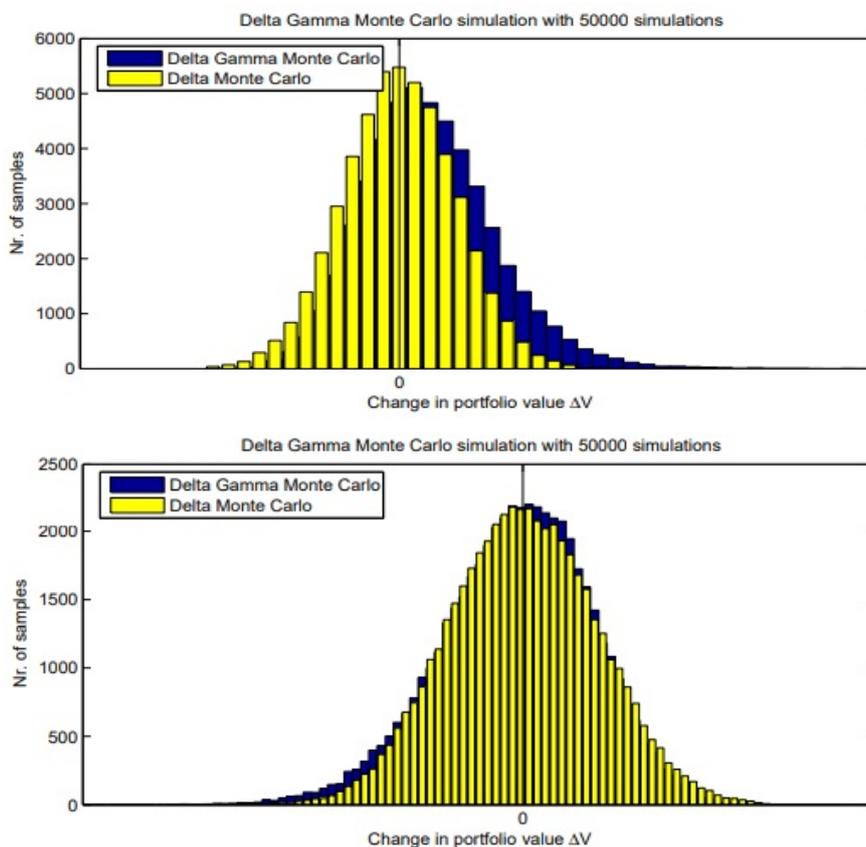
Considerato che i prodotti assicurativi vita spesso contengono una quantità significativa di opzioni implicite, dal valore anche non trascurabile, risulta allora evidente che per questi prodotti, per cui l'andamento della funzione di perdita non è lineare, la standard formula può portare ad errori anche molto elevati.

- Pfeifer & Strassburger (2008) mostrano come, anche se i capitali di solvibilità per i singoli moduli fossero noti e si possa assumere la simmetria della distribuzione congiunta, è possibile che la struttura di aggregazione della standard formula sottostimi significativamente l'SCR complessivo (anche del 20-25% in alcuni casi). Inoltre, Pfeifer e Strassburger dimostrano che a partire da alcune particolari strutture di dipendenza come le grid type copulas, anche in caso di rischi incorrelati l'approccio seguito dalla standard formula può risultare inadeguato.
- Sandström (2007) mostra come l'approccio adottato dalla standard formula porta ad una forte distorsione del valore dell'SCR in presenza di asimmetria nelle distribuzioni dei fattori di rischio. Sandström propone di utilizzare l'espansione di Cornish-Fisher per approssimare il quantile al 99.5% della distribuzione asimmetrica tramite il quantile al 99.5% della distribuzione normale, corretto in modo da tenere conto dell'asimmetria. Sandström addega questo approccio in modo da tener conto del principio di consistenza⁸ seguito dall'EIOPA, ottenendo una nuova versione della standard formula che dipenderà da un fattore di calibrazione legato, in particolare, all'asimmetria.
- Cardi & Rusnak (2007) confrontano il modello Delta-Normal della standard formula con il modello Delta-Gamma, che anziché basarsi su un'approssimazione di tipo lineare ammette anche una dipendenza quadratica⁹. Cardi e

⁸Il principio di consistenza richiede che ogni rischio della standard formula, segua, ove possibile, gli stessi principi e le stesse assunzioni di fondo, e che i diversi moduli siano calcolati in maniera consistente tra di loro.

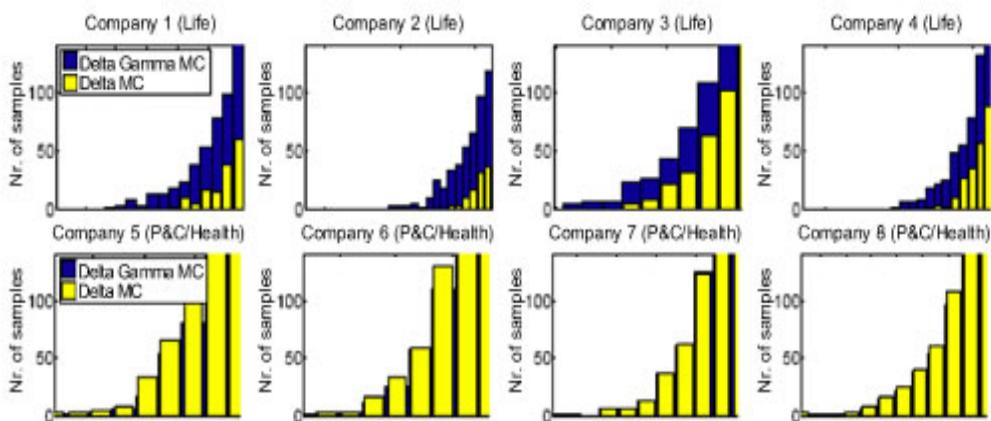
⁹I dettagli teorici sul modello Delta-Gamma sono disponibili in Cardi & Rusnak (2007); sostanzialmente si ripercorre lo schema che abbiamo seguito nella sezione 1.1.3 considerando l'approssimazione con il polinomio di Taylor al secondo anziché al primo ordine. Osserviamo che il termine Gamma con cui questo modello è conosciuto non si riferisce alla distribuzione Gamma, ma all'Hessiano di un portafoglio finanziario (così come il gradiente di un portafoglio è conosciuto come Delta).

Rusnak si basano sui dati di 38 compagnie di assicurazione, di cui 12 operanti nei rami vita, 12 nei rami salute e 14 nei rami danni. Essi ricavano la distribuzione simulata della variazione nel valore del portafoglio nei due modelli:



Confronto tra distribuzioni simulate della variazione nel valore del portafoglio per una compagnia danni (in alto) e una compagnia vita (in basso). Fonte: Cardi & Rusnak (2007)

Osserviamo che, per la compagnia danni considerata, la standard formula risulta in una stima molto prudente in quanto porta ad una coda sinistra più leggera. Per la compagnia vita invece, la coda sinistra è nettamente più pesante nel più accurato modello Delta-Gamma. Vediamo più in dettaglio l'andamento della coda:



Confronto tra distribuzioni simulate, coda sinistra in dettaglio per otto compagnie. Fonte: Cardi & Rusnak (2007)

Si verifica chiaramente che la standard formula conduce ad una drastica sottostima della coda sinistra della distribuzione per le compagnie operanti nei rami vita, mentre sovrastima le perdite per le compagnie danni o salute. Nell'articolo inoltre viene precisato che i portafogli vita considerati sono relativamente standard, senza effetti significativi dovuti alla presenza di opzioni implicite: anche in un contesto così semplice dunque la standard formula risulta spesso inadeguata. Nell'articolo sono contenute ulteriori analisi, anche in relazione all'utilizzo dell'espansione di Cornish-Fisher, che, pur portando a risultati migliori di quelli ottenuti con la standard formula, comunque comporta una sottostima significativa per le compagnie di assicurazione vita.

Ricordiamo infine che abbiamo posto l'attenzione unicamente alle ipotesi sottostanti alla struttura di aggregazione della standard formula, e non siamo mai entrati nei dettagli relativi al calcolo del requisito di capitale relativo a ciascun sotto-modulo, per i quali la standard formula effettua ulteriori ipotesi semplificatrici: ad esempio, diversi studi hanno mostrato l'inadeguatezza della standard formula nel cogliere adeguatamente il longevity risk nei prodotti di rendita vitalizia, oppure l'equity risk in portafogli particolarmente poco diversificati.

1.1.5 Verso un modello interno

Nelle sezioni precedenti abbiamo analizzato la standard formula e le sue principali limitazioni, che hanno spinto il regolatore a permettere e, anzi, ad incentivare l'utilizzo di un modello interno da parte delle compagnie di assicurazione. Qualora un'impresa decidesse di avvalersi di questa possibilità, il modello interno sarà soggetto all'approvazione da parte dell'autorità di vigilanza preposta.

Il modello interno può discostarsi - in tutto o in parte - dalla struttura alla base della standard formula, né vi sono altri vincoli strutturali. Tuttavia, il modello deve soddisfare una pluralità di requisiti, descritti dagli articoli 120-125 della direttiva Solvency II e riguardanti standard di qualità statistica (appropriatezza, accuratezza e completezza dei dati utilizzati, realistica e coerenza delle assunzioni alla base del modello), di calibrazione (l'SCR risultante dal modello deve essere riconducibile al VaR al 99.5% o comunque garantire almeno quel livello di protezione agli assicurati), di validazione (il modello deve essere sottoposto ad un processo di validazione interna che includa un monitoraggio della performance del modello, testando periodicamente che le ipotesi alla sua base siano ancora valide, anche in relazione ai nuovi dati), di documentazione (il modello deve essere adeguatamente documentato e la documentazione deve dimostrare che il modello verifichi gli altri standard nonché fornire dettagli relativi alla teoria, alle ipotesi matematiche, ai limiti relativi al modello), relativi all'assegnazione di profitti e perdite (la categorizzazione dei rischi nel modello deve riflettere il profilo di rischio e spiegare le cause e le fonti di perdite e profitti), di verifica dell'utilizzo (il modello deve essere ampiamente utilizzato nella governance dell'impresa, in particolare per quanto riguarda il risk management e i processi decisionali).

Tuttavia, la costruzione di un modello interno efficiente richiede una quantità ingente di lavoro e risorse, per cui nel breve termine per le piccole e medie imprese di assicurazione non risulta una strada percorribile. Qualora la standard formula risultasse inadeguata, al fine di non incorrere nel capital add-on è possibile seguire l'approccio dei parametri specifici d'impresa (Undertaking-Specific Parameters o USP), per cui viene mantenuta inalterata la struttura della standard formula, ma alcuni dei parametri standard forniti dall'EIOPA vengono sostituiti con parametri calibrati sui dati interni dell'impresa (o su dati di mercato opportuni qualora

non sia possibile utilizzare dati interni). Esistono diversi approcci di calcolo degli USP, definiti dall'EIOPA; l'impresa dovrà dimostrare che l'approccio scelto è il più adeguato oppure accettare l'approccio che fornisce il risultato più conservativo. In ogni caso l'utilizzo degli USP è soggetto ad approvazione da parte dell'autorità di vigilanza che verificherà l'adeguatezza dei dati utilizzati (art. 104 della direttiva Solvency II). L'autorità di vigilanza può inoltre obbligare l'impresa a utilizzare l'approccio USP per calibrare un sottoinsieme dei parametri se ritiene che la standard formula non sia adeguata a descrivere la rischiosità specifica dell'impresa (art. 110 della direttiva Solvency II).

Considerato che il focus di questa tesi verte sul modello interno, non tratteremo ulteriormente la tematica degli USP. Ulteriori informazioni su questo approccio possono essere rinvenute in CEIOPS (2010).

Un altro passo intermedio tra la standard formula e il modello interno pieno è costituito dal modello interno parziale. Grazie alla struttura modulare della standard formula, è infatti possibile utilizzare un modello interno per alcuni rischi, la standard formula (eventualmente con USP) per altri. Ciò permette un passaggio graduale dalla standard formula al modello interno pieno. Le norme relative all'approvazione del modello interno parziale ripercorrono sostanzialmente quelle per il modello interno pieno, e inoltre dovrà essere dimostrata la compatibilità dell'approccio con la standard formula e, qualora il modello parziale copra solo alcuni sotto-moduli di una classe di rischi oppure solo alcune linee di business della compagnia, può essere richiesto un progetto per la transizione ad un modello interno parziale più completo. L'autorità di vigilanza può inoltre imporre all'impresa di utilizzare un modello interno totale o parziale qualora ritenga che la standard formula non sia adeguata a descrivere la rischiosità specifica dell'impresa (art. 119 della direttiva Solvency II).

1.2 Approccio stocastico al calcolo dell'SCR

Volendo adottare un modello interno, per calcolare l'SCR come VaR al 99.5% si renderà necessario ottenere la distribuzione di probabilità della variazione di fondi propri su un orizzonte temporale di un anno. Vista la complessità di tale obiettivo, un approccio analitico non risulta sostenibile senza ingenti semplificazioni e in letteratura vi è un'intesa comune che il miglior metodo per risolvere questo problema è tramite simulazioni stocastiche: si generano un gran numero di scenari sull'evoluzione in un anno delle grandezze influenti sul bilancio (mortalità, andamento del mercato finanziario, riscatti, etc...) e in ogni scenario si ricavano i valori degli elementi dello stato patrimoniale e quindi anche un valore del patrimonio netto. In tal modo, si otterrà una distribuzione empirica del patrimonio netto ad un anno dalla quale si potrà agevolmente calcolare il VaR al 99.5%.

Andremo ora a descrivere in dettaglio questo approccio, che è anche la base fondante per tutti i modelli proxy analizzati in seguito.

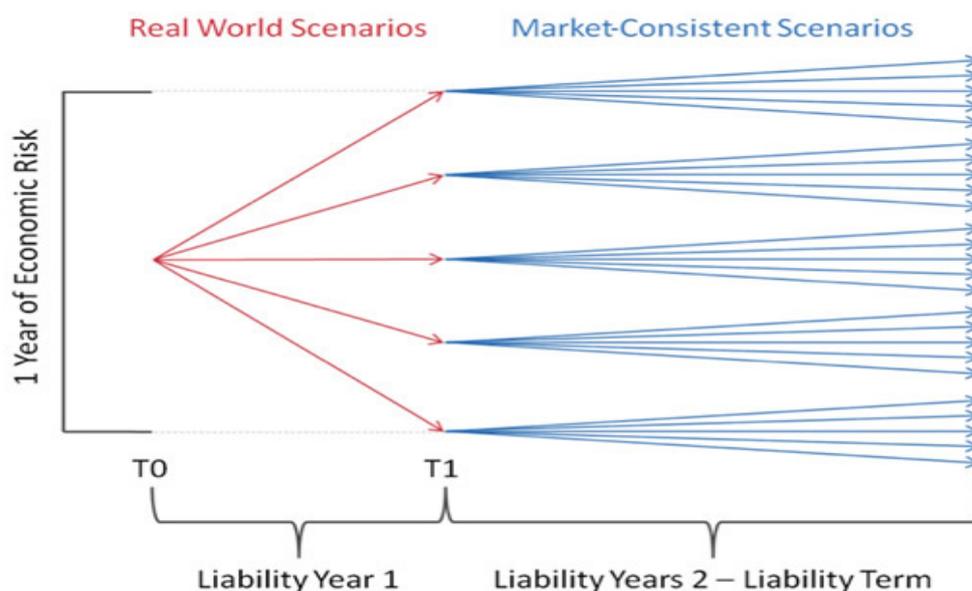
1.2.1 Simulazioni stocastiche annidate

Come abbiamo anticipato, l'approccio stocastico consiste nel generare via simulazione stocastica degli scenari e calcolare in ogni scenario il valore degli elementi dello stato patrimoniale. Fissiamo ora l'attenzione su un portafoglio di attività o di passività assicurative. Una volta determinati gli elementi esterni il cui andamento futuro influenza il valore del portafoglio (fattori di rischio o risk factor¹⁰) e una volta stimata la distribuzione della loro variazione congiunta in un anno¹¹, si procede a simulare da questa distribuzione. In ogni simulazione (outer scenario) a partire dal vettore delle realizzazioni dei fattori di rischio dovremo calcolare il valo-

¹⁰Vi è in letteratura una certa ambiguità relativa alla terminologia da utilizzare in merito: di norma, infatti, viene chiamato risk factor sia l'elemento fonte di rischiosità, sia la sua variazione nell'intervallo di tempo considerato, che è ciò che effettivamente impatta sul valore del portafoglio. Per mancanza di termini migliori, ci adegueremo a questa terminologia, lasciando al lettore il compito di discernere il concetto dalla variazione nel suo valore.

¹¹In realtà la determinazione della distribuzione congiunta dei fattori di rischio è un argomento estremamente complesso, e trattarne in maniera più approfondita non è scopo di questa tesi, nella quale ci poniamo nelle condizioni di avere già a disposizione la distribuzione da cui simulare. Tendenzialmente, per ottenere la distribuzione congiunta si calibra un modello univariato per ogni fattore di rischio e si procede all'aggregazione tramite copule.

re del portafoglio¹². Tuttavia, se ciò risulta spesso facile per portafogli di passività danni e per diverse categorie di attivi, per le quali esistono formule chiuse di valutazione a partire dalle determinazioni dei fattori di rischio, la notevole complessità di molti prodotti assicurativi vita - di durata elevata e contenenti opzioni implicite path-dependent - comportano la necessità di ricorrere a tecniche più complesse. In ogni scenario esterno sarà opportuno ottenere il valore del portafoglio utilizzando nuovamente un approccio stocastico, questa volta adottando un'ottica di run-off per simulare un gran numero di volte (inner scenarios) l'andamento del portafoglio per tutta la sua durata residua, condizionatamente alla realizzazione dei fattori di rischio nello scenario esterno, in modo da ottenere, per ogni scenario esterno, una distribuzione condizionata simulata del valore del portafoglio. Questo metodo, che si basa su un ciclo esterno di simulazioni e, per ogni simulazione esterna, su un ciclo interno di simulazioni condizionate, prende il nome di simulazioni stocastiche annidate (fully nested simulations o stochastic-on-stochastic).



Lo schema alla base delle simulazioni stocastiche annidate. Fonte: Koursaris (2010)

¹²Caratteristica fondamentale di questo passo affinché i principi di Solvency II siano rispettati è la market-consistency, per cui i prezzi attuali teorici ottenuti tramite il modello dovranno essere quanto più vicini possibile a quelli osservati sul mercato. Per approfondimenti su questi aspetti si consiglia Society of Actuaries (2016), un testo di carattere introduttivo che fornisce ampie basi sull'argomento.

Osserviamo che mentre le simulazioni esterne ad un anno devono essere realistiche (quindi prendono il nome di *real-world scenarios*) in modo da catturare al meglio l'evoluzione dei fattori di rischio nel prossimo anno, per essere coerenti con il principio di *market-consistency* gli scenari interni devono essere neutrali al rischio. La tematica della valutazione neutrale al rischio è ampiamente trattata nella maggior parte dei più recenti manuali di finanza quantitativa, nell'Appendice A sono in ogni caso descritti i principali risultati che utilizzeremo in seguito per una trattazione più formale delle simulazioni stocastiche annidate.

Osserviamo che ottenere una calibrazione *market-consistent* nell'ambito delle passività assicurative vita può in realtà essere una sfida notevole: infatti le passività derivanti da alcune polizze possono estendersi per decenni, e una calibrazione neutrale al rischio richiede un mercato liquido, che non è presente né per i portafogli assicurativi, né per le opzioni a lunghissima durata con le quali si può cercare di replicarne i payoff. Nel nostro lavoro non ci occuperemo di questa tematica, supponendo che il modello considerato sia in grado di ottenere una valutazione neutrale al rischio sufficientemente adeguata.

Schematicamente, l'implementazione di un modello *nested stochastic* può essere riassunta dai seguenti passi:

1. Identificazione dei rischi che influenzano gli elementi del bilancio;
2. Determinazione di una distribuzione marginale (analiticamente o empiricamente su base di dati storici) per la variazione di ciascun fattore di rischio in un anno;
3. Determinazione di una distribuzione congiunta della variazione dei *risk factor* a partire dalle marginali tramite copule;
4. Simulazione di un gran numero di scenari relativi alla variazione ad un anno dei *risk factor*;
5. Rivalutazione degli elementi del bilancio in ognuno degli scenari trovati al punto precedente. Per attività e passività più complesse (in particolare in ambito vita) ciò sarà effettuato seguendo nuovamente un approccio di simulazione stocastica;

6. Calcolo della variazione di patrimonio netto in ogni scenario;
7. Determinazione dell'SCR come VaR al 99.5% della distribuzione empirica trovata al punto precedente.

L'approccio stocastico qui illustrato è, da un punto di vista teorico, il più adeguato per risolvere il problema della determinazione del capitale di solvibilità. Esso infatti presenta notevoli vantaggi: l'idea dietro alla struttura di calcolo è intuitiva e, trattandosi di un approccio non parametrico, non richiede ipotesi matematiche aggiuntive, rendendo questa tecnica estremamente flessibile. Peraltro, esso rispetto alla Standard Formula consente di ottenere l'intera distribuzione della perdita, e quindi consente di calcolare anche altre misure di rischio o altri quantili (requisito fondamentale per un modello interno). Tuttavia, lo svantaggio principale del metodo è dato dallo sforzo computazionale richiesto per avere stime robuste: l'SCR infatti è il quantile al 99.5% della distribuzione, e per ottenere una stima affidabile della coda possono servire centinaia di migliaia di scenari esterni (infatti la maggior parte degli scenari in cui si valuta il portafoglio non contribuirà alla coda e quindi non sarà significativo per il calcolo dell'SCR). Inoltre, ricordando che anche le passività assicurative spesso sono di natura complessa, in ogni scenario esterno saranno necessarie quantomeno migliaia di scenari interni in modo da catturare correttamente il valore della passività. Pertanto, il numero totale di simulazioni che devono essere fatte per una valutazione accurata dell'SCR è estremamente elevato. Considerato poi che l'impresa dovrà effettuare diverse analisi di sensitività, che in questo modello richiedono di ripetere l'intero procedimento, il budget computazionale complessivo raggiunge anche i miliardi di scenari, molto al di sopra delle capacità di calcolo persino delle macchine moderne, motivo per il quale l'approccio nella sua forma più classica non è diffuso nella pratica assicurativa (sebbene possa comunque venir utilizzato per singoli prodotti particolarmente complessi oppure come strumento di validazione per altri modelli). Sono quindi state sviluppate tecniche per ridurre lo sforzo computazionale necessario, cui dedicheremo i successivi capitoli di questa tesi, dopo aver posto le basi formali dell'approccio stocastico.

1.2.2 L'impostazione formale del modello stocastico

Riprendiamo l'impostazione formale di Bauer et al. (2012) che abbiamo utilizzato per descrivere l'SCR nella sezione 1.1.2 e consideriamo un mercato finanziario aperto nel continuo in $[0, T]$ ove T rappresenta la massima scadenza presente nel nostro portafoglio. Rappresentiamo l'incertezza sul mercato tramite lo spazio di probabilità filtrato $(\Omega, \mathcal{A}, P, \mathcal{F})$ ove la filtrazione $\mathcal{F} = (\mathcal{F}_t)_{0 \leq t < T}$ con $\mathcal{F}_t \in \mathcal{A} \forall t$ è completa ($\forall t$ \mathcal{F}_t contiene gli eventi di probabilità nulla di \mathcal{A}) e continua a destra ($\forall t$ $\mathcal{F}_t = \bigcap_{u>t} \mathcal{F}_u$). Ω rappresenta lo spazio degli stati possibili del mercato, \mathcal{A} è una σ -algebra di eventi di Ω , P una misura realistica di probabilità, \mathcal{F}_t la σ -algebra degli eventi osservabili in t (ossia l'informazione disponibile in t).

Consideriamo poi un processo markoviano d -dimensionale sufficientemente regolare $\mathbf{Y} = (\mathbf{Y}_t)_{0 \leq t < T} = (Y_{t1}, \dots, Y_{td})_{0 \leq t < T}$, in grado di modellare l'incertezza sul mercato finanziario (per cui tutte le attività rischiose presenti sul mercato potranno essere espresse in termini di \mathbf{Y}). Supponiamo inoltre che esista un'attività localmente risk-free (bank account) $B = (B_t)_{0 \leq t < T}$ ove $B_t = e^{\int_0^t r_u du}$ e ove r_t è l'intensità istantanea di interesse al tempo t (per quanto ipotizzato, r_t sarà funzione di \mathbf{Y}_t) e anche che esista una misura di probabilità neutrale al rischio Q equivalente a P sotto alla quale i flussi di pagamento possono essere valutati tramite i flussi scontati attesi rispetto al numerario B (sotto ipotesi non molto restrittive questa assunzione equivale all'assenza di opportunità di arbitraggio sul mercato).

Infine, assumiamo che esistano T funzionali f_1, \dots, f_T in cui f_t deriva i profitti futuri del portafoglio in-force al tempo t a partire dallo sviluppo del mercato finanziario fino a t ¹³. A partire dagli f_t possiamo definire i profitti futuri dal portafoglio in-force come una sequenza di variabili aleatorie $\mathbf{X} = (X_1, \dots, X_T)$ ove $X_t = f_t(\mathbf{Y}_s, s \leq t)$.

Osserviamo che per il momento, per rientrare nel modello dei mercati finanziari, ci siamo limitati a considerare i rischi di mercato, ma anche altri fattori di rischio tipici dell'attività assicurativa (quali ad es. il rischio di mortalità) possono essere considerati all'interno del processo di stato e quindi rientrare nella trattazione seguita. Per semplicità di esposizione trascureremo tuttavia la componente relativa

¹³Ciò interpretativamente equivale ad ipotizzare che la compagnia abbia un modello di proiezione dei propri flussi.

al costo del capitale.

Ricordiamo che nella sezione 1.1.2 abbiamo definito il requisito di capitale come $SCR = \arg \min_x \{P(AC_0 - \frac{AC_1}{1+r(0,1)} > x) \leq 1 - \alpha\}$ ove $AC_0 = MCEV_0$, $AC_1 = MCEV_1 + X_1$ e $MCEV_t = ANAV_t + PVFP_t$. Essendo $r(0,1)$ noto in $t = 0$, dobbiamo procedere alla valutazione probabilistica di AC_0 e AC_1 .

Sfruttando la formula di valutazione neutrale al rischio (per una derivazione formale della formula si veda l'appendice A) possiamo determinare il PVFP al tempo 0 come speranza matematica sotto la misura neutrale al rischio Q della somma dei profitti futuri X_t attualizzati:

$$PVFP_0 = \mathbb{E}_Q \left[\sum_{t=1}^T X_t e^{-\int_0^t r_u du} \right] \quad (1.10)$$

Come abbiamo già anticipato, $PVFP_0$ spesso non è calcolabile analiticamente per via della complessità dell'interazione tra gli utili X_t e l'andamento del mercato finanziario (infatti $X_t = f_t(\mathbf{Y}_s, s \leq t)$). Pertanto seguiamo un approccio di tipo Monte Carlo e simuliamo, tramite la misura neutrale al rischio Q , K_0 traiettorie fino a T del processo di stato, $\tilde{\mathbf{Y}}^{(k)}$ per $k = 1, \dots, K_0$. A partire dalle determinazioni simulate del processo di stato possiamo ottenere determinazioni simulate degli utili X_t e dell'intensità d'interesse risk-free r_t e quindi stimare:

$$\widetilde{PVFP}_0 = \frac{1}{K_0} \sum_{k=1}^{K_0} \sum_{t=1}^T \tilde{X}_t^{(k)} e^{-\int_0^t \tilde{r}_u^{(k)} du} \quad (1.11)$$

Considerato poi che l' $ANAV_0$ è noto al tempo $t = 0$, potendo essere ricavato a partire dal capitale netto (NAV) definito a bilancio tramite opportuni aggiustamenti, avremo che:

$$\widetilde{AC}_0 = ANAV_0 + \widetilde{PVFP}_0 \quad (1.12)$$

Procediamo ora a stimare la distribuzione di $AC_1 = ANAV_1 + PVFP_1 + X_1$. A tal fine, procediamo a simulare N traiettorie (questa volta tramite la probabilità realistica P) per l'andamento del processo di stato nel primo anno, $(\tilde{\mathbf{Y}}_s^{(i)})_{s \leq 1}$ per $i = 1, \dots, N$. A partire da queste possiamo agevolmente trovare $\widetilde{ANAV}_1^{(i)}$ e $\tilde{X}_1^{(i)}$. Sempre in base alla formula di valutazione neutrale al rischio si ha poi:

$$PVFP_1 = \mathbb{E}_Q \left[\sum_{t=2}^T X_t e^{-\int_1^t r_u du} \mid (\mathbf{Y}_s)_{s \leq 1} \right] \quad (1.13)$$

Nella pratica, il $PVFP_1$ non dipende da tutta l'evoluzione del mercato finanziario in $[0, 1]$, ma soltanto da un vettore finito-dimensionale di variabili di stato $\mathbf{D}_1 = (D_{11}, \dots, D_{1m})$. Ne segue:

$$PVFP_1 = \mathbb{E}_Q \left[\sum_{t=2}^T X_t e^{-\int_1^t r_u du} \mid \mathbf{Y}_1, \mathbf{D}_1 \right] \quad (1.14)$$

E in particolare:

$$PVFP_1^{(i)} = \mathbb{E}_Q \left[\sum_{t=2}^T X_t e^{-\int_1^t r_u du} \mid (\mathbf{Y}_1, \mathbf{D}_1) = (\mathbf{Y}_1^{(i)}, \mathbf{D}_1^{(i)}) \right] \quad (1.15)$$

Come per il $PVFP_0$, il problema di valutazione dell'equazione (1.14) spesso non può essere risolto in via analitica e richiede un nuovo approccio simulativo. Fissato lo scenario esterno i simuliamo $K_1^{(i)}$ traiettorie dell'andamento del mercato in $[1, T]$ da cui possiamo ricavare determinazioni simulate degli X_t e r_t futuri. Allora:

$$\widetilde{PVFP}_1^{(i)} = \frac{1}{K_1^{(i)}} \sum_{k=1}^{K_1^{(i)}} \sum_{t=2}^T \widetilde{X}_t^{(ik)} e^{-\int_1^t \widetilde{r}_u^{(ik)} du} \quad (1.16)$$

E quindi:

$$\widetilde{AC}_1^{(i)} = \widetilde{ANAV}_1^{(i)} + \widetilde{PVFP}_1^{(i)} + \widetilde{X}_1^{(i)} \quad (1.17)$$

Ripetendo questo procedimento $\forall i$ otteniamo un campione di N determinazioni simulate di AC_1 e quindi anche N determinazioni di $L(0, 1) = AC_0 - \frac{AC_1}{1+r(0,1)}$. Osserviamo che essendo $r(0, 1)$ noto e AC_0 approssimato tramite la formula nell'equazione (1.12), l'unica fonte di aleatorietà della variabile $L(0, 1)$ è data da AC_1 e quindi che per trovare l'SCR sarà sufficiente trovare il quantile al 99.5% della distribuzione di $-AC_1$, che indichiamo con $q_{0.995}^{-AC_1}$. Infatti:

$$P \left(\widetilde{AC}_0 - \frac{AC_1}{1+r(0,1)} > x \right) = 1 - \alpha \iff P \left(-AC_1 \leq (x - \widetilde{AC}_0)(1+r(0,1)) \right) = \alpha$$

Da cui:

$$SCR = \widetilde{AC}_0 + \frac{q_{0.995}^{-AC_1}}{1 + r(0, 1)} \quad (1.18)$$

Per trovare $q_{0.995}^{-AC_1}$ un metodo semplice è basarsi sul corrispondente quantile empirico, $q_{(m)}^{-AC_1}$ con $M = \lfloor 0.995N + 0.5 \rfloor$. In alternativa si potrebbe pensare di utilizzare stime più robuste basate sulla teoria dei valori estremi oppure di sfruttare tecniche di smoothing stimando $q_{0.995}^{-AC_1}$ con una media ponderata su un insieme contiguo di valori della statistica d'ordine, centrato sul quantile empirico (ad esempio considerando i cento scenari con valore immediatamente superiore al quantile empirico e i cento scenari con valore immediatamente inferiore). Per quanto riguarda la scelta dei pesi, sia $2n$ l'ampiezza dell'intervallo di smoothing considerato e x è il quantile empirico al livello 0.995, una scelta semplice è data dal metodo di Gaussian Smoothing:

$$w_{(x+i)}^G = \frac{\phi\left(\frac{d(2i+1)}{2n}\right) - \phi\left(\frac{d(2i-1)}{2n}\right)}{\phi\left(\frac{d(2n+1)}{2n}\right) - \phi\left(\frac{d(2n-1)}{2n}\right)} \quad \forall i = -n, \dots, n$$

ove d è un parametro legato alla varianza della gaussiana che si vuole considerare. In alternativa si può considerare il Triangular Smoothing per cui $w_{(x+i)}^T = \frac{n+1-|i|}{(n+1)^2} \quad \forall i = -n, \dots, n$.

Si rimanda a Bauer et al. (2012) per ulteriori dettagli sulla variabilità delle stime e per la costruzione di un intervallo di confidenza per l'SCR.

Osserviamo che essendo AC_0 e AC_1 stimati sulla base di un numero finito di simulazioni, negli N scenari esterni non si ottiene propriamente una determinazione di $L(0, 1) = ANAV_0 + PVFP_0 - \frac{ANAV_1 + PVFP_1 + X_1}{1+r(0,1)}$ ma di $\widetilde{L}(0, 1) = ANAV_0 + \widetilde{PVFP}_0 - \frac{ANAV_1 + \widetilde{PVFP}_1 + X_1}{1+r(0,1)}$, introducendo quindi una distorsione nello stimatore dell'SCR anche al divergere di N . Per la legge dei grandi numeri si verifica tuttavia che \widetilde{L} converge quasi certamente a L al divergere del numero di simulazioni utilizzate per la stima di $PVFP_0$ e $PVFP_1$.

In particolare, Gordy & Juneja (2010) derivano una formula per l'errore quadratico medio $MSE = \mathbb{E} \left[(\widetilde{SCR} - SCR)^2 \right]$ e verificano che in generale il bias è positivo e che quindi l'SCR calcolato con questo approccio dovrebbe sovrastimare il "vero"

SCR. Ci si potrebbe allora aspettare che in media la probabilità che la compagnia sia solvibile sia superiore al 99.5% e per ottenere una distorsione discretamente ridotta è necessario uno sforzo computazionale immenso. Per l'impostazione tecnica di questo problema si rimanda a Gordy & Juneja (2010) e Bauer et al. (2012).

1.2.3 Ottimizzazione del modello stocastico

Come abbiamo notato, l'approccio delle simulazioni stocastiche annidate è astrattamente adeguato per il calcolo del capitale di solvibilità, ma richiede uno sforzo computazionale non sostenibile al giorno d'oggi. Si è cercato quindi di estendere e ottimizzare il modello in modo da ridurre il numero di scenari necessari per ottenere una stima sufficientemente precisa dell'SCR. Ci proponiamo in questa sezione di analizzare brevemente alcuni dei principali risultati e alcune delle più diffuse tecniche proposte in letteratura, che sebbene non siano sufficienti a garantire la possibilità di implementare pienamente questo approccio in pratica, torneranno in parte utili nella trattazione dei modelli proxy che analizziamo in seguito.

a) Selezione ottimale del numero di scenari

Gordy & Juneja (2010) e Bauer et al. (2012) analizzano la scelta ottimale per l'allocazione del budget computazionale nell'ipotesi semplificata $K_1^{(i)} = K_1 \forall i$. Essi considerano la seguente espressione approssimata dell'errore quadratico medio:

$$\begin{aligned} MSE_{SCR}(K_0, K_1, N) &= \mathbb{E} \left[\left(\widetilde{SCR} - SCR \right)^2 \right] = V(\widetilde{SCR}) + \mathbb{E} \left(\widetilde{SCR} - SCR \right)^2 \\ &\simeq \frac{\sigma_0}{K_0} + \frac{\theta_\alpha^2}{K_1^2 l^2(SCR)} + \frac{\alpha(1-\alpha)}{(N+2) l^2(SCR)} \end{aligned}$$

ove σ_0 è la radice della varianza di $PVFP_0$, $l(\cdot)$ è la densità della distribuzione della perdita, $\theta_\alpha = -\frac{1}{2} \int_{-\infty}^{+\infty} z^2 \frac{\partial}{\partial u} g_{K_1}(u, z) dz \Big|_{u=SCR}$ con $g_{K_1}(\cdot, \cdot)$ distribuzione congiunta di L, Z dove $Z = \frac{\sqrt{K_1}}{1+r(0,1)} (\widetilde{PVFP}_1 - PVFP_1)$ e $r(0, 1)$ indica il tasso privo di rischio sull'intervallo $[0, 1]$.

A partire dall'equazione approssimata sopra, consideriamo il problema di minimizzazione del MSE con il vincolo computazionale $\Gamma = K_0 + N K_1$ (si trascura lo

sfuerzo richiesto per la simulazione degli N scenari esterni, in quanto solitamente trascurabile rispetto a quello necessario per le simulazioni interne, e si trascura il fatto che le K_0 simulazioni hanno un anno di run-off in meno rispetto alle $N K_1$ in quanto solitamente T è molto elevato).

Tramite moltiplicatori di Lagrange fissato K_1 si trova:

$$N \simeq \frac{\alpha(1-\alpha)K_1^2}{2\theta_\alpha^2} \quad K_0 \simeq \frac{\sigma_0 K_1 l(SCR)}{\theta_\alpha} \sqrt{\frac{N K_1}{2}} \quad (1.19)$$

Tale risultato nella pratica non è immediatamente applicabile in quanto θ_α , $l(\cdot)$ e σ_0 sono quantità ignote. Bisognerebbe allora effettuare una simulazione con un numero ridotto di scenari in modo da stimare queste quantità prima della simulazione ufficiale, ma la stima di θ_α sarà inaccurata in quanto si tratta di una quantità che dipende dalla coda della distribuzione e quindi difficile da stimare accuratamente con un numero ridotto di simulazioni.

b) Tecniche di riduzione della varianza

Considerato uno stimatore per il valore atteso basato su n simulazioni Monte Carlo, sia σ^2 la varianza del campione ottenuto, allora per il Teorema del Limite Centrale lo standard error risultante è $\frac{\sigma}{\sqrt{n}}$, quindi per dimezzare la deviazione standard (ossia ad es. l'ampiezza dell'intervallo di confidenza) dobbiamo quadruplicare il numero di simulazioni. Si dicono tecniche di riduzione della varianza i metodi utilizzati per diminuire l'errore campionario senza aumentare il numero di simulazioni. Nel fornire un quadro sintetico relativo a queste tecniche, ci si basa principalmente su Glasserman (2003) e Higham (2004).

Il modello dall'implementazione più semplice è quello delle variabili antitetiche. Consideriamo $\mathbf{Z} = (Z_1, \dots, Z_m) \sim \mathcal{N}^{(m)}(\mathbf{0}, I_m)$ e il funzionale $f : \mathbb{R}^m \rightarrow \mathbb{R}$, sia $Y = f(\mathbf{Z})$ la quantità aleatoria d'interesse e $\mu = \mathbb{E}(Y)$ la sua speranza matematica. Allora utilizzando un approccio simulativo con n simulazioni i.i.d di \mathbf{Z} , lo stimatore più naturale per μ sarebbe $\tilde{\mu} = \frac{1}{n} \sum_{i=1}^n f(\mathbf{Z}_i)$. Consideriamo invece il seguente stimatore alternativo (stimatore delle variabili antitetiche):

$$\tilde{\mu}_{AV} = \frac{1}{n} \sum_{i=1}^n \frac{f(\mathbf{Z}_i) + f(-\mathbf{Z}_i)}{2} \quad (1.20)$$

Essendo $-\mathbf{Z}_i$ ancora un vettore di normali standard, lo stimatore così costruito non introduce distorsioni. Inoltre, la sua varianza sarà data da:

$$\begin{aligned} V(\tilde{\mu}_{AV}) &= \frac{1}{4n} (V(f(\mathbf{Z})) + V(f(-\mathbf{Z})) + 2Cov(f(\mathbf{Z}), f(-\mathbf{Z}))) \\ &= \frac{1}{2n} V(f(\mathbf{Z})) + \frac{1}{2n} Cov(f(\mathbf{Z}), f(-\mathbf{Z})) \end{aligned}$$

Allora, supponendo che il costo computazionale per generare una determinazione simulata della coppia $(f(\mathbf{Z}), f(-\mathbf{Z}))$ sia il doppio del costo per la generazione di una determinazione standard di Y ¹⁴, si verifica che se $Cov(f(\mathbf{Z}), f(-\mathbf{Z}))$ è negativa, lo stimatore delle variabili antitetiche ha varianza minore dello stimatore classico ottenibile con lo stesso budget computazionale.

Inoltre, se in particolare f fosse perfettamente lineare rispetto a \mathbf{Z} , allora la varianza dello stimatore così costruito sarebbe nulla (ciò deriva dal fatto che nel caso della normale standard la media aritmetica di due variabili antitetiche coincide con la speranza matematica della distribuzione). Tanto più f è vicina alla linearità, tanto maggiore sarà il beneficio dell'utilizzo di variabili antitetiche.

Condizione sufficiente affinché la covarianza sia negativa è la monotonia di f , condizione spesso verificata nell'ambito finanziario. Un esperimento di applicazione del metodo delle variabili antitetiche al computo dell'SCR con le simulazioni stocastiche annidate è stato effettuato da Bauer et al. (2012), ove per un fissato budget computazionale l'utilizzo di variabili antitetiche ha ridotto del 70% il MSE associato allo stimatore dell'SCR. Potenzialmente vi è dunque un vantaggio significativo nell'implementazione di questa tecnica al contesto in esame.

Finora siamo partiti da variabili normali standard, per cui la coppia di variabili antitetiche è $(Z, -Z)$. Il metodo in realtà ha una validità più generale, estendendosi in particolare a tutte le distribuzioni simmetriche rispetto all'origine e alle distribuzioni uniformi. Nel caso $U \sim Unif(0, 1)$ si ha la coppia di variabili antitetiche $(U, 1 - U)$ e valgono le stesse considerazioni che abbiamo visto nel caso di normalità.

¹⁴In realtà il costo è leggermente minore del doppio, infatti anziché dover generare $2n$ scenari è sufficiente generarne n e poi cambiare il segno. Tuttavia, possiamo considerare questa differenza come trascurabile in quanto solitamente il costo computazionale per la generazione degli scenari è irrisorio rispetto a quello per la valutazione del funzionale f negli scenari.

Un'altra comune tecnica di riduzione della varianza è il metodo delle variabili di controllo: considerato sempre il problema di stima per $\mu = \mathbb{E}(Y)$ con $Y = f(\mathbf{Z})$, supponiamo che a partire dalla generica simulazione \mathbf{Z}_i oltre ad Y_i possiamo ottenere anche la determinazione di un'altra variabile aleatoria X_i dipendente da \mathbf{Z}_i e di speranza matematica nota. Lo stimatore delle variabili di controllo sarà:

$$\tilde{\mu}_{CV} = \frac{1}{n} \sum_{i=1}^n (Y_i - b(X_i - \mathbb{E}(X))) = \bar{Y} - b(\bar{X} - \mathbb{E}(X)) \quad (1.21)$$

con $b \in \mathbb{R}$. Lo stimatore è non-distorto e consistente (vedi Glasserman (2003)). Inoltre si ha:

$$V(\tilde{\mu}_{CV}) = \frac{1}{n} (V(Y) - 2b \text{Cov}(X, Y) + b^2 V(X))$$

Pertanto, se $b^2 V(X) < -2b \text{Cov}(X, Y)$, allora lo stimatore con il metodo delle variabili di controllo ha varianza minore dello stimatore classico.

Si verifica che il valore ottimale di b è dato da $\frac{\text{Cov}(X, Y)}{V(X)}$. Il valore minimo della varianza dello stimatore ottenibile con questo metodo è quindi: $V(\tilde{\mu}_{CV}) = V(Y) - \frac{\text{Cov}(X, Y)^2}{V(X)}$

Affinché l'utilizzo di questa tecnica sia ammissibile, X e Y devono necessariamente essere correlate. Inoltre, la correlazione deve essere elevata se si vuole sperare di ottenere una riduzione significativa della varianza. Interpretativamente, infatti, questo metodo utilizza la conoscenza relativa all'errore nella stima di $\mathbb{E}(X)$ per migliorare la stima di $\mathbb{E}(Y)$, e quindi tanto più sono correlate X e Y , tanto più utili sono le informazioni ottenute su X . Le principali difficoltà del metodo sono la ricerca di una variabile di controllo X adeguata e di speranza matematica nota e il fatto che la scelta ottimale di b richiede la conoscenza della covarianza tra X e Y che difficilmente sarà nota se non conosciamo nemmeno il valore atteso di Y .

Ulteriori tecniche di riduzione della varianza sono presentate in Glasserman (2003). Osserviamo che tutti questi metodi comportano la riduzione dello standard error moltiplicandolo per un coefficiente minore di 1, ma l'ordine di infinitesimo rimane sempre $O(n^{-1/2})$. Vedremo nel prossimo paragrafo una tecnica con cui potenzialmente possiamo portare l'ordine di infinitesimo a $O(n^{-1})$.

c) Metodi quasi-Monte Carlo

Se gli algoritmi di generazione dei numeri pseudo-casuali hanno l'obiettivo di imitare la casualità calcolando tramite sequenze deterministiche numeri tali da sembrare determinazioni di una variabile aleatoria uniforme, gli algoritmi di generazione di numeri quasi-casuali (detti anche a bassa discrepanza) cercano di migliorare la precisione degli stimatori tramite numeri troppo regolari per essere davvero casuali, ma in grado di coprire in maniera più uniforme lo spazio. Per introdurre questi concetti, seguiamo ancora una volta Glasserman (2003).

Definiamo la discrepanza di un insieme di punti $\{x_1, \dots, x_n\}$ rispetto ad una famiglia \mathcal{A} di sottoinsiemi di $[0, 1]^d$ misurabili secondo Lebesgue come:

$$Dis_{\mathcal{A}}(x_1, \dots, x_n) = \sup_{A \in \mathcal{A}} \left| \frac{\#\{x_i \in A\}}{n} - v(A) \right| \quad (1.22)$$

ove indichiamo con $\#\{x_i \in A\}$ il numero di punti contenuti in A e con $v(A)$ la misura di volume di A . Inoltre, la discrepanza è detta ordinaria se \mathcal{A} è la famiglia dei rettangoli di $[0, 1]^d$ del tipo $\prod_{j=1}^d [a_j, b_j[$ con $0 \leq a_j < b_j \leq 1$; stellata (star discrepancy) se \mathcal{A} è la famiglia dei rettangoli di $[0, 1]^d$ del tipo $\prod_{j=1}^d [0, b_j[$ con $0 < b_j \leq 1$. Interpretativamente, la discrepanza di un insieme di punti rappresenta una misura di lontananza dall'uniformità.

Indicata con D la discrepanza ordinaria e con D^* quella stellata, Niederreiter (1992) dimostra che:

$$D^*(x_1, \dots, x_n) \leq D(x_1, \dots, x_n) \leq 2^d D^*(x_1, \dots, x_n)$$

Pertanto, fissato d , l'ordine di grandezza delle due misure di discrepanza è lo stesso. Per avvicinarci al problema delle simulazioni in esame, consideriamo in particolare il caso in cui fissiamo una successione $(x)_{n \in \mathbb{N}}$ e vogliamo misurare la discrepanza dei primi n termini della successione.

Niederreiter (1992) dimostra che per $d = 1$ vale:

$$D(x_1, \dots, x_n) \geq D^*(x_1, \dots, x_n) \geq B n^{-1} \log(n)$$

Inoltre egli ipotizza che per un generico d valga:

$$D(x_1, \dots, x_n) \geq D^*(x_1, \dots, x_n) \geq B_d n^{-1} \log(n)^d$$

con B_d dipendente solo dalla dimensione d . Tuttavia, a conoscenza di chi scrive, tale ipotesi è stata verificata soltanto per $d = 2$, rimanendo quindi indimostrata per $d \geq 3$.

La discrepanza riveste un ruolo chiave anche nella determinazione del limite superiore dell'errore di approssimazione: non ci proponiamo, in questa tesi, di fornire un quadro relativo al delicato modello teorico sottostante a questi risultati, per il quale rimandiamo a Niederreiter (1992). Sia sufficiente sapere che in relazione al problema di stima di $\mathbb{E}(g(\mathbf{U}))$ con $g : [0, 1]^d \rightarrow \mathbb{R}$ finita, $\mathbf{U} = (U_1, \dots, U_d)$ e $U_j \sim Unif(0, 1) \forall j$, supposti gli U_j indipendenti e ottenute n simulazioni di \mathbf{U} - indicate con $\mathbf{u}_1, \dots, \mathbf{u}_n$ - tramite un metodo quasi-Monte Carlo, si verifica la disuguaglianza di Koksma-Hlawka:

$$\left| \frac{1}{n} \sum_{i=1}^n g(\mathbf{u}_i) - \mathbb{E}(g(\mathbf{U})) \right| \leq V(g) D^*(\mathbf{u}_1, \dots, \mathbf{u}_n) \quad (1.23)$$

ove $V(g)$, supposto finito, dipende solo dalla funzione g e non da \mathbf{U} (formalmente si tratta della variazione di Hardy-Krause della funzione). Tale elegante risultato teorico trova poche applicazioni pratiche, in quanto calcolare le quantità necessarie a trovare la limitazione superiore dell'errore comporta uno sforzo computazionale enorme. Tuttavia, questo risultato fa capire quanto la minimizzazione della discrepanza possa potenzialmente giovare alla precisione dello stimatore.

Definiamo a bassa discrepanza le sequenze per cui la discrepanza stellata corrispondente ha ordine $O(\log(n)^d/n)$. Sebbene qualsiasi potenza di $\log(n)$ sia asintoticamente trascurabile rispetto ad n , nelle applicazioni pratiche, considerando valori realistici di n , storicamente la letteratura suggerisce che queste tecniche siano efficaci soltanto per un numero relativamente basso di dimensioni (alcuni autori pongono il limite a 12 o 15, altri a 40). Glasserman (2003) tuttavia argomenta a favore dell'efficacia dei metodi quasi-Monte Carlo - per quanto concerne le applicazioni finanziarie - anche con un valore di d molto più elevato.

Un primo semplice metodo per la generazione di numeri quasi-casuali è dato dalle sequenze di Van der Corput, che si basano sull'espansione in base b dei numeri naturali:

$$\forall k \in \mathbb{N} \exists! (a_j(k))_{j \in \mathbb{N}} : a_j(k) \in \{0, 1, \dots, b-1\} \forall j \wedge k = \sum_{j=0}^{+\infty} a_j(k) b^j \quad (1.24)$$

Ossia per una data base b ogni numero naturale ammette un'unica rappresentazione come combinazione lineare di potenze non negative di b con coefficienti in $\{0, 1, \dots, b-1\}$. Inoltre $\exists h : a_j(k) = 0 \forall j > k$, quindi la somma nella proposizione (1.24) è finita. La sequenza di Van der Corput è data dalle quantità $\phi_b(k) = \sum_{j=0}^{+\infty} a_j(k) b^{-(j+1)}$ al variare di k in \mathbb{N} . Ad esempio, posto $b = 2$ i primi otto valori della sequenza di Van der Corput sono $\{0, \frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{5}{8}, \frac{3}{8}, \frac{7}{8}\}$: i punti sono alternativamente maggiori e minori di 0.5, e quelli maggiori sono alternativamente maggiori e minori di 0.75, quelli minori alternativamente maggiori e minori di 0.25; la proprietà si ripete nello stesso modo per tutta la successione, e quindi mira a bilanciare al massimo la disposizione dei numeri nell'intervallo. Niederreiter (1992) dimostra che le sequenze di Van der Corput sono a bassa discrepanza.

In letteratura sono proposti diversi altri metodi per la generazione di numeri quasicasuali (si veda Glasserman (2003) per un'introduzione alle varie metodologie), noi in particolare analizzeremo ancora soltanto le sequenze di Sobol, introdotte dall'omonimo matematico russo nel 1967 e spesso utilizzate nelle applicazioni per le loro importanti proprietà e la relativa semplicità d'implementazione.

Per introdurre le sequenze di Sobol, definiamo innanzitutto per b intero ≥ 2 l'intervallo elementare in base b come sottoinsieme di $[0, 1]^d$ del tipo $\prod_{i=1}^d [\frac{a_i}{b^{j_i}} \frac{a_i+1}{b^{j_i}}[$ con $j_i \in 0, 1, 2, \dots$ e $a_i \in 0, 1, \dots, b^{j_i} - 1$ e il suo volume $b^{-(j_1+\dots+j_d)}$.

Si definisce poi (t, m, d) -rete in base b ($0 \leq t \leq m$) ogni insieme di b^m punti in $[0, 1]^d$ tale per cui b^t punti cadono in ogni intervallo elementare di volume b^t , ossia ogni insieme per il quale la frazione b^t/b^m di punti che cade in un generico intervallo elementare coincide con il volume dell'intervallo.

Definiamo infine una (t, d) -sequenza in base b come una successione (x_1, x_2, \dots) di punti di $[0, 1]^d$ tale per cui $\forall m > t$ ogni segmento $\{x_i : j b^m < i \leq (j+1) b^m\}$ per $j = 0, 1, 2, \dots$ è una (t, d, m) -rete in base b .

Data una (t, d) -sequenza in base b , minore è t e più si ha uniformità, infatti anche intervalli elementari piccoli contengono il numero corretto di punti. Inoltre, vi è interesse a mantenere b il più piccolo possibile in modo che la proprietà valga anche in insiemi con un numero b^m di punti più ridotto.

Niederreiter (1992) dimostra che ogni (t, d) -sequenza è una sequenza a bassa discrepanza. Per ulteriori risultati sulle proprietà delle (t, m, d) -reti e delle (t, d) -sequenze si rimanda sempre a Niederreiter (1992).

Le sequenze di Sobol sono storicamente il primo esempio di (t, d) -sequenza. Si tratta di (t, d) -sequenze in base 2 in cui t dipende da d . Esse nascono come permutazioni di sequenze di Van der Corput in base 2, ottenute moltiplicando espansioni binarie di numeri naturali consecutivi (si veda la proposizione (1.24)) con un set di d generatori $V^{(1)}, \dots, V^{(d)}$.

Il generico generatore $V^{(i)}$, che per comodità indicheremo nel seguito con V , sarà una matrice triangolare superiore i cui elementi sono 0 o 1 e le cui colonne sono le espansioni binarie di un set di direction numbers v_1, \dots, v_r , in cui r è arbitrariamente grande e nella costruzione del k -esimo punto della sequenza può essere pensato come numero di termini dell'espansione binaria di k .

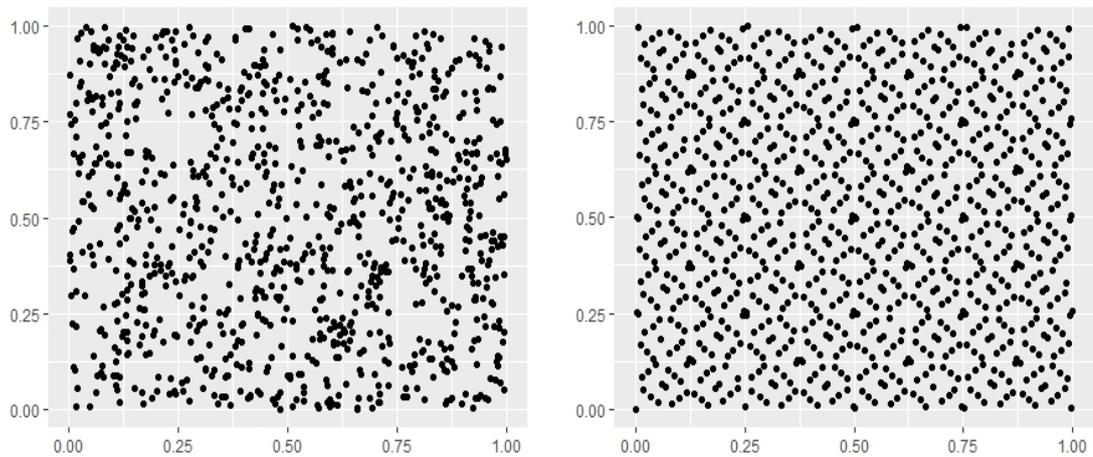
Dato \mathbf{a}_k vettore degli r coefficienti della rappresentazione binaria di k (se la costruzione di k richiede meno di r coefficienti, per come è costruita la proposizione (1.24) sarà sufficiente porre quelli successivi uguali a 0), troviamo i coefficienti $\mathbf{v}(k)$ dell'espansione binaria della coordinata i -esima del k -esimo elemento della sequenza di Sobol tramite:

$$\mathbf{v}(k) = V \mathbf{a}_k \tag{1.25}$$

Ripeteremo questo procedimento per ogni dimensione $i = 1, \dots, d$ del nostro problema, ogni volta con un generatore diverso.

L'utilizzo esclusivo della base 2, caratteristico dell'approccio di Sobol, porta a notevoli vantaggi computazionali rispetto a molte altre tecniche di costruzione di numeri quasi-casuali. Si rimanda a Sobol (1967) per i dettagli formali sulla costruzione dei generatori e a Glasserman (2003) per ulteriori approfondimenti e per suggerimenti relativi all'implementazione pratica di un generatore di sequenze di Sobol.

Concludiamo mostrando graficamente in due dimensioni quanto l'utilizzo di una sequenza a bassa discrepanza quale quella di Sobol copra lo spazio in maniera più uniforme rispetto ai numeri pseudo-casuali.



A sinistra 1000 simulazioni da un'uniforme bivariata su $[0, 1]^2$ tramite l'algoritmo di generazione di numeri pseudo-casuali Mersenne-Twister e a destra 1000 elementi di una sequenza di Sobol con $d = 2$.

d) Procedure di Screening

Come abbiamo già detto, le simulazioni stocastiche annidate sono un metodo estremamente inefficiente, in quanto richiedono uno sforzo computazionale significativo per calcolare la perdita in ogni scenario esterno, ma soltanto lo 0.5% di questi scenari sarà situato nella coda d'interesse. L'idea intuitiva dietro alla procedura di screening è quella di effettuare il ciclo interno di simulazione soltanto per gli scenari per cui effettivamente vi è una ragionevole probabilità che siano nella coda. In questo modo sarà possibile fornire una valutazione più accurata negli scenari di maggiore interesse per il calcolo dell'SCR, per quanto tramite questa tecnica si introduce un nuovo errore nel modello, infatti vi sarà sempre la possibilità che uno scenario escluso tramite la procedura di screening fornisca in realtà un contributo significativo per l'SCR.

L'idea dietro questo approccio ricorda la tecnica delle simulazioni subordinate (in particolare, subordinate alla coda di una distribuzione), talvolta utilizzata nella letteratura attuariale; esse prevedono di ottenere un campione più rappresentativo degli eventi estremi effettuando simulazioni condizionatamente al fatto di trovarsi sulla coda della distribuzione. Nel nostro contesto, non avendo ipotesi distributive, gli scenari esterni sono generati come nel modello di simulazioni stocastiche

annidate classico, ma le simulazioni interne sono generate soltanto per gli scenari esterni ritenuti plausibilmente sulla coda della distribuzione.

Jan et al. (2010) e Bauer et al. (2012) formalizzano quest'idea. Il procedimento si basa sulla costruzione di un primo set di simulazioni con $N^{(1)}$ scenari esterni e $K_1^{(1)}$ scenari interni per ogni scenario esterno e si determina la statistica d'ordine $(L_{(1)}, \dots, L_{(N^{(1)})})$ e l'SCR seguendo la metodologia esposta nella sezione 1.2.2. Si procede inoltre a determinare un intervallo di confidenza $[LB_{SCR}, UB_{SCR}]$ opportuno per l'SCR, seguendo, ad esempio, la metodologia proposta da Bauer et al. (2012). Si procede poi a fissare una probabilità di errore accettabile per la procedura di selezione, α_s e si effettua lo screening: fissato uno scenario esterno i , lo si confronta con ogni altro scenario esterno j tramite un test di verifica d'ipotesi al livello $\delta = \frac{\alpha_s}{(N_1 - q + 1)(q - 1)}$ ove q indica l'elemento della statistica d'ordine corrispondente al Lower Bound dell'intervallo di confidenza calcolato sulla base del primo set di simulazioni, ossia tale che $LB_{SCR} = L_{(q)}$.¹⁵ Il test considerato avrà $H_0^{(j)} = L_i \geq L_j$ e verrà ripetuto $\forall j$. Infine, lo scenario i viene mantenuto nel set soltanto se il numero di volte in cui $H_0^{(j)}$ viene rifiutata è minore di $N_1 - q + 1$. In formule quindi il set di scenari mantenuti sarà dato da:

$$I = \left\{ i : \sum_{j \neq i} \mathbb{1}_{\{L_i < L_j - t_{ij} \sigma_{ij}\}} < N_1 - q + 1 \right\} \quad (1.26)$$

Ove t_{ij} è il quantile di ordine $(1 - \delta)$ della distribuzione t di Student con gradi di libertà pari al minimo intero minore di $(K_1^{(1)} - 1) \left(1 + \frac{2}{(\sigma_1^{(i)}/\sigma_1^{(j)})^2 + (\sigma_1^{(j)}/\sigma_1^{(i)})^2} \right)$ e $\sigma_{ij} = \sqrt{\frac{(\sigma_1^{(i)})^2 + (\sigma_1^{(j)})^2}{K_1^{(1)}(1+r(0,1))^2}}$ con $\sigma_1^{(k)}$ radice della varianza di $PVFP_1$ nello scenario k .

Inoltre, considerato che la procedura di screening richiede una quantità estremamente elevata di confronti e quindi comunque uno sforzo computazionale non indifferente, viene proposta una procedura di pre-screening in modo da effettuare una prima selezione tramite un test più severo basato su un solo confronto e quindi ridurre il numero di confronti da effettuare nella successiva procedura di screening. Definita una permutazione $\pi(\cdot)$ degli indici tale che $L_{\pi(i)}$ è non-decrescente, sia

¹⁵Tale scelta di δ garantisce infatti di determinare opportunamente il livello di confidenza complessivo della stima intervallare per l'SCR risultante dall'applicazione della procedura. Per maggiori dettagli sugli intervalli di confidenza si rimanda a Bauer et al. (2012).

$\sigma_{i,\max} = \max_{j=q,\dots,N_1} \{\sigma_1^{(j)}\}$ e $t_{\max} = \max_{i=1,\dots,N_1} \{\max_{j=q,\dots,N_1} \{t_{ij}\}\}$. Il set di scenari mantenuti dalla procedura di pre-screening sarà:

$$I' = \{i : L_i \geq L_{\pi(q)} - t_{\max} \sigma_{i,\max}\} \quad (1.27)$$

In Bauer et al. (2012) si dimostra che $I \subseteq I'$, ossia che tutti gli scenari eliminati dal pre-screening sarebbero eliminati anche dallo screening. In altre parole, il pre-screening, pur riducendo lo sforzo computazionale dello screening, non introduce un ulteriore errore nel modello.

In Liu et al. (2010) si espande la procedura vista finora e si propone in maniera euristica un approccio iterativo di screening per il calcolo dell'Expected Shortfall, facilmente adattabile anche per il calcolo del VaR, ma che non consente di ricavare intervalli di confidenza né garantisce una soglia massima per la probabilità di errore.

Una volta completato il processo di screening, possiamo effettuare un nuovo step interno del processo di simulazione, generando ${}^{(i)}K_1^{(2)}$ simulazioni interne $\forall i \in I$. Una scelta immediata per la ripartizione del budget computazionale tra gli scenari rimasti è ${}^{(i)}K_1^{(2)} = K_1^{(2)} \forall i$, in alternativa si può dividere il budget in maniera proporzionale alla varianza negli scenari, ponendo ${}^{(i)}K_1^{(2)}$ pari al massimo intero minore di $\frac{(\Gamma - N_1 K_1^{(1)} - K_0) (\sigma_1^{(i)})^2}{\sum_{j \in I} (\sigma_1^{(j)})^2}$.

Se $K_1^{(2)}$ è significativamente maggiore di K_1 (numero di scenari interni non effettuando lo screening), allora avremo la possibilità di stimare l'SCR in maniera molto più precisa. Tuttavia, come abbiamo già anticipato, la procedura introduce nel modello una nuova fonte di errore, ossia la possibilità che alcuni tra gli scenari scartati facciano in realtà parte della coda. Ciò comporta un bias negativo per lo stimatore dell'SCR, potenzialmente non compensato dal bias positivo insito nelle simulazioni stocastiche annidate, e quindi il metodo potrebbe portare ad una sottostima, in media, dell'SCR. Anche per questo motivo, l'approccio basato sullo screening ha riscosso poco successo nell'implementazione pratica nel contesto assicurativo, ove si favorisce invece l'utilizzo di modelli proxy.

1.3 Modelli proxy

Abbiamo studiato nella sezione precedente l'approccio stocastico al calcolo del requisito di capitale e abbiamo visto che il modello delle simulazioni stocastiche annidate, pur adeguato da un punto di vista teorico per risolvere il problema, richiede uno sforzo computazionale non attuabile al giorno d'oggi. In particolare, il problema è dovuto alla necessità di un numero molto elevato sia di simulazioni esterne, sia di simulazioni interne per ogni simulazione esterna in modo da ottenere risultati affidabili.

La peculiarità dei modelli proxy è che anziché descrivere in via semplificata la realtà, come tutti i modelli, descrivono in via approssimata un modello più complesso, nel nostro caso quello delle simulazioni stocastiche annidate: i modelli proxy mantengono l'impostazione dell'approccio stocastico, attuando però delle semplificazioni in modo da superare i problemi computazionali. In particolare, l'obiettivo è eliminare la necessità di effettuare il ciclo interno di simulazioni, ottenendo una formula chiusa per la valutazione del portafoglio in ogni scenario, sulla base delle informazioni portate dagli scenari real-world.

Analizziamo nel seguito le tre principali tecniche sviluppate in questo ambito. Le tecniche del Curve Fitting e del Least-Squares Monte Carlo mirano ad ottenere il valore del portafoglio in ogni scenario tramite una funzione polinomiale nei risk factor, utilizzando tecniche statistiche sulla base di una prima fase di simulazioni annidate, attuata con un numero ridotto di scenari. La tecnica dei portafogli replicanti invece, anziché utilizzare funzioni astratte, si basa su strumenti finanziari preziosi tramite formule chiuse.

In tutti e tre i casi, avremo un metodo per valutare in via approssimata gli elementi del bilancio in ogni scenario real-world, eliminando il bisogno di un ciclo interno di simulazioni, e quindi riducendo in maniera significativa il costo computazionale necessario, permettendo anche l'utilizzo di un numero significativamente maggiore di scenari esterni. Naturalmente però questi metodi comportano una prima fase di calibrazione volta ad ottenere un'approssimazione funzionale del valore del portafoglio sufficientemente precisa. Inoltre, trattandosi comunque di tecniche di approssimazione, la funzione trovata dovrà essere sottoposta ad un opportuno processo di validazione prima di poter essere utilizzata per il calcolo dell'SCR.

1.3.1 Portafogli replicanti

Un portafoglio replicante è un insieme di attività finanziarie utilizzate per riprodurre approssimativamente i flussi di pagamento di un insieme di passività oppure di un altro insieme di attività. Infatti, se i cash flow di due portafogli coincidono esattamente, per la legge del prezzo unico dovranno coincidere anche i rispettivi valori di mercato. Nella pratica, riprodurre esattamente i flussi di un portafoglio relativamente complesso in ogni scenario non è pensabile, ma, ottenendo flussi sufficientemente simili, potenzialmente si può comunque conseguire una buona approssimazione del profilo di rischio del portafoglio.

Ne segue che, qualora si riuscisse a costruire un portafoglio replicante, il suo valore potrebbe fungere da proxy per il valore del portafoglio replicato. Se inoltre il portafoglio replicante è costituito da attività relativamente semplici da apprezzare, allora valutare tale portafoglio in un generico scenario permetterà di ottenere una valutazione (approssimata) market-consistent in maniera molto più rapida che procedendo con le simulazioni risk-neutral su, potenzialmente, decine di anni di run-off.

L'utilizzo di portafogli replicanti non è nuovo nella letteratura attuariale, essendo già stato introdotto a fini di hedging (si veda ad es. Pelsser (2003)): investendo in un portafoglio di attivi i cui flussi futuri replicano quelli delle proprie passività, l'assicuratore può infatti ridurre il rischio di non essere in grado di coprire i benefici aleatori che dovrà erogare in futuro. Mentre per l'hedging l'obiettivo è la riproduzione dei flussi nella maniera più fedele possibile, facendo utilizzo di strumenti finanziari anche complessi, un requisito fondamentale per utilizzare i portafogli replicanti come proxy del modello nested stochastic è la semplicità di valutazione degli strumenti utilizzati per la replicazione.

Gli strumenti più utilizzati in pratica sono Zero Coupon Bonds e Coupon Bonds per replicare i cash-flow costanti, Swaps e Swaptions per replicare le opzioni e le garanzie sui tassi d'interesse, Index Derivatives e Credit Derivatives per tenere conto rispettivamente delle esposizioni azionarie e del rischio di credito.

Una volta determinata la famiglia di strumenti finanziari da utilizzare per la replicazione, bisognerà determinare i coefficienti da associare ad ogni strumento in modo da formare il portafoglio. Ciò di regola avviene tramite metodi statistici

di regressione o di ottimizzazione con un approccio relativamente simile a quanto vedremo per il metodo LSMC.

Nell'applicazione del metodo dei portafogli replicanti, non vi è unicità d'opinione in letteratura. Le principali decisioni da prendere riguardano:

- La scelta tra l'utilizzo di un portafoglio replicante statico e più complessi metodi che consentono strategie dinamiche d'investimento.
- La scelta degli strumenti finanziari da utilizzare per la replicazione. Un numero troppo esiguo comporta scarsa capacità di cogliere con precisione il valore di passività complesse; un numero troppo elevato porta a problemi di overfitting.
- La scelta della misura di probabilità sotto alla quale effettuare il matching dei portafogli. Nella pratica e in gran parte della letteratura viene utilizzata la misura neutrale al rischio Q , ma vi è chi propende per la probabilità real-world P .
- La scelta della norma da utilizzare per determinare la misura di distanza. Scelte classiche sono la norma $L^{(1)}$ (somma dei valori assoluti delle componenti del vettore), la norma $L^{(2)}$ (norma euclidea) e la norma $L^{(2)}$ al quadrato.
- La scelta del criterio da considerare per risolvere il problema di matching, ossia della grandezza da minimizzare nell'algoritmo di ottimizzazione. Le proposte maggiormente implementate sono:
 - Il criterio del Cash-Flow Matching, che considera la somma (sulle epoche di pagamento) della differenza dei cash-flows attualizzati di ogni epoca, misurata tramite una delle tre norme sopra.
 - Il criterio del Terminal Value Matching, che trascura il tempismo dei flussi, considerando sufficiente che sia simile il loro valore aggregato a scadenza: si considera quindi come distanza la differenza tra i Terminal Values dei due portafogli.

Il problema di ottimizzazione del portafoglio replicante secondo il Cash-Flow Matching è di maggiore complessità, in quanto allo stesso tempo è necessario replicare l'importo e l'epoca del pagamento. Considerato anche che per la determinazione del rischio siamo interessati soltanto alla variazione nel valore di mercato del portafoglio, il criterio più utilizzato nella pratica è quello del Terminal Value Matching, per la sua confrontabilità con il modello LSMC e la relativa semplicità d'implementazione. Per entrambi i criteri sono state dimostrate, sotto opportune condizioni, proprietà di convergenza al modello nested stochastic, ma il Cash-Flow Matching è un metodo più robusto: anche se qui non vi è l'esigenza di replicare l'epoca dei flussi futuri, ma solo il valore complessivo, portafogli con troppi sfasamenti nelle epoche avranno anche, presumibilmente, dipendenze diverse dall'andamento dei fattori di rischio.

Possiamo impostare il problema della replicazione matematicamente a partire da Natolski & Werner (2014). Fissato l'orizzonte temporale $T \in \mathbb{N}$ e posto $\mathcal{T} = 0, 1, \dots, T$, consideriamo lo spazio di probabilità filtrato $(\Omega, \mathcal{A}, P, \mathcal{F})$ con le stesse ipotesi già introdotte nella sezione 1.2. In particolare, assumiamo che $\mathcal{A} = \mathcal{F}_T$ e che esistano la misura neutrale al rischio Q e il numerario $(B_t)_{t \in \mathcal{T}}$.

Indichiamo allora con $\mathbf{R}_t = (R_{1,t}, \dots, R_{m,t})$ i flussi di cassa (attualizzati in 0) corrisposti al tempo t dagli m strumenti finanziari da utilizzare per la replicazione e con L_t il flusso di cassa attualizzato in 0 derivante al tempo t dalle passività da replicare. Indichiamo ancora con $T_L = \sum_{t=1}^T L_t$ il Terminal Value attualizzato in 0 dei flussi di passività e con $\mathbf{T}_R = \sum_{t=1}^T \mathbf{R}_t$ il Terminal Value attualizzato degli strumenti presenti sul mercato.

Possiamo allora formulare il problema di ottimizzazione del portafoglio replicante secondo il criterio del Cash-Flow Matching e utilizzando la norma euclidea come:

$$\inf_{\boldsymbol{\alpha} \in \mathbb{R}^m} \sqrt{\sum_{t=1}^T \mathbb{E}_Q \left((L_t - \boldsymbol{\alpha}^T \mathbf{R}_t)^2 \mid \mathcal{F}_0 \right)} \quad (1.28)$$

Analogamente, il problema secondo il criterio del Terminal Value Matching, sempre utilizzando la norma euclidea, sarà:

$$\inf_{\boldsymbol{\alpha} \in \mathbb{R}^m} \sqrt{\mathbb{E}_Q \left((T_L - \boldsymbol{\alpha}^T \mathbf{T}_R)^2 \mid \mathcal{F}_0 \right)} \quad (1.29)$$

A partire da questi problemi di ottimizzazione si troverà la composizione ottimale del portafoglio replicante, α^* . Il problema in realtà è di notevole complessità, in quanto L_t è un numero aleatorio potenzialmente molto articolato e bisognerà far sì che il portafoglio replicante sia sufficientemente preciso in un numero il più elevato possibile di scenari. Il problema andrà risolto in via numerica.

Vista la semplicità di interpretazione di questo modello, già ampiamente sviluppato in finanza, e la vicinanza alle strategie di hedging, non sorprende che l'approccio basato sui portafogli replicanti per il calcolo dell'SCR sia quello più utilizzato dalla maggior parte dei grandi gruppi assicurativi europei anche per quanto concerne le passività in ambito vita.¹⁶

Tuttavia, l'approccio ha anche significative limitazioni: quella più immediata è data dalla difficoltà nel trovare un portafoglio costituito da elementi relativamente semplici da prezzare e allo stesso tempo in grado di replicare con sufficiente precisione il profilo di rischio di prodotti di assicurazione vita più complessi, in cui vi è, ad esempio, interazione tra attivi e passivi. La tecnica quindi richiede un profondo know-how finanziario e in ogni caso solitamente non ammette unica soluzione. La seconda grande limitazione del modello è la relativa incapacità nel cogliere rischi di natura non finanziaria¹⁷ e le interazioni tra rischi finanziari senza ricorrere a prodotti particolarmente complessi. Sarà pertanto necessario integrare il modello tramite altri metodi, il che apre a tutta una serie di problemi relativi alla calibrazione e alla coerenza tra le metodologie utilizzate.

Non tratteremo oltre la tematica dei portafogli replicanti. Per un'introduzione relativa all'implementazione pratica di questo modello al nostro contesto si consiglia Koursaris (2011), mentre per una trattazione matematica più approfondita del metodo si rimanda a Natolski & Werner (2014) e Beutner et al. (2016). Ancora, in Pelsser & Schweizer (2016) si studiano le relazioni tra la tecnica dei portafogli replicanti e il metodo Least-Squares Monte Carlo che tratteremo in seguito.

¹⁶La fonte per quanto affermato in questo paragrafo è DAV (2015), i dati sull'approccio utilizzato dai gruppi assicurativi sono relativi a fine 2014.

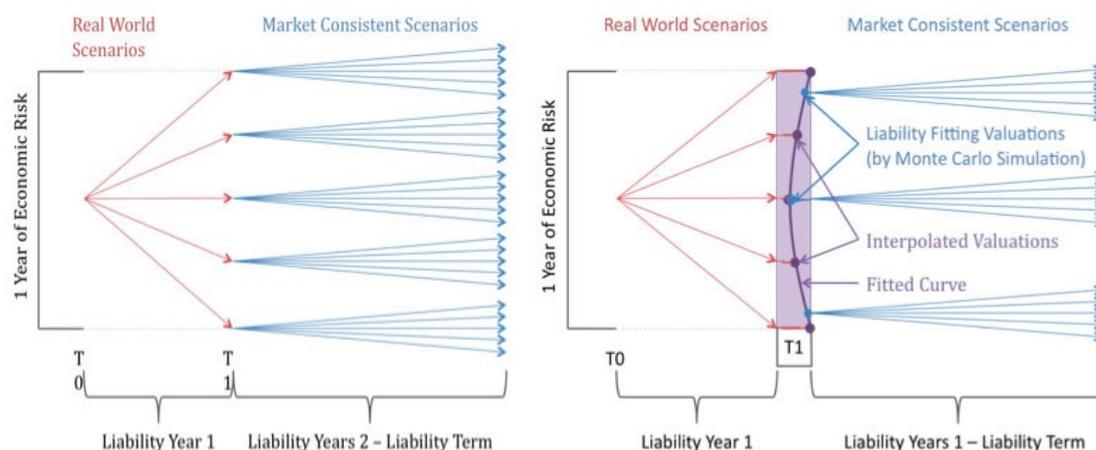
¹⁷Si parla dei rischi biometrici: anche se esistono in finanza strumenti come i mortality bonds, il loro mercato non è liquido e non sono semplici da prezzare, per cui non sono includibili nel portafoglio replicante. Per altri rischi, come quello di riscatto, il metodo risulta ancora più critico.

1.3.2 Curve fitting

Il metodo del curve fitting si propone di ottenere una proxy per il valore del portafoglio (o, equivalentemente, per la funzione di perdita ad esso associata) non tramite strumenti finanziari, come nel metodo dei portafogli replicanti, ma attraverso l'utilizzo di funzioni astratte, tipicamente polinomiali nei risk factor.

Mentre nel modello delle simulazioni stocastiche annidate otteniamo direttamente la distribuzione empirica del valore del portafoglio tramite un gran numero di scenari real-world e un gran numero di scenari interni per ogni scenario esterno, il curve fitting si contraddistingue per una prima fase, che come struttura ricorda le simulazioni stocastiche annidate, in cui però gli scenari esterni spesso non sono pseudo- o quasi-casuali, ma scelti deterministicamente; una seconda fase in cui a partire dai risultati ottenuti si ottengono i coefficienti della funzione di perdita (in funzione dei risk factor) tramite metodi di interpolazione o di regressione; una terza fase in cui, a partire da un gran numero di simulazioni real-world dei risk factor, si ottiene la stima della funzione di perdita procedendo con l'approccio stocastico classico, ma non vi sarà più bisogno del ciclo interno di simulazioni avendo ora una formula chiusa per valorizzare il portafoglio in ogni scenario.

L'obiettivo, quindi, rispetto alle simulazioni stocastiche annidate, è quello di utilizzare il modello annidato soltanto con un numero ridotto di scenari esterni, in modo da ridurre significativamente il costo computazionale richiesto, e a partire da ciò calibrare una funzione da utilizzare per valorizzare facilmente gli elementi del bilancio in un numero maggiore di scenari esterni. Il costo computazionale passa da $\Gamma_{NS} = K_0 + N K_1$ a $\Gamma_{CF} = K_0 + M K_1 + C(M, K_1)$ ove $C(M, K_1)$ è il costo relativo al fitting della proxy. Considerato ad esempio $C(M, K_1)$ trascurabile, $K_0 = 1'000$, $K_1 = 10'000$, $N = 200'000$, $M = 500$ si avrà $\Gamma_{NS} = 200'010'000$, mentre $\Gamma_{CF} = 510'000$. Nella realtà $C(M, K_1)$ può essere anche significativo, ma il risparmio in termini di scenari lo compensa facilmente, in particolare all'aumentare della dimensione del problema.



Rappresentazione schematica del metodo fully nested (a sinistra) e del metodo curve fitting (a destra). Fonte: Koursaris (2010)

Schematicamente, l'implementazione di un modello di stampo curve fitting può essere riassunta dai seguenti passi:

1. Identificazione dei risk factor che influenzano gli elementi del bilancio;
2. Determinazione di una distribuzione marginale (analiticamente o empiricamente su base di dati storici) per la variazione di ciascun risk factor in un anno;
3. Determinazione di una distribuzione congiunta della variazione dei risk factor a partire dalle marginali tramite copule;
4. Scelta di un numero ridotto di scenari relativi alla variazione ad un anno dei risk factor, giudicati rappresentativi di valori cruciali dei risk-driver. Ciò spesso avverrà scegliendo i valori d'interesse dei risk-driver presi singolarmente e poi considerando gli scenari risultanti dalle diverse loro combinazioni;
5. Rivalutazione degli elementi del bilancio in ognuno degli scenari trovati al punto precedente. Per attività e passività più complesse ciò sarà effettuato seguendo nuovamente un approccio di simulazione stocastica;
6. Determinazione dei coefficienti della funzione polinomiale che lega la variazione di valore del portafoglio alla variazione dei fattori di rischio;

7. Simulazione di un gran numero di scenari relativi alla variazione ad un anno dei risk factor tramite la distribuzione real-world;
8. Rivalutazione degli elementi del bilancio in ognuno degli scenari trovati al punto precedente. Ciò ora sarà possibile tramite una formula chiusa sulla base della funzione trovata al punto 6;
9. Calcolo della variazione di patrimonio netto in ogni scenario;
10. Determinazione dell'SCR come VaR al 99.5% della distribuzione empirica trovata al punto precedente.

Qualora negli M scenari esterni utilizzati per calibrare la funzione fossimo in grado di calcolare esattamente il valore del portafoglio in esame, potremmo creare un fit esatto interpolando con un polinomio di grado $M - 1$. Osservato però che anche con K_1 elevato per i prodotti assicurativi vita più complessi vi sarà ancora, potenzialmente, un notevole margine d'errore e che, volendo ottenere una funzione relativamente affidabile, M dovrà essere comunque relativamente elevato (nella pratica, spesso si considerano tra 100 e 1000 scenari), il polinomio interpolante non risulta una scelta adeguata. Anche altri metodi di interpolazione, quali l'interpolazione lineare o l'utilizzo di spline interpolanti risultano comunque poco adeguati. Si propende pertanto per tecniche di approssimazione in modo da ottenere una funzione più semplice e ridurre il problema dell'overfitting. In tale ambito, la scelta più comune è data dalla regressione polinomiale tramite il criterio dei minimi quadrati. In questo caso, quantomeno da un punto di vista formale, possiamo far rientrare la tecnica del curve fitting nell'impostazione del modello LSMC. Tratteremo meglio in seguito di queste somiglianze tra i due metodi.

La scelta più comune è l'utilizzo di una funzione di perdita con struttura polinomiale nei risk factor. Considerato il caso bidimensionale con fattori di rischio X_1 , X_2 , poniamo il valore del portafoglio $Y = f(X_1, X_2)$ pari a:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \beta_4 X_1^2 + \beta_5 X_2^2 + \dots \quad (1.30)$$

Oltre ai semplici fattori di rischio X_1 , X_2 vengono quindi considerate anche le potenze in modo da catturare gli effetti di non-linearità e i termini misti (*cross – terms*) in modo da catturare gli effetti di interazione tra i diversi fattori di rischio.

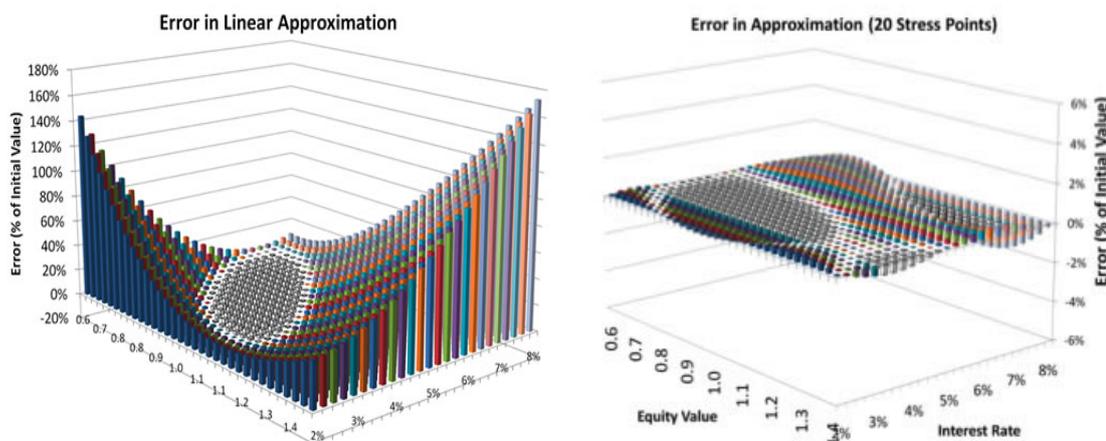
In un'applicazione pratica, spesso i fattori di rischio da considerare sono molteplici (potenzialmente anche diverse decine), per cui al crescere del grado massimo del polinomio lo sforzo computazionale necessario per stimare i coefficienti aumenta rapidamente, così come si incrementa anche il rischio di overfitting. Bisognerà quindi scegliere il grado massimo del polinomio valutando con cura il tradeoff tra semplicità e capacità di adattamento. Altre limitazioni possono essere imposte in aggiunta o, eventualmente, in sostituzione alla limitazione del grado massimo del polinomio; scelte comuni in tal senso sono limitare separatamente il grado massimo dei termini singoli e dei cross-terms, limitare il grado massimo di un singolo fattore di rischio all'interno dei cross-terms, limitare il numero massimo di termini del polinomio.

Bisognerà poi scegliere un criterio per determinare $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4 \dots)$. La struttura dell'equazione (1.30) ricorda fortemente l'ipotesi di linearità della funzione di regressione alla base del modello di regressione lineare. Potremmo allora pensare di utilizzare lo stimatore dei minimi quadrati:

$$\tilde{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \mathbf{Y} \quad (1.31)$$

ove come usuale abbiamo indicato con X la matrice di regressione (design matrix) in cui ogni riga si riferisce ad una simulazione e ogni colonna ad un singolo regressore (per semplicità in questo contesto consideriamo come regressore la variabile relativa ad ogni addendo della funzione di regressione, quindi X_1 e X_1^2 sono visti come due regressori diversi). Un'analisi più accurata della struttura della funzione di regressione, delle ipotesi alla base del modello, e del problema di selezione dei regressori sarà effettuata in seguito nella trattazione relativa al modello LSMC nella sezione 2.1.

Riprendiamo infine l'esempio di valutazione di un'opzione put tratto da Koursaris (2010) ed esposto nella sezione 1.1.4.



Errore di approssimazione (in termini percentuali rispetto al valore iniziale dell'opzione put), ottenuto utilizzando l'approssimazione lineare alla base della standard formula (a sinistra) e il metodo del curve fitting (a destra). Fonte: Koursaris (2010)

Notiamo dal grafico sopra che anche su un titolo così semplice si riesce ad ottenere una riduzione enorme dell'errore rispetto a quello commesso dalla standard formula. Va osservato che nel grafico sopra le valutazioni ottenute a partire dalla standard formula e dal curve fitting sono comparate con i risultati del modello di simulazioni stocastiche annidate. Il modello proxy, essendo approssimazione di un altro modello (c.d. heavy model), funziona bene solo se funziona bene il modello di partenza; nel caso in esame dovrà funzionare bene il modello di simulazioni stocastiche annidate, cioè dovranno essere affidabili i modelli attuariali o di asset pricing utilizzati per la determinazione del valore del portafoglio in ogni simulazione del ciclo interno.

Il curve fitting è un metodo computazionalmente implementabile e generalmente in grado di ottenere buoni risultati, ma soffre dell'elevato grado di arbitrarietà nella scelta degli scenari utilizzati per calibrare la funzione, che potenzialmente possono impattare in maniera estremamente significativa sul risultato. Il metodo, infatti, tende a non essere molto robusto, e se gli scenari di training falliscono nell'avvicinarsi al VaR al 99.5%, l'SCR ottenuto può essere significativamente distorto. Inoltre, un numero significativo di risk factor (come spesso accade nella pratica), può far sorgere nuovamente problemi computazionali: per catturare in maniera precisa l'effetto congiunto dei fattori di rischio, infatti, servirà un numero cospicuo

cuo di scenari, che cresce significativamente con la dimensionalità del problema. Il metodo nella sua forma generale non è stato ampiamente studiato e non sono presenti in letteratura molti risultati teorici. Ciò è dovuto all'estrema generalità della tecnica, che rappresenta un approccio (o, meglio, un insieme di approcci) estremamente pratico, studiato appositamente per approssimare il risultato delle simulazioni stocastiche annidate con un budget computazionale significativamente inferiore. Come abbiamo anticipato, il modello sotto opportune ipotesi può essere ricondotto ad un caso particolare di applicazione della tecnica Least-Squares Monte Carlo, che analizzeremo nel prossimo capitolo e che garantisce una base teorica più affidabile per risolvere il problema del computo stocastico dell'SCR.

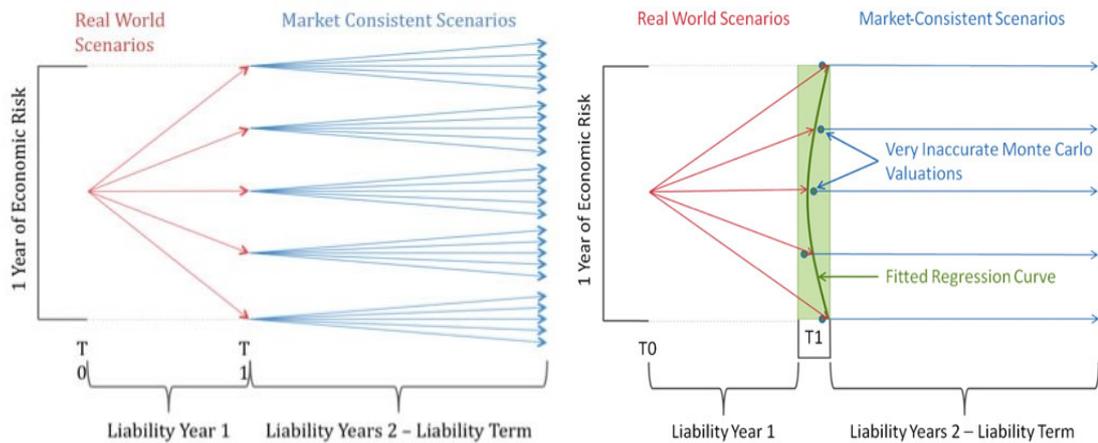
Per testare l'affidabilità della funzione di perdita trovata mediante il curve fitting, sarà necessario un adeguato sistema di validazione mediante scenari fuori campione (out-of-sample scenarios) in cui calcolare il valore del portafoglio mediante un ciclo interno di simulazioni tramite i modelli attuariali o di asset pricing e confrontarlo con il valore prodotto dalla funzione approssimante utilizzata. Solo qualora l'errore commesso dall'approssimazione sia sufficientemente piccolo, il modello potrà essere giudicato adeguato. Il tema della validazione della funzione di perdita approssimata è comune anche al metodo LSMC e sarà trattato con maggiore dettaglio in seguito.

1.3.3 Least-Squares Monte Carlo

La tecnica Least-Squares Monte Carlo (brevemente LSMC) è stata introdotta da Longstaff & Schwartz (2001) come metodo di valutazione delle opzioni americane in finanza. Tale modello ha riscosso fin da subito un forte successo, e risalgono al 2006 le prime proposte di applicazione al problema delle simulazioni stocastiche annidate, proposte successivamente formalizzate da Bauer (e altri) in una serie di articoli tra il 2008 e il 2012.

Il metodo LSMC segue la stessa logica del curve fitting, ponendosi l'obiettivo di utilizzare una funzione astratta, tipicamente polinomiale nei risk factor, per calcolare in via semplificata il valore del portafoglio in ogni scenario, e quindi ridurre lo sforzo computazionale necessario per ottenere una distribuzione empirica della funzione di perdita in un gran numero di simulazioni. Rispetto al curve

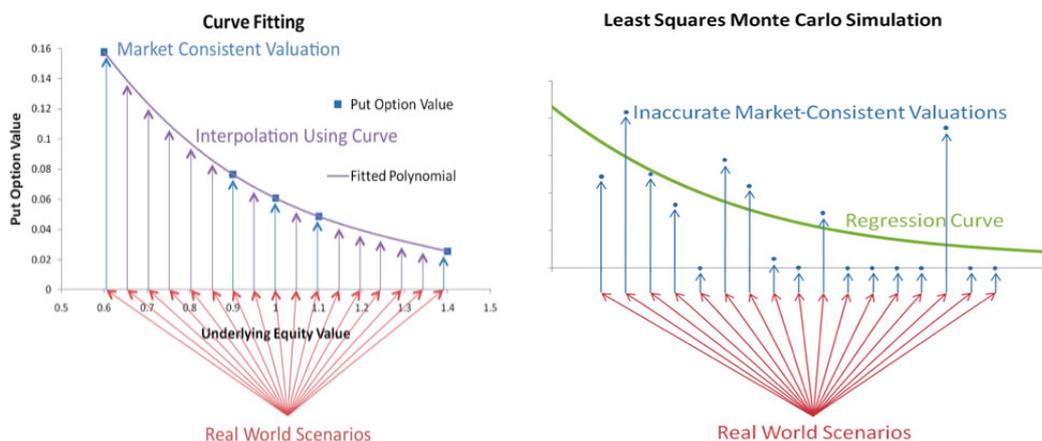
fitting, però, il modello LSMC poggia su più solide basi teoriche, che assicurano la maggiore affidabilità dei risultati.



Rappresentazione schematica del metodo fully nested (a sinistra) e del metodo LSMC (a destra). Fonte: Koursaris (2010)

Come accadeva per il curve fitting, il metodo LSMC si basa su una prima fase di simulazioni annidate utilizzate per calibrare la forma funzionale della funzione di perdita e una seconda fase in cui la funzione trovata viene valutata in un gran numero di scenari real-world. L'applicazione dei due metodi coincide nella seconda fase, mentre nella prima vi sono alcune differenze:

- Il curve fitting contempla, solitamente, la possibilità di scelta degli scenari esterni da utilizzare per la calibrazione, mentre il metodo LSMC si basa esclusivamente su numeri pseudo- o quasi-casuali.
- È diversa l'allocazione delle risorse computazionali: nel curve fitting si ottiene una valutazione accurata in un numero ridotto di scenari esterni, mentre il metodo LSMC si propone una valutazione più approssimata (pochissimi scenari interni) in un gran numero di scenari.
- Il curve fitting nasce con un obiettivo di interpolazione, mentre il metodo LSMC si basa esclusivamente su metodi di regressione.



Confronto tra la logica del curve fitting e quella del modello Least-Squares Monte Carlo. Fonte: Koursaris (2010)

Anche se il curve fitting nasce con l'obiettivo dell'interpolazione, abbiamo visto nella sezione precedente che è molto comune calibrare la funzione tramite metodi di regressione. Lasciando da parte, per il momento, il problema della determinazione degli scenari esterni (che con il curve fitting può avvenire in maniera più arbitraria), la vera differenza tra i due metodi si basa esclusivamente su una diversa allocazione del budget computazionale. Ne segue che l'impostazione teorica che vedremo in seguito potrà - in generale - essere estesa anche al curve fitting, tenendo però conto che le scelte arbitrarie che esso consente potrebbero invalidare i risultati ottenuti. Nel modello LSMC per calibrare la funzione proxy non si simula dalla distribuzione congiunta real-world dei risk factor, ma si suppone che la distribuzione sia uniforme sul supporto¹⁸: dato che gli scenari generati non rientrano in quelli utilizzati in via diretta per il calcolo dell'SCR, ma servono unicamente alla calibrazione della funzione di regressione, non è d'interesse rappresentare fedelmente la realtà, ma piuttosto coprire uniformemente lo spazio, in particolare per quanto riguarda la coda della distribuzione real-world, sulla quale altrimenti avremmo poche ben poche informazioni e per la quale è particolarmente importante che l'approssimazione

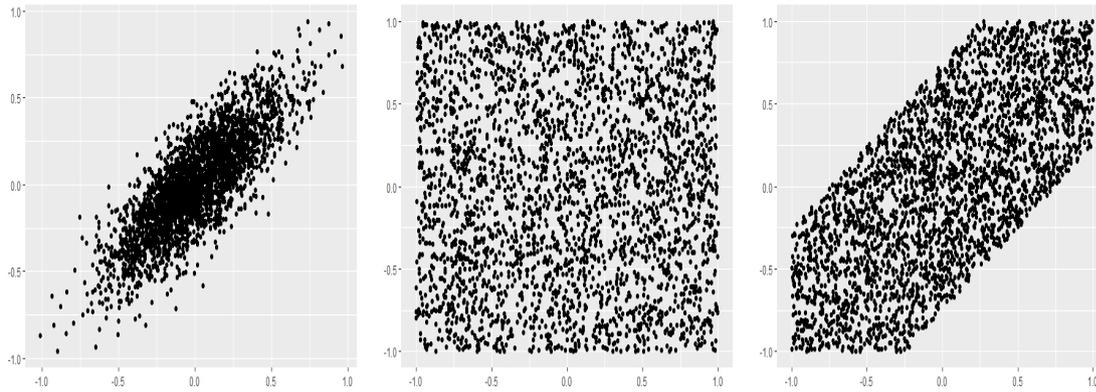
¹⁸In realtà, spesso il modello per la distribuzione della variazione di un risk factor avrà supporto pari, ad es., all'intera retta reale. Quando consideriamo la distribuzione uniforme del supporto, allora, considereremo in realtà la distribuzione uniforme su un intervallo sufficientemente ampio da contenere valori estremi, ma plausibili, dei risk factor. Si potrebbe pensare di prendere come supporto della distribuzione uniforme, ad esempio, l'intervallo dal quantile 0.01% al quantile 99.9% della distribuzione di ogni fattore di rischio. Il supporto considerato per la calibrazione sarà poi il range di validità della funzione proxy.

tramite la funzione proxy sia precisa. Per aumentare la regolarità, si potrebbero utilizzare sequenze a bassa discrepanza come le sequenze di Sobol introdotte nella sezione 1.2.3.c anziché numeri pseudocasuali.

Il problema della scelta tra distribuzione uniforme e real-world in questo contesto è stato studiato nel dettaglio da Cathcart (2012). Egli in un esperimento ottiene che il Mean Absolute Error (MAE) su tutto il set di scenari utilizzando la distribuzione real-world è leggermente inferiore che con la distribuzione uniforme (circa 10%), ma considerando invece il MAE su un set ridotto di scenari nella regione vicina all'SCR (in particolare, Cathcart considera una regione di 400 scenari centrati sul quantile 99.5% rispetto ai 50'000 scenari complessivi), si ha che l'utilizzo della distribuzione real-world conduce ad un MAE significativamente maggiore (circa +250%).

Un ulteriore vantaggio dell'utilizzo della distribuzione uniforme anziché quella real-world per la calibrazione della funzione proxy è dato dal fatto che in questo modo la proxy non dipende dalla distribuzione real-world dei risk factor, ma solo dal legame che essi hanno con la funzione di perdita. Allora, in seguito ad una variazione nell'economia che porta ad un cambiamento nella distribuzione real-world, se volessimo ristimare l'SCR potremmo utilizzare la proxy trovata al punto precedente e quindi ottenere il risultato con uno sforzo computazionale minimo (purché la variazione nell'economia sia sufficientemente contenuta per cui a cambiare sono solo le probabilità degli stati possibili dei risk factor e non il supporto utilizzato per la distribuzione uniforme).

Sono state esplorate altre strade oltre alla distribuzione uniforme e all'uso di sequenze a bassa discrepanza per la distribuzione degli scenari utilizzati per il fitting della proxy. In particolare, considerato l'obiettivo dell'analisi, ossia la determinazione dell'SCR, vi sono state alcune proposte per una collocazione mirata degli scenari in modo da coprire più pesantemente regioni di importanza primaria per l'SCR. Una scelta di questo tipo, tuttavia, oltre ad introdurre un certo grado di arbitrarietà nel modello, richiede conoscenze preliminari sulle dinamiche di rischio e limita la validità globale della funzione trovata, per cui non sarà da noi ulteriormente considerata.



Diversi metodi per la selezione degli scenari esterni nel fitting della proxy: distribuzione real-world (a sinistra), distribuzione uniforme su tutto lo spazio (al centro), distribuzione uniforme su valori plausibili (a destra).

Schematicamente, l'implementazione di un modello di stampo Least-Squares Monte Carlo può essere riassunta dai seguenti passi:

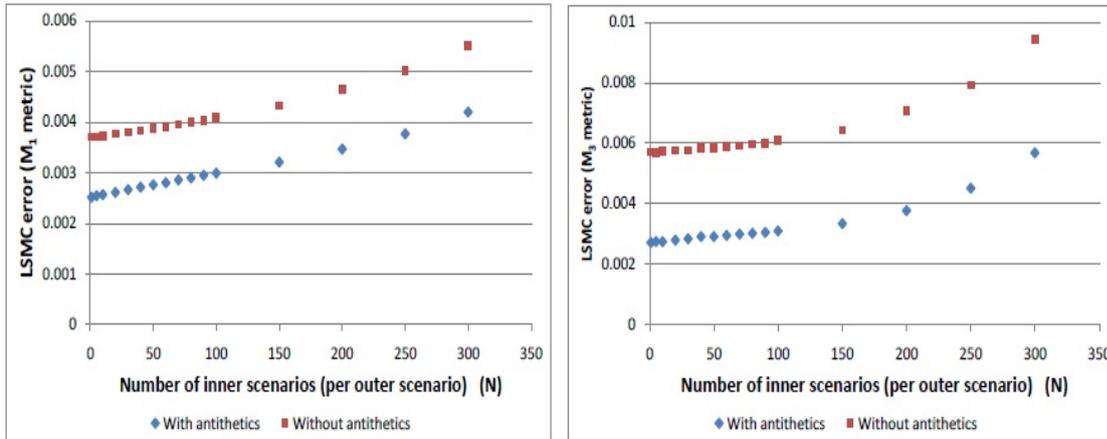
1. Identificazione dei risk factor che influenzano gli elementi del bilancio;
2. Determinazione di una distribuzione marginale (analiticamente o empiricamente su base di dati storici) per la variazione di ciascun risk factor in un anno;
3. Determinazione di una distribuzione congiunta della variazione dei risk factor a partire dalle marginali tramite copule;
4. Simulazione di un numero elevato di scenari relativi alla variazione dei risk factor, ipotizzando distribuzione uniforme su tutti i valori plausibili;
5. Rivalutazione approssimativa degli elementi del bilancio in ognuno degli scenari trovati al punto precedente. Per attività e passività più complesse ciò sarà effettuato seguendo nuovamente un approccio di simulazione stocastica, ma dato che ci accontentiamo di una valutazione approssimata saranno sufficienti poche simulazioni;
6. Applicazione del modello di regressione lineare (polinomiale nei risk factor) ai dati trovati ai punti 4 e 5 in modo da ottenere una stima del vettore dei coefficienti;

7. Simulazione di un gran numero di scenari relativi alla variazione ad un anno dei risk factor tramite la distribuzione real-world;
8. Rivalutazione degli elementi del bilancio in ognuno degli scenari trovati al punto precedente. Ciò ora sarà possibile tramite una formula chiusa sulla base della funzione di regressione trovata al punto 6;
9. Calcolo della variazione di patrimonio netto in ogni scenario;
10. Determinazione dell'SCR come VaR al 99.5% della distribuzione empirica trovata al punto precedente.

Osserviamo che rispetto al metodo curve fitting, qui abbiamo aumentato il numero di scenari esterni generati al punto 4 e utilizzati per la fase di calibrazione della funzione di regressione. Pertanto, a parità di budget computazionale, in ogni scenario esterno potremo considerare soltanto pochi scenari risk neutral. Per visualizzare le differenze tra i due metodi, consideriamo ad esempio un budget computazionale complessivo di 50'000 scenari. Con il curve fitting si potrebbe pensare di utilizzare 100 scenari esterni e 500 scenari interni per ogni scenario esterno; con il metodo LSMC 25'000 scenari esterni e, per ognuno, 2 scenari interni.

Le prime proposte di applicazione del metodo LSMC al contesto del risk management assicurativo contemplavano un singolo scenario risk neutral per ogni scenario esterno, al giorno d'oggi la configurazione più adottata prevede due scenari interni per ogni scenario esterno. Inoltre, per ridurre la varianza dell'errore, solitamente i due scenari sono presi antitetici (per dettagli si veda la sezione 1.2.3.d).

Anche il problema della configurazione ottimale degli scenari è stato analizzato da Cathcart (2012) nell'ambito della valutazione di un'opzione put. Egli, fissato il budget computazionale, studia l'andamento del MAE complessivo (M_1) e del MAE sulla regione vicina all'SCR (M_3) al variare del numero di scenari interni utilizzato. Come desumibile dalla figura sotto, la configurazione migliore nell'esperimento di Cathcart in entrambi i casi è data da una singola coppia di scenari antitetici.



MAE complessivo (a sinistra) e sulla regione vicina all'SCR (a destra) al variare del numero di scenari interni utilizzato, fissato il budget computazionale complessivo pari a 10'000. Fonte: Cathcart (2012)

Come abbiamo già accennato per gli altri metodi, una volta determinata la funzione proxy (la cui forma funzionale seguirà ancora la struttura delineata alla formula (1.30)) al punto 6, prima di procedere ad utilizzarla, saranno necessarie delle opportune analisi volte a validare la funzione di perdita trovata testandone l'affidabilità su un set di scenari (validation scenarios), diversi da quelli utilizzati per calibrarla (fitting scenarios).

Ci siamo limitati, in questa sede, a fornire una descrizione intuitiva del metodo LSMC e a fornire alcuni risultati in merito all'impostazione degli scenari da utilizzare per il fitting della proxy, concludendo che, per il problema in esame, è consigliato simulare dalla distribuzione uniforme (o da una sequenza di Sobol) anziché dalla distribuzione real-world e che l'allocazione migliore del budget computazionale sembra prevedere una singola coppia antitetica di scenari interni per ogni scenario esterno. Rimandiamo invece al capitolo 2 lo studio dell'impostazione teorica alla base del modello di regressione e delle principali proprietà, del tema della validazione della funzione proxy, nonché un'analisi dei punti critici e alcune soluzioni per l'implementazione.

1.4 I modelli a confronto

Nel corso di questo capitolo abbiamo analizzato i principali metodi attualmente presenti in letteratura per risolvere il problema del calcolo del capitale di solvibilità. Ricapitoliamo brevemente i principali risultati:

- Il modello nested stochastic è il più accurato tra i metodi proposti, tuttavia richiede uno sforzo eccessivo in termini di tempo e di costo computazionale per fornire risultati adeguatamente precisi, infatti i risultati sono significativamente distorti se il budget in termini di scenari è troppo limitato. Inoltre, ogni forma di prognosi, sensitività, o altra analisi richiede nuovamente un giro completo del modello, con tutto il costo computazionale ad esso annesso. Nella pratica, quindi, il modello ha riscontrato scarsa implementabilità.
- La standard formula di Solvency II può essere vista come il più semplice tra i modelli proxy: al ciclo esterno di simulazioni stocastiche si sostituiscono valutazioni deterministiche in specifici scenari, alla ricerca della distribuzione della perdita la semplice ricerca del quantile; la struttura di aggregazione inoltre assume in via semplificata la linearità della funzione di perdita e la distribuzione ellittica dei rischi, oltre che basarsi su una matrice di correlazione predefinita. Abbiamo già analizzato nella sezione 1.1.4 la sostanziale inadeguatezza della standard formula a descrivere strutture di rischio più complesse quali possono essere quelle di imprese di assicurazione vita. Ciò nonostante, la standard formula rimane il metodo più utilizzato tra le piccole e medie imprese di assicurazione, che spesso non dispongono delle risorse necessarie per costruire un modello interno.
- I modelli proxy mirano ad approssimare il procedimento di simulazioni stocastiche annidate, stimando il valore del portafoglio nell' i -esimo scenario real-world, Y_i , tramite una funzione dei risk factor $\tilde{Y}_i = f(X_{i,1}, \dots, X_{i,n})$. Tali modelli comportano una significativa riduzione del tempo di esecuzione rispetto al modello nested stochastic, e allo stesso tempo superano le stringenti limitazioni della standard formula. In particolare:

- Il metodo dei replicating portfolios considera una struttura di tipo $\tilde{Y}_i = \sum_{k=1}^K \alpha_k g_k(X_{i,1}, \dots, X_{i,n})$, ove g_k rappresenta il prezzo del k -esimo strumento replicante in funzione dei risk factor. Il metodo è tuttora il più utilizzato per la costruzione di un modello interno, presumibilmente per la sua semplicità d'interpretazione e per la vicinanza con i modelli più classici della finanza, tuttavia la sua applicazione incontra diverse limitazioni nel nostro contesto: da un lato la difficoltà ad ottenere una buona replicazione di complessi prodotti assicurativi vita tramite strumenti sufficientemente semplici da apprezzare, dall'altro il rischio di arbitrarietà nella scelta e nella calibrazione del portafoglio, oltre alla difficoltà di tale modello nel tenere conto di rischi non finanziari e delle inter-relazioni tra rischi finanziari e non-finanziari.
- Il metodo del curve fitting stima il valore del portafoglio in ogni scenario tramite una funzione, tipicamente polinomiale nei risk factor. Il metodo può essere visto come una via di mezzo tra il modello nested stochastic e la standard formula: come quest'ultima, si basa su specifici scenari di variazione dei fattori di rischio per calibrare la funzione di perdita, ma in quegli scenari calcola il valore del portafoglio in maniera molto più precisa, tramite modelli stocastici. Le limitazioni della standard formula sono superate: il metodo non necessita di particolari ipotesi distributive né impone di per sé limitazioni alla forma funzionale della funzione di perdita. Semplice da capire e da implementare, il metodo permette una pluralità di configurazioni possibili, e, anche a causa di tale generalità, oltre che per il rischio di arbitrarietà nella scelta degli scenari da utilizzare, non possiede solide basi teoriche che ne assicurino la bontà.
- Il metodo Least-squares Monte Carlo riprende la logica di fondo del curve fitting, ma si basa su un'idea diversa: anziché effettuare una valutazione precisa in un numero ristretto di scenari prescelti, si effettua una valutazione approssimata in un gran numero di scenari generati (quasi-)casualmente e si deriva la funzione di perdita tramite tecniche di regressione. Il metodo gode di buone proprietà statistiche, che analiz-

zeremo nel prossimo capitolo, nonché riduce la necessità di ricorrere ad arbitrari expert judgements e consente una maggiore automatizzazione della procedura di calibrazione. Rispetto al curve fitting, inoltre, il metodo LSMC è computazionalmente più efficiente e consente di ricavare facilmente intervalli di confidenza e misure statistiche oggettive per la bontà di adattamento. Il metodo si è affermato negli ultimi anni come best practice per la valutazione delle passività vita in un approccio stocastico al calcolo dell'SCR¹⁹.

Tra i vari metodi analizzati, nel seguito ci concentreremo esclusivamente sul LSMC. Riteniamo infatti che le limitazioni imposte dall'elegante soluzione dei portafogli replicanti siano significative, e rispetto al curve fitting, il metodo LSMC è più efficiente, meno arbitrario e gode di solide basi matematiche. Ciò non toglie che, attuate le dovute precauzioni, essendo l'output delle due procedure lo stesso, molti dei temi che saranno trattati guardando al modello LSMC - come le tecniche di validazione e la selezione dei regressori - potranno essere traslati pari pari anche nel contesto del curve fitting e, talvolta, anche alle applicazioni relative ai portafogli replicanti. Peraltro, i metodi sono tra loro consistenti, per cui si potrebbe pensare di utilizzare il metodo LSMC per alcune categorie di elementi del bilancio (ad es. le passività in ambito vita), il curve fitting per altre. Inoltre, alcune imprese hanno implementato forme ibride, avvicinando gli esistenti modelli di curve fitting al metodo LSMC, ad esempio utilizzando l'algoritmo dei minimi quadrati per trovare i coefficienti della funzione proxy o aggiungendo scenari esterni stocastici alla procedura di calibrazione. In tal caso, il curve fitting può essere visto come una configurazione - meno efficiente - del metodo LSMC classico.

¹⁹Per la valutazione degli attivi, invece, solitamente vi sono minori difficoltà nel prezzamento in scenari stressati, per cui l'utilizzo di un heavy model è considerato teoricamente attuabile ed eviterebbe di ricorrere all'errore di approssimazione dovuto all'utilizzo di modelli proxy, i quali, tuttavia, nella pratica spesso sono utilizzati lo stesso al fine di semplificare il modello, ridurre i tempi di calcolo e garantire coerenza con l'approccio utilizzato per la modellizzazione dei passivi. Risulta quindi sensato lo studio di un modello di stampo LSMC anche per gli attivi, tenendo comunque in considerazione che per una gran parte delle attività in realtà non sarà necessario procedere alla generazione di scenari interni tramite misure neutrali al rischio: in tal caso, la valutazione effettuata in ogni scenario sarà molto più precisa e, pertanto, saranno necessari molti meno scenari esterni per ottenere un fit adeguato: ci avviciniamo, in sostanza, ad una forma di modello LSMC che ricorda, per l'impostazione, più la logica del curve fitting.

Capitolo 2

Il modello Least-Squares Monte Carlo

Nel corso di questo capitolo si analizzerà il modello LSMC nel contesto della determinazione del requisito di capitale di solvibilità richiesto alle imprese operanti nel settore assicurativo.

Abbiamo già introdotto la tecnica LSMC nella sezione 1.3.3, nonché riportato alcune analisi relative alla configurazione ottimale degli scenari da utilizzare per la calibrazione, osservando che, come studiato da Cathcart (2012), risulta opportuno considerare una distribuzione uniforme per la generazione degli scenari esterni e utilizzare due scenari interni antitetici per ogni scenario esterno. Nella prima parte di questo capitolo studieremo con maggiore dettaglio il modello LSMC, approfondendone gli aspetti teorici e le proprietà di convergenza. Tratteremo in particolare il problema della scelta dei regressori e il tema della validazione della funzione di perdita trovata. Successivamente ci soffermeremo sulle principali estensioni del modello che sono state finora proposte in letteratura.

Nell'ultima parte del capitolo, infine, applicheremo il modello in un contesto assicurativo, analizzando in particolare l'effetto di diverse configurazioni sull'errore generato dall'utilizzo delle proxy.

2.1 Costruzione di un modello LSMC

Nella sezione 1.3.3 si è fornita un'introduzione discorsiva alla tecnica LSMC. Andremo ora a studiare più in dettaglio alcuni aspetti teorici relativi all'approccio. In particolare, considereremo come dati gli scenari di fitting e il valore ottenuto in ciascuno degli scenari interni corrispondenti, e analizzeremo il procedimento di costruzione e validazione della funzione proxy.

2.1.1 L'impostazione del metodo

Fissato un portafoglio, che per comodità penseremo essere costituito da passività derivanti da prodotti sufficientemente omogenei¹ di assicurazione vita (ma potrebbe anche essere un portafoglio di attivi o di passivi non-vita), consideriamo i d fattori di rischio che si ritiene possano impattare sul suo valore e generiamo N scenari esterni rappresentanti traiettorie monoannuali² dei fattori di rischio. Avremo dunque i vettori d -dimensionali $\mathbf{Z}_1, \dots, \mathbf{Z}_N$, rappresentanti i valori assunti dai fattori di rischio negli N scenari. Consideriamo poi la funzione obiettivo p che rappresenta il valore del portafoglio (ossia, più precisamente, il valore dei discounted cash flows al tempo $t = 1$ degli impegni relativi alle polizze del portafoglio), che potenzialmente dipenderà dal valore assunto dai risk factor su tutta la durata di run-off. Un primo obiettivo sarà la stima di $V^{(i)} = \mathbb{E}(p | \mathbf{Z}_i)$, alla quale procederemo tramite un processo Monte Carlo.

Applichiamo allora un ciclo interno di K_1 simulazioni in ottica di run-off a ciascuno degli N scenari esterni e otteniamo il valore del portafoglio al tempo 1, $p^{(i,j)}$ per $i = 1, \dots, N, j = 1, \dots, K_1$. Coerentemente con quanto affermato nella sezione 1.3.3.

¹Non parleremo troppo, in questo lavoro, della granularità alla quale sia opportuno utilizzare l'approccio, in quanto dipenderà principalmente dalla struttura della singola compagnia, tuttavia riteniamo opportuno considerare una granularità non meno fine della singola line of business, in modo da avere un effetto coerente dei fattori di rischio su tutte le polizze del portafoglio; potenzialmente, in ogni caso, il metodo funziona anche considerando un unico portafoglio per tutti gli impegni dell'assicuratore. Ovviamente, all'aumentare della granularità utilizzata, si renderà necessaria l'implementazione di opportune tecniche di aggregazione per ottenere la perdita complessiva della compagnia.

²Spesso, nella pratica, si considera una semplificazione del modello qui presentato, utilizzando, anziché traiettorie monoannuali, stress istantanei che rappresentano la traiettoria monoannuale attualizzata. Tale variazione non dovrebbe avere un impatto significativo sul modello, ma in letteratura non sono presenti numerose analisi in merito.

vorremo avere K_1 relativamente piccolo, in particolare terremo in considerazione i valori $K_1 = 1$ e $K_1 = 2$.

Dobbiamo ora stimare il valore del portafoglio in ciascuno degli N scenari esterni. A tal fine, considereremo semplicemente la media aritmetica sugli scenari interni:

$$\tilde{V}^{(i)} = \frac{1}{K_1} \sum_{j=1}^{K_1} p^{(i,j)} \quad (2.1)$$

Ovviamente, nel caso $K_1 = 1$, prenderemo come $\tilde{V}^{(i)}$ il valore dell'unico scenario disponibile. Per una trattazione più formale degli aspetti relativi alla determinazione di $\tilde{V}^{(i)}$ rimandiamo alla sezione 1.2.2, in cui abbiamo delineato con maggiore dettaglio questo processo nell'ambito del modello stocastico annidato.

Abbiamo quindi N scenari in cui sono note le determinazioni dei risk factor \mathbf{Z}_i e in cui possiamo pensare di aver osservato il valore del portafoglio $V^{(i)}$. Notiamo che lo stimatore Monte Carlo necessita di molte simulazioni per fornire risultati relativamente precisi, e che per vincoli computazionali abbiamo scelto K_1 estremamente ridotto. Sarà allora necessario un procedimento che riesca a trattare l'incertezza dovuta alla stima di $\tilde{V}^{(i)}$, estrapolando l'informazione contenuta nella stima e separandola dal rumore statistico. A tal fine, utilizzeremo il modello di regressione lineare.

Il modello di regressione lineare è un modello di stima della funzione di regressione³ di una variabile Y rispetto alle variabili X_1, \dots, X_m , con l'obiettivo, quindi, di modellare la relazione funzionale che intercorre tra la variabile dipendente Y e i regressori X_1, \dots, X_m a partire dall'osservazione di N dati, che indichiamo con $(y_i, x_{i,1}, \dots, x_{i,m})$ per $i = 1, \dots, N$, con la condizione $N \leq m$.

Il modello si basa sulle seguenti ipotesi:

1. $\mathbb{E}(Y_i | X_{i,1}, \dots, X_{i,m}) = \beta_0 + \beta_1 X_{i,1} + \dots + \beta_m X_{i,m} \quad \forall i$ con $\boldsymbol{\beta} = (\beta_0, \dots, \beta_m) \in \mathbb{R}^{m+1}$, ossia ipotizziamo la linearità della funzione di regressione rispetto ai coefficienti $\boldsymbol{\beta}$.

³In questa trattazione chiameremo funzione di regressione di Y su X la variabile aleatoria $\mathbb{E}(Y|X)$. In molti testi, per assonanza con il corrispondente termine inglese *conditional expectation*, tale quantità viene chiamata semplicemente speranza matematica condizionata, sottintendendo che si tratta di un condizionamento alla σ -algebra indotta dalla variabile aleatoria X , e non di un condizionamento ad un evento.

2. $(Y_i, X_{i,1}, \dots, X_{i,m})$ e $(Y_j, X_{j,1}, \dots, X_{j,m})$ siano stocasticamente indipendenti e identicamente distribuiti per $i \neq j$.
3. $X = (X_{i,j})_{i=1, \dots, N, j=1, \dots, m}$ sia una matrice a rango pieno (cioè con colonne linearmente indipendenti), ossia ipotizziamo che non vi sia multicollinearità perfetta.
4. I momenti quarti di tutti i regressori siano finiti e non nulli.
5. $V(Y_i | X_{i,1}, \dots, X_{i,m}) = \sigma^2 \forall i$, ossia ipotizziamo valga l'omoschedasticità (condizionata).

Posto ancora $\mathbf{Y} = (Y_1, \dots, Y_N)$, e inserita l'intercetta nella design matrix X , possiamo scrivere il modello in forma matriciale come $\mathbb{E}(\mathbf{Y} | X) = X\boldsymbol{\beta}$ o equivalentemente $\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ con $\mathbb{E}(\boldsymbol{\varepsilon} | X) = 0$. Possiamo allora considerare lo stimatore dei minimi quadrati:

$$\tilde{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} (\mathbf{Y} - X\boldsymbol{\beta})^T (\mathbf{Y} - X\boldsymbol{\beta}) = (X^T X)^{-1} X^T \mathbf{Y} \quad (2.2)$$

Lo stimatore $\tilde{\boldsymbol{\beta}}$ sotto alle ipotesi 1.-4. è corretto, consistente e asintoticamente normale. Inoltre, aggiungendo l'ipotesi 5., per il Teorema di Gauss-Markov lo stimatore è anche BLUE, ossia quello con varianza minima tra tutti gli stimatori lineari corretti. Ancora, osserviamo che la presenza di regressori fortemente correlati tra loro, pur non invalidando la teoria sottostante al modello, potrebbe causare problemi tecnici nell'inversione della matrice $X^T X$. Il problema della multicollinearità sarà potenzialmente significativo nel nostro modello, considerato che stiamo considerando come regressori diverse potenze degli stessi fattori di rischio.

Un metodo agevole per calcolare lo stimatore dei minimi quadrati $\tilde{\boldsymbol{\beta}}$ in maniera da garantire stime più stabili si basa sulla decomposizione QR della matrice X , un risultato per il quale una matrice $N \times m$ a rango pieno X può essere scomposta in $X = QR$ ove Q è una matrice $N \times m$ ortogonale (quindi per cui $Q^T Q = I$) e R è una matrice $m \times m$ triangolare superiore. Ne segue che possiamo riscrivere l'equazione dei minimi quadrati come:

$$\begin{aligned} X^T X \boldsymbol{\beta} = X^T \mathbf{Y} &\Leftrightarrow (QR)^T (QR) \boldsymbol{\beta} = (QR)^T \mathbf{Y} \\ &\Leftrightarrow R^T (Q^T Q) R \boldsymbol{\beta} = R^T Q^T \mathbf{Y} \end{aligned}$$

da cui, per l'ortogonalità di Q , si ha:

$$R^T R\boldsymbol{\beta} = R^T Q^T \mathbf{Y}$$

E quindi l'equazione dei minimi quadrati diventa:

$$R\boldsymbol{\beta} = Q^T \mathbf{Y} \tag{2.3}$$

Essendo R triangolare, l'equazione (2.3) è significativamente più semplice da risolvere, e garantisce una stabilità molto maggiore rispetto alle stime ottenute risolvendo in via diretta l'equazione (2.2). Il metodo di soluzione dell'equazione dei minimi quadrati tramite il procedimento di decomposizione QR è, di norma, implementato anche dalla maggior parte dei moderni software statistici. Per una trattazione più formale relativa alla teoria sottostante a questa tecnica, così come per potenziali alternative si veda Björk (1996).

Nel nostro contesto, abbiamo indicato con $V^{(i)}$ la variabile dipendente Y_i ; inoltre, i regressori X_1, \dots, X_m sono funzioni deterministiche dei fattori di rischio Z_1, \dots, Z_d . In particolare, vedremo utilizzate funzioni polinomiali, in modo da tenere conto anche degli effetti di non-linearità e dei fattori di interazione. Spesso, come giustificazione teorica dell'uso di funzioni polinomiali in questo contesto, si cita il teorema di approssimazione di Weierstrass, per cui ogni funzione reale continua definita su un intervallo chiuso e limitato può essere approssimata uniformemente a piacere tramite un polinomio di grado opportuno. Nella letteratura, si dicono poi Main Effects i regressori contenenti al più un singolo fattore di rischio, Interaction Effects o Cross Terms i regressori contenenti due o più fattori.

Osserviamo in particolare che nella nostra applicazione del modello lineare, risulta particolarmente importante evitare il fenomeno dell'overfitting, in quanto, per come abbiamo impostato il modello LSMC, ogni osservazione, generata da soltanto pochissimi scenari interni, presenterà verosimilmente un notevole errore, per cui è necessario scongiurare che la funzione di perdita stimata presenti un adattamento troppo accentuato ai singoli dati utilizzati per la calibrazione. Sarà necessario quindi un adeguato procedimento di selezione dei regressori, che sia in grado di bilanciare la necessità di utilizzare tutti i termini necessari a spiegare gli effetti statisticamente significativi con quella di evitare problematiche legate all'overfit-

ting. Maggiori dettagli sulla struttura e un'analisi dei principali procedimenti di selezione dei regressori sarà fornita nell'ambito della sezione 2.1.3.

Considereremo, in realtà, una semplificazione del modello qui esposto, assumendo che i regressori non siano stocastici: ciò potrebbe, a primo impatto, apparire sorprendente, in quanto nell'approccio di simulazione stocastica considerato finora sono trattati come variabili aleatorie, e i relativi scenari generati casualmente. Qui tuttavia ci si riferisce unicamente al comportamento dei regressori all'interno del modello stocastico per la regressione. Tale scelta comporta che possiamo considerare le osservazioni relative ai regressori come valori fissati e costanti in tutti i calcoli relativi al modello. Ciò, in particolare, ha come conseguenza che l'ipotesi 4. del modello sarà superflua e che l'ipotesi 2. si riferirà soltanto agli Y_i e quindi non sarà più necessario limitarsi ad un campionamento casuale degli scenari esterni dalla distribuzione real-world, permettendo, tra l'altro, il campionamento da una distribuzione diversa, quale quella uniforme, e l'utilizzo di tecniche non prettamente casuali (quali le sequenze a bassa discrepanza) per la generazione degli scenari da utilizzare per la calibrazione. A conseguenza di ciò, tuttavia, la regressione non fornirà alcuna informazione sulla funzione di regressione di Y rispetto ai risk factor Z_1, \dots, Z_d , forzando quindi la necessità di effettuare successivamente una seconda fase di simulazioni, questa volta dalla distribuzione real-world, al fine di ottenere il VaR. Tale procedimento è coerente con quello esposto nella sezione 1.3.3.

Per quanto riguarda le altre ipotesi alla base del modello, osserviamo che la linearità della funzione di regressione non è molto restrittiva, in quanto essa è richiesta esclusivamente rispetto ai coefficienti; inoltre, l'indipendenza stocastica delle Y_i è garantita dal fatto che, nel modello da noi considerato, come usuale nei modelli di tipo LSMC, gli scenari neutrali al rischio generati in ogni scenario esterno sono tra loro indipendenti. Anche l'ipotesi 3 non risulta restrittiva (fermo restando quanto osservato sulla multicollinearità non perfetta), mentre, in generale, potrebbe esserci eteroschedasticità tra le osservazioni relative a diversi scenari. Fortunatamente, però, l'ipotesi di omoschedasticità non è fondamentale: anche in sua assenza, infatti, lo stimatore dei minimi quadrati è corretto e consistente.

2.1.2 Le basi teoriche dell'approccio

Riprendiamo ora l'approccio teorico del modello stocastico che abbiamo impostato nel capitolo precedente e cerchiamo di inquadrare entro tale approccio anche il metodo LSMC. I principali riferimenti per la trattazione seguente sono Bauer et al. (2009) e Ha & Bauer (2018). Osserviamo che l'impostazione considerata in tali articoli si basa su una calibrazione mediante un unico scenario interno per ogni scenario esterno; noi, invece, coerentemente con quanto esposto nella sezione 1.2.2, definiremo l'algoritmo generale permettendo un numero arbitrario di scenari interni per ogni scenario esterno.

Nella sezione 1.1.2 abbiamo definito il requisito di capitale di solvibilità come:

$$SCR = \arg \min_x \left\{ P \left(AC_0 - \frac{AC_1}{1 + r(0, 1)} > x \right) \leq 1 - \beta \right\} \quad (2.4)$$

Abbiamo osservato poi che le uniche grandezze aleatorie sono AC_0 e AC_1 , e nella sezione 1.2.2 abbiamo definito un metodo stocastico per la stima di tali quantità. Per quanto riguarda AC_0 , il processo di stima non comporta, in realtà, l'utilizzo di simulazioni annidate, per cui possiamo mantenere invariato il metodo esposto nella sezione 1.2.2. Per quanto riguarda $AC_1 = ANAV_1 + PVFP_1 + X_1$, abbiamo visto che $ANAV_1$ e X_1 possono essere ricavati a partire unicamente dal ciclo esterno monoannuale di simulazioni real-world, il quale viene eseguito anche dal modello LSMC una volta calibrate le funzioni proxy, per cui anche per queste quantità la metodologia esposta nella sezione 1.2.2 rimane invariata. L'unico elemento a richiedere il computazionalmente impegnativo ciclo di simulazioni stocastiche annidate è infatti $PVFP_1$.

Sempre nella sezione 1.2.2 abbiamo poi considerato lo spazio di probabilità filtrato $(\Omega, \mathcal{A}, P, \mathcal{F})$ e definito $PVFP_1$ tramite la formula di valutazione neutrale al rischio:

$$PVFP_1 = \mathbb{E}_Q \left[\sum_{t=2}^T X_t e^{-\int_1^t r_u du} \mid \mathbf{Y}_1, \mathbf{D}_1 \right] \quad (2.5)$$

ove \mathbf{Y}_1 rappresenta il valore del processo markoviano d -dimensionale utilizzato per modellare l'incertezza nel mercato e \mathbf{D}_1 è un vettore finito-dimensionale contenente tutte le informazioni relative all'evoluzione del mercato in $[0, 1]$ necessarie per la determinazione di $PVFP_1$ tramite la formula sopra.

Come avevamo osservato, la difficoltà nella stima della quantità nell'equazione (2.5) tramite metodi Monte Carlo risiede nella natura condizionata della speranza matematica, che costringe a ricorrere ad un procedimento annidato di simulazioni. Una difficoltà analoga è stata incontrata relativamente al problema di prezzamento delle opzioni bermuda e, in tale ambito, è stato inventato il metodo Least-Squares Monte Carlo (i cui dettagli si trovano in Longstaff & Schwartz (2001)), come alternativa computazionalmente più leggera al procedimento di simulazioni stocastiche annidate. Tale procedimento si basa su un'approssimazione della speranza matematica in (2.5) tramite una combinazione lineare di funzioni base (basis functions), i cui coefficienti saranno poi stimati tramite simulazioni Monte Carlo e tecniche di regressione. Seguendo l'impostazione di Bauer et al. (2009), applicheremo tale logica al contesto del calcolo stocastico dei requisiti di capitale.

Consideriamo H funzioni base linearmente indipendenti $g_h(\mathbf{Y}_1, \mathbf{D}_1)$, $h = 1, \dots, H$, in modo da poter scrivere approssimativamente la speranza matematica in (2.5) tramite una loro combinazione lineare:

$$PVFP_1 \approx f(\mathbf{Y}_1, \mathbf{D}_1) = \sum_{h=1}^H \beta_h g_h(\mathbf{Y}_1, \mathbf{D}_1) \quad (2.6)$$

Come nel modello stocastico annidato, generiamo poi N traiettorie per l'andamento del processo di stato nel primo anno, in modo da ottenere, in particolare, $(\mathbf{Y}_1^{(i)}, \mathbf{D}_1^{(i)})$ per $i = 1, \dots, N$. Utilizzando lo stesso procedimento delle simulazioni stocastiche annidate, per ogni scenario i simuliamo poi $K_1^{(i)}$ traiettorie del processo di stato su $[1, T]$ tramite la misura neutrale al rischio Q , a partire dalle quali possiamo ricavare determinazioni dei flussi di cassa e dei fattori di attualizzazione futuri, e quindi calcolare la determinazione simulata di $PVFP_1$ nello scenario i come:

$$\widetilde{PVFP}_1^{(i)} = \frac{1}{K_1^{(i)}} \sum_{k=1}^{K_1^{(i)}} \sum_{t=2}^T \tilde{X}_t^{(i,k)} e^{-\int_0^t \tilde{r}_u^{(i,k)} du} \quad (2.7)$$

A partire da queste realizzazioni, possiamo determinare i coefficienti $\boldsymbol{\beta} = (\beta_1, \dots, \beta_H)$ in (2.6) tramite il metodo dei minimi quadrati:

$$\tilde{\beta} = \arg \min_{\beta \in \mathbb{R}^H} \left\{ \sum_{i=1}^N \left(\widetilde{PVFP}_1^{(i)} - \sum_{h=1}^H \beta_h g_h(\mathbf{Y}_1^{(i)}, \mathbf{D}_1^{(i)}) \right)^2 \right\} \quad (2.8)$$

Otteniamo in questo modo la funzione proxy \tilde{f} , che dipenderà soltanto dai valori $\mathbf{Y}_1, \mathbf{D}_1$:

$$\tilde{f}(\mathbf{Y}_1, \mathbf{D}_1) = \sum_{h=1}^H \tilde{\beta}_h g_h(\mathbf{Y}_1, \mathbf{D}_1) \quad (2.9)$$

A questo punto, la metodologia proposta in Bauer et al. (2009) e Ha & Bauer (2018) prevede che si riciclino gli N scenari monoannuali $(\mathbf{Y}_1^{(i)}, \mathbf{D}_1^{(i)})$, nei quali tramite la (2.9) possiamo calcolare il valore della f , che utilizziamo per stimare $PVFP_1$ e quindi anche AC_1 e la funzione di perdita in ogni scenario. Dagli N valori trovati, avremo ottenuto una distribuzione empirica dalla quale possiamo ottenere una stima dell'SCR sulla base delle tecniche esposte nella sezione 1.2.2.

Sulla base di quanto detto nelle sezioni 1.3.3 e 2.1.1, invece, noi consideriamo un nuovo set di M simulazioni monoannuali $(\mathbf{Y}_1^{(j)}, \mathbf{D}_1^{(j)})$ per $j = 1, \dots, M$. Ciò consente da un lato di scegliere M significativamente maggiore di N (infatti in questo secondo set di scenari non dovremo effettuare il ciclo interno di simulazioni né valutare il portafoglio se non tramite la proxy, per cui il costo computazionale ad essi associato è molto più basso), dall'altro consente di simulare gli N scenari di fitting tramite una misura di probabilità diversa da quella real-world al fine di ottenere una calibrazione più precisa della funzione proxy, dato che tali scenari in questo modo non influiscono più direttamente sulla distribuzione della perdita.

A questo punto possiamo procedere con la valutazione di $f(\mathbf{Y}_1^{(j)}, \mathbf{D}_1^{(j)})$ per ogni scenario $j = 1, \dots, M$ tramite la formula (2.9) e ottenere $PVFP_1^{(j)} \approx f(\mathbf{Y}_1^{(j)}, \mathbf{D}_1^{(j)})$. In questo modo come prima potremo calcolare il capitale disponibile in ogni scenario e quindi la funzione di perdita. Avremo così una distribuzione empirica della perdita, basata su M scenari, sulla base della quale possiamo procedere con la determinazione di una stima dell'SCR.

2.1.3 Analisi di convergenza dell'algoritmo

Seguiamo dapprima Bauer et al. (2009). Considereremo pertanto l'algoritmo esposto nella sezione 2.1.2, mantenendo invariata l'impostazione originale di Bauer (quindi simulando gli N scenari esterni di calibrazione a partire dalla distribuzione real-world e riciclandoli per la successiva fase di calcolo dell'SCR). Per semplicità inoltre considereremo un singolo scenario interno per ogni scenario esterno nella fase di calibrazione.

L'obiettivo è dimostrare la convergenza in distribuzione di \tilde{f} a f per $N \rightarrow +\infty$ e di f a $PFVP_1$ per $H \rightarrow +\infty$. Tuttavia, sorge una difficoltà dovuta al cambio di misura effettuato in $t = 1$, che impone cautela nell'adottare i classici risultati della letteratura statistica. In Bauer et al. (2009) si propone una possibile soluzione, che consiste nella costruzione di una misura di probabilità alternativa P' su una copia esatta $(\Omega', \mathcal{A}', \mathcal{F}')$ dello spazio di probabilità filtrato $(\Omega, \mathcal{A}, \mathcal{F})$ utilizzato finora, in modo che siano soddisfatte le relazioni:

$$\mathbb{E}_Q(X | \mathcal{F}_1) = \mathbb{E}_{P'}(X | \mathcal{F}_1) \quad (2.10)$$

per ogni variabile aleatoria $X \in \mathcal{A}$ e:

$$P(Z \leq z) = P'(Z \leq z) \quad \forall z \in \mathbb{R} \quad (2.11)$$

per ogni variabile aleatoria Z che sia \mathcal{F}_t -misurabile con $t \in [0, 1]$.

Consideriamo ad esempio un modello di Black-Scholes per l'evoluzione del mercato finanziario. Assumiamo quindi che il titolo S si sviluppi secondo le equazioni differenziali:

$$\begin{aligned} dS_t &= S_t(\mu dt + \sigma dW_t) \\ dS_t &= S_t(r dt + \sigma dZ_t) \end{aligned}$$

con $S_0 > 0$ e ove $\mu \in \mathbb{R}$ rappresenta il termine di drift del titolo, W, Z sono moti browniani rispettivamente sotto P e Q , $r \in \mathbb{R}$ è il tasso d'interesse e $\sigma > 0$ la volatilità. Allora se W' è un moto browniano nello spazio $(\Omega', \mathcal{A}', P')$ e S si sviluppa secondo:

$$dS_t = S_t(\mu' dt + \sigma dW'_t)$$

ove poniamo:

$$\mu' = \mu \mathbb{1}_{\{0 \leq t \leq 1\}} + r \mathbb{1}_{\{1 < t < +\infty\}}$$

ne segue che P' soddisfa alle condizioni (2.10) e (2.11). Pertanto, in questo specifico caso siamo riusciti a costruire la misura di probabilità P' semplicemente tramite una manipolazione del termine di drift.

Un metodo generale per la costruzione della probabilità P' è invece considerato in Ha & Bauer (2018). Ivi si introduce la misura P' tramite la sua derivata di Radon-Nikodym:

$$\frac{\partial P'}{\partial P} = \frac{\frac{\partial Q}{\partial P}}{\mathbb{E}_P\left(\frac{\partial Q}{\partial P} \mid \mathcal{F}_1\right)} \quad (2.12)$$

Proposizione 2.1.

Introducendo la misura di probabilità P' tramite la formula (2.12), sono verificate le seguenti proprietà:

1. $P'(A) = P(A) \quad \forall A \in \mathcal{F}_t, 0 \leq t \leq 1.$
2. $\mathbb{E}_Q(X \mid \mathcal{F}_1) = \mathbb{E}_{P'}(X \mid \mathcal{F}_1) \quad \forall v.a. X \mathcal{A}\text{-misurabile.}$

Dimostrazione.

1. Consideriamo $A \in \mathcal{F}_t, 0 \leq t \leq 1.$ Si ha allora:

$$P'(A) = \mathbb{E}_{P'}(\mathbb{1}_{\{A\}}) = \mathbb{E}_P\left(\frac{\partial P'}{\partial P} \mathbb{1}_{\{A\}}\right) = \mathbb{E}_P\left[\mathbb{E}_P\left(\frac{\partial P'}{\partial P} \mathbb{1}_{\{A\}} \mid \mathcal{F}_1\right)\right]$$

ove abbiamo sfruttato il fatto che la probabilità di un evento coincide con la speranza matematica dell'indicatore, le proprietà della derivata di Radon-Nikodym e la proprietà iterativa della speranza matematica condizionata. Tramite la definizione nella formula (2.12) e la \mathcal{F}_1 -misurabilità di $\mathbb{1}_{\{A\}}$ e di $\mathbb{E}_P\left(\frac{\partial Q}{\partial P} \mid \mathcal{F}_1\right)$ otteniamo poi:

$$\begin{aligned}
P'(A) &= \mathbb{E}_P \left[\mathbb{E}_P \left(\frac{\frac{\partial Q}{\partial P}}{\mathbb{E}_P \left(\frac{\partial Q}{\partial P} \mid \mathcal{F}_1 \right)} \mathbb{1}_{\{A\}} \mid \mathcal{F}_1 \right) \right] \\
&= \mathbb{E}_P \left[\frac{\mathbb{1}_{\{A\}}}{\mathbb{E}_P \left(\frac{\partial Q}{\partial P} \mid \mathcal{F}_1 \right)} \mathbb{E}_P \left(\frac{\partial Q}{\partial P} \mid \mathcal{F}_1 \right) \right] \\
&= \mathbb{E}_P(\mathbb{1}_{\{A\}}) = P(A)
\end{aligned}$$

2. Consideriamo la variabile aleatoria $X : \Omega \rightarrow \mathbb{R}$. Seguendo lo stesso ragionamento del punto sopra si ha:

$$\mathbb{E}_{P'}(X \mid \mathcal{F}_1) = E_P \left(\frac{\partial P'}{\partial P} X \mid \mathcal{F}_1 \right) = E_P \left(\frac{\frac{\partial Q}{\partial P}}{\mathbb{E}_P \left(\frac{\partial Q}{\partial P} \mid \mathcal{F}_1 \right)} X \mid \mathcal{F}_1 \right)$$

Ancora, per la \mathcal{F}_1 -misurabilità si ottiene:

$$\begin{aligned}
\mathbb{E}_{P'}(X \mid \mathcal{F}_1) &= \frac{1}{\mathbb{E}_P \left(\frac{\partial Q}{\partial P} \mid \mathcal{F}_1 \right)} \mathbb{E}_P \left(\frac{\partial Q}{\partial P} X \mid \mathcal{F}_1 \right) \\
&= \mathbb{E}_P \left(\frac{\partial Q}{\partial P} X \mid \mathcal{F}_1 \right) = \mathbb{E}_Q(X \mid \mathcal{F}_1)
\end{aligned}$$

ove abbiamo sfruttato il fatto che $\mathbb{E}_P \left(\frac{\partial Q}{\partial P} \mid \mathcal{F}_1 \right) = 1$.

□

Pertanto, la misura di probabilità P' così costruita soddisfa alle proprietà (2.10), (2.11) ed è quindi ammissibile per lo studio del problema in esame.

Proposizione 2.2.

\tilde{f} converge a f P' -q.c. per $N \rightarrow +\infty$.

Dimostrazione.

Posto per semplicità $\mathbf{Z}_1 = (\mathbf{Y}_1, \mathbf{D}_1)$ e definito $\mathbf{g} = (g_1, \dots, g_H)$ il vettore delle funzioni base, si nota (Ha & Bauer (2018)) che possiamo scrivere:

$$\tilde{\boldsymbol{\beta}} = A^{-1} \frac{1}{N} \sum_{i=1}^N \mathbf{g}(\mathbf{Z}_1^{(i)}) \widetilde{PVFP}_1^{(i)}$$

ove $A = \left[\frac{1}{N} \sum_{i=1}^N g_r(\mathbf{Z}_1^{(i)}) g_s(\mathbf{Z}_1^{(i)}) \right]_{r,s=1,\dots,M}$. Avendo assunto la lineare indipendenza delle funzioni base, possiamo assumere che, per N sufficientemente grande, A sia invertibile.

Posto poi $A' = [\mathbb{E}_{P'}(g_r(\mathbf{Z}_1)g_s(\mathbf{Z}_1))]_{r,s=1,\dots,M}$, si ha:

$$\boldsymbol{\beta} = (A')^{-1} \mathbb{E}_{P'} \left(\mathbf{g}(\mathbf{Z}_1) \sum_{t=1}^T X_t \right)$$

Per la legge forte dei grandi numeri si ha che $\tilde{\boldsymbol{\beta}} \rightarrow \boldsymbol{\beta}$ P' -q.c. e quindi anche: $\tilde{f} = \mathbf{g}^T \tilde{\boldsymbol{\beta}} \rightarrow \mathbf{g}^T \boldsymbol{\beta} = f$ P' -q.c. \square

Proposizione 2.3.

L'algoritmo LSMC come definito nella sezione 2.1.2 gode delle seguenti proprietà:

1. f converge a $PVFP_1$ in media quadratica per $H \rightarrow +\infty$
2. $\sqrt{N} (\tilde{f} - f)$ tende in distribuzione ad una normale $N(0, \xi_H)$.

Si rimanda a Ha & Bauer (2018) per la dimostrazione di tale risultato. Inoltre, nello stesso lavoro si possono trovare risultati più formali relativi alla convergenza congiunta e alla velocità di convergenza.

Proposizione 2.4.

\tilde{f} converge in distribuzione a $PVFP_1$, ossia:

$$P(\tilde{f} \leq l) \rightarrow P(PVFP_1 \leq l) \quad \forall l \in \mathbb{R} \quad (N, H \rightarrow +\infty)$$

Inoltre, indicate con $F_{\tilde{f}}^{\leftarrow}$ e $F_{PVFP_1}^{\leftarrow}$ le inverse generalizzate delle funzioni di ripartizione delle due variabili, si ha anche la convergenza dei quantili della distribuzione per ogni α punto di continuità di $F_{PVFP_1}^{\leftarrow}$:

$$F_{\tilde{f}}^{\leftarrow}(\alpha) \rightarrow F_{PVFP_1}^{\leftarrow}(\alpha) \quad (N, H \rightarrow +\infty)$$

Dimostrazione.

La prima parte del risultato discende immediatamente dalle proposizioni 2.2 e 2.3, ricordato che la convergenza quasi certa e la convergenza in media quadratica implicano la convergenza in probabilità, che a sua volta implica la convergenza in distribuzione. La seconda parte del risultato discende dalla prima, in quanto

la convergenza in distribuzione implica quella dei quantili nei punti di continuità dell'inversa. Verificheremo tale risultato nel caso considerando un solo limite (in questo modo abbiamo verificato il risultato generale se si considerano i due limiti sequenzialmente).

Fissato α punto di continuità per F^{\leftarrow} , possiamo considerare una sequenza $(e_n)_{n \in \mathbb{N}}$ convergente a 0 tale che, per ogni n , $F^{\leftarrow}(\alpha) - e_n$ e $F^{\leftarrow}(\alpha) + e_n$ siano ancora punti di continuità e che:

$$F(F^{\leftarrow}(\alpha) - e_n) < \alpha < F(F^{\leftarrow}(\alpha) + e_n)$$

Sia $(F_n)_{n \in \mathbb{N}}$ la successione di funzioni di ripartizione convergente a F . Allora, fissato n , possiamo prendere \bar{n} tale che $\forall n > \bar{n}$ valga:

$$F_n(F^{\leftarrow}(\alpha) - e_n) < \alpha < F_n(F^{\leftarrow}(\alpha) + e_n)$$

e quindi per ogni $n \geq \bar{n}$ si ha:

$$F^{\leftarrow}(\alpha) - e_n < F_n^{\leftarrow}(\alpha) < F^{\leftarrow}(\alpha) + e_n$$

da cui si ottiene che $F_n^{\leftarrow}(\alpha) \rightarrow F^{\leftarrow}(\alpha)$. □

Questo risultato garantisce quindi la convergenza del Value-at-Risk ottenuto tramite il metodo LSMC qualora consideriamo i limiti derivanti dalle due approssimazioni in maniera sequenziale. Risultati relativi alla convergenza congiunta sotto condizioni più restrittive sono disponibili in Ha & Bauer (2018).

Osserviamo infine che lo stimatore LSMC dell'SCR avrà la stessa componente di bias positivo dello stimatore del modello di simulazioni stocastiche annidate, della quale abbiamo trattato nell'ultima parte della sezione 1.2.2. Per una derivazione formale di questo risultato si veda ancora Ha & Bauer (2018), in cui l'approccio di Gordy & Juneja (2010), brevemente descritto nella sezione 1.2.2, viene adattato al contesto del modello LSMC.

2.1.4 Determinazione della struttura del polinomio

Torniamo all'implementazione pratica del modello LSMC illustrata nella sezione 2.1.1 e consideriamo N scenari di fitting, nei quali sia calcolato il valore del portafoglio (o, equivalentemente, della perdita), $y^{(i)}$, $i = 1, \dots, N$ e siano disponibili i valori assunti dai d fattori di rischio, $\mathbf{x}^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})$ per $i = 1, \dots, N$.⁴

Come abbiamo già illustrato nelle sezioni precedenti, possiamo interpretare $(y^{(i)}, \mathbf{x}^{(i)})$ come un campione con il quale stimare i coefficienti di un modello di regressione. Il modello considerato sarà polinomiale, in modo da catturare la non-linearità della relazione e l'interazione tra i regressori. Ad esempio, considerato il caso bidimensionale con fattori di rischio X_1, X_2 la struttura del modello sarà data da:

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \beta_3 X_{i,1} X_{i,2} + \beta_4 X_{i,1}^2 + \beta_5 X_{i,2}^2 + \dots + \varepsilon_i \quad (2.13)$$

Senza alcun vincolo al modello così sviluppato, il numero di termini sarebbe infinito; anche semplici limitazioni al grado o al numero dei termini potrebbero condurre ad eccessivo overfitting, che, per quanto detto nella sezione 2.1.1, è particolarmente importante evitare in questa applicazione. Servirà quindi un procedimento di selezione per includere nel modello soltanto i termini effettivamente rilevanti. Vediamo, nel seguito, i principali procedimenti iterativi utilizzati nella letteratura relativa ai modelli di regressione:

- Procedimento forward: partendo dal modello nullo, ad ogni passo si considerano tutti i modelli ottenuti aggiungendo al modello del passo precedente (modello base) un singolo nuovo termine tra quelli non ancora presenti. Tutti i modelli così creati vengono comparati secondo un qualche criterio e il termine che porta al modello migliore viene inserito; tale procedimento viene ripetuto nel passo successivo, questa volta considerando nel modello base anche il termine scelto nello step precedente. Il procedimento termina quando

⁴Rispetto alla simbologia utilizzata nelle sezioni precedenti, per adeguarci alla notazione standard relativa ai modelli di regressione, indichiamo ora con x anziché con z i risk factor nel nostro modello. Nella sezione 2.1.1 dovevamo distinguere infatti i risk factor dai regressori del modello; in questa sezione, invece, considerato che, per via dell'impostazione polinomiale seguita, i regressori sono monomi dei risk factor, andando ad investigare più nel dettaglio la struttura del modello possiamo unificare le due notazioni.

in un dato step nessun modello è preferito al modello base oppure quando tutti i termini possibili sono stati inseriti.

- Procedimento backward: partendo dal modello contenente tutti i regressori, ad ogni passo si considerano i modelli ottenuti togliendo al modello del passo precedente (modello base) un singolo termine tra quelli presenti. I modelli così creati vengono comparati secondo un qualche criterio e il termine che porta al modello peggiore viene eliminato, per poi passare al passo successivo con il modello base aggiornato. Il procedimento termina quando in un dato passo nessun modello ridotto è giudicato migliore di quello base oppure quando si è giunti al modello nullo.
- Procedimento stepwise: combina il procedimento forward con il procedimento backward. La configurazione più utilizzata (forward-stepwise) parte dal modello nullo e segue l'algoritmo forward, facendo tuttavia un passo backward dopo ogni passo forward: rispetto all'algoritmo forward, quindi, dopo ogni inserimento di un nuovo termine del modello viene effettuato un controllo per verificare che la nuova variabile non abbia reso obsoleta una variabile precedentemente inserita, che, in caso, verrebbe eliminata prima di procedere con lo step forward successivo. Un'altra versione (backward-stepwise) parte dal modello completo ed esegue sempre prima uno step backward, poi uno step forward per controllare che l'eliminazione di una nuova variabile non abbia reso nuovamente rilevante una variabile scartata in precedenza.

Tali processi sono stati pensati per la selezione dei regressori nel contesto del modello lineare classico. Nel caso della regressione polinomiale, in realtà, il procedimento backward risulta poco sensato - in quanto non è definito a priori un modello completo, a meno di non fissare un grado massimo del polinomio - e in ogni caso computazionalmente non attuabile: anche se fissassimo arbitrariamente limitazioni in modo da poter definire un modello completo, infatti, per come è costruito l'algoritmo ogni suo passo richiederebbe un numero di valutazioni irragionevolmente elevato (ad es., con solo 8 fattori di rischio il polinomio completo di sesto grado avrebbe 3003 termini). Non considereremo quindi procedimenti backward o backward-stepwise.

Tra i tre procedimenti, lo stepwise è tendenzialmente quello considerato più adeguato. Tuttavia, anche qui è necessario bilanciare l'adeguatezza del criterio con i vincoli computazionali: infatti, richiedendo uno step backward dopo ogni step forward, il numero di confronti necessario per l'implementazione è sicuramente superiore. Considerato che il problema computazionale è molto sentito nella nostra applicazione, non è chiaro se il beneficio derivante dall'utilizzo di questa tecnica porti in realtà a significative migliorie nel modello.

Oltre al tipo di procedimento, considerando in particolare la natura polinomiale del modello, bisognerà scegliere se e in che modo implementare il principio statistico della marginalità, per il quale i termini di interazione non devono essere inseriti nel modello se non lo sono i corrispondenti termini principali. Considerato che, inoltre, il nostro approccio è polinomiale e che, con una leggera forzatura, ai fini del principio potremo considerare le potenze come interazione tra un termine e sé stesso, avremo diverse possibilità per la selezione dei candidate terms da considerare ad ogni passo. Per semplicità, enuncieremo tali possibilità in relazione ad un modello forward o forward-stepwise, fermo restando che la stessa logica è applicabile anche a modelli backward (in tal caso, l'enunciato del principio di marginalità implica che i termini principali non possono essere tolti dal modello se non lo sono prima i termini di interazione ad essi associati).

- Algoritmo unconstrained: ad ogni passo dell'algoritmo, tutti i termini non ancora inseriti nel modello sono ammissibili. Il principio di marginalità non viene soddisfatto.
- Algoritmo adaptive: inizialmente sono considerati ammissibili tutti i termini lineari (ossia di grado 1). Successivamente, ogni volta che un termine viene inserito nel modello, vengono inseriti tra i candidati tutti i termini per i quali tutti i termini ad essi marginali sono presenti nel modello. A titolo di esempio, il termine x^2yz sarà candidabile per l'algoritmo solo se sono stati inseriti nel modello i termini $x, y, z, x^2, xy, xz, yz, xyz, x^2y, x^2z$. Il modello così costruito soddisfa il principio di marginalità.
- Algoritmo layered: si procede come nell'algoritmo adaptive, con l'ulteriore restrizione che i termini di grado k saranno inseribili nel modello solo una

volta che si sarà selezionato il miglior modello di grado $k - 1$. Pertanto, a titolo di esempio, il termine x^2yz sarà candidabile solo se sono stati inseriti nel modello i termini $x, y, z, x^2, xy, xz, yz, xyz, x^2y, x^2z$ e se tutti i termini di grado < 4 non ancora inseriti nel modello sono stati definitivamente scartati. In sostanza, rispetto all'algoritmo adapted, qui la candidabilità di nuovi termini viene verificata solo al raggiungimento del miglior polinomio di un dato grado massimo anziché dopo ogni passo del modello. Anche questo modello soddisfa il principio di marginalità.

Osserviamo che l'algoritmo unconstrained, oltre a non soddisfare il principio statistico di marginalità, comporta costi computazionali immensi, in quanto il numero di termini candidati ad ogni passo è estremamente elevato (infinito se non si fissa il grado massimo del polinomio). Pertanto, tale algoritmo non sarà più considerato. Inoltre, osserviamo che tendenzialmente l'algoritmo layered produrrà polinomi con trattamento più omogeneo dei diversi fattori di rischio rispetto all'algoritmo adaptive, che con maggiore facilità si adatterà meglio ai fattori più significativi a discapito di quelli meno significativi. L'algoritmo adaptive risulta computazionalmente più veloce, pertanto sarà quello da noi principalmente considerato.

Infine, dovremo scegliere il criterio con il quale confrontare i diversi modelli ad ogni passo. Il criterio dovrà essere in grado di catturare in maniera sintetica, ma adeguata, la bontà del modello e allo stesso tempo essere computazionalmente semplice da implementare. Di seguito sono presentate le principali proposte in letteratura (per semplicità nuovamente riferite ad un approccio di tipo forward):

- Test di Wald: ad ogni passo, si effettua un test di Wald per la significatività del nuovo coefficiente per ogni termine candidato. Il termine che minimizza il p -value associato al test sarà incluso nel modello. L'algoritmo si ferma quando nessun nuovo termine risulta significativo ad un fissato livello α . Tale criterio viene solitamente utilizzato nei procedimenti di selezione di tipo backward, eliminando via via le variabili non significative; il suo utilizzo in procedimenti forward è meno comune, nonché di dubbia validità.
- Adjusted R^2 : si tratta di una variante del coefficiente di determinazione $R^2 = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2 / \sum_{i=1}^N (y_i - \bar{y})^2$, cui viene applicata una penalizzazione

che aumenta con il numero di regressori del modello. In particolare, si ha: $R_{adj}^2 = 1 - (1 - R^2) \frac{N-1}{N-k-1}$ ove N come sempre è il numero di osservazioni e k il numero di termini nel modello. Ad ogni passo, si inserisce il termine che porta all' R^2 aggiustato più alto; l'algoritmo si ferma quando tutti i possibili termini non migliorano l' R^2 aggiustato del modello.

- Out-of-sample R^2 : si tratta del coefficiente di determinazione R^2 , calcolato tuttavia non sui dati utilizzati per il fitting, ma su un secondo campione, in modo da determinare la capacità di previsione del modello prescelto.
- Akaike Information Criterion: si tratta del più comune tra i criteri informativi, utilizzati per confrontare in maniera sintetica diversi modelli statistici. Esso è calcolato come $AIC = 2k - 2l(\hat{\beta})$ dove k è il numero di termini del modello e $l(\hat{\beta})$ è la funzione di log-verosimiglianza calcolata nel punto di massimo. Il modello con AIC minore è considerato migliore: il valore dell'AIC migliora all'aumentare del valore massimo della funzione di verosimiglianza e peggiora in modo lineare all'aumentare del numero di parametri.
- Bayesian Information Criterion: un altro tra i più comuni criteri informativi, è calcolato come $BIC = \log(N)k - 2l(\hat{\beta})$. La logica è molto simile all'AIC, ma la penalizzazione questa volta dipende da N e, in particolare per campioni consistenti, risulta molto più forte di quella corrispondente dell'AIC.

Nel modello finora considerato, delineato nella sezione 2.1.1, non abbiamo mai parlato di ipotesi distributive, limitandoci a considerare lo stimatore $\tilde{\beta}$ ricavato tramite il metodo dei minimi quadrati. Per calcolare i criteri informativi, in questa impostazione si considera:

$$l(\hat{\beta}) = N \log \left(\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \right) \quad (2.14)$$

il che corrisponde ad assumere la normalità degli errori del modello. Qualora non si possa accettare l'ipotesi di normalità, non sarà più possibile calcolare la verosimiglianza tramite la formula (2.14). Per calcolare i criteri informativi, allora, bisognerà ipotizzare un'altra distribuzione, stimare β con il metodo della massima

verosimiglianza e utilizzare poi la verosimiglianza del modello prescelto. In tal modo ci si allontana tuttavia dall'impostazione least-squares tipica del modello LSMC.

L'utilizzo dei soli criteri informativi per il confronto tra modelli è stato fortemente criticato nella letteratura statistica, in quanto essi sono ritenuti sostanzialmente incapaci di condurre di per sé alla scelta di modelli statisticamente adeguati per l'inferenza (si veda ad esempio Spanos (2006)). Ciò nonostante, si tratta delle tecniche di selezione più utilizzate nella pratica, a causa della loro semplicità di calcolo e ampia sfera di applicazione. In ogni caso, tuttavia, i criteri calcolati con la formula (2.14) nell'approccio dei minimi quadrati sono validi esclusivamente se sono soddisfatte le restrittive ipotesi di normalità e omoschedasticità. Alcuni approfondimenti sullo studio di un criterio più generale e robusto per la selezione del modello, nonché ulteriori informazioni sui criteri informativi presentati sono presenti in Baek et al. (2005).

Possiamo infine introdurre delle limitazioni alla struttura del modello, in modo da forzare il procedimento automatico a non considerare termini eccessivamente complessi e/o a interrompersi dopo il raggiungimento di un certo grado di complessità, in modo da ridurre ulteriormente il rischio di overfitting. Di seguito i vincoli più diffusi:

- limitazione del grado del polinomio;
- limitazione del grado dei main effects;
- limitazione del grado dei cross terms;
- limitazione del grado dei singoli risk factor all'interno dei cross terms;
- limitazione del numero di risk factor all'interno dei cross terms;
- limitazione del numero di termini del polinomio.

Oltre ai procedimenti qui elencati vi sono, nella letteratura statistica, anche tecniche più avanzate per effettuare la selezione delle variabili esplicative da inserire nel modello. A conoscenza di chi scrive sono tuttavia inesistenti o molto carenti le applicazioni al modello LSMC. Forniremo alcuni cenni e riferimenti bibliografici a queste tecniche nella sezione 2.2.3.

2.1.5 La validazione delle funzioni proxy

Abbiamo visto nelle sezioni precedenti che i modelli proxy si basano su un'approssimazione funzionale della funzione di perdita, ottenuta a partire da un set di simulazioni stocastiche annidate; inoltre, abbiamo osservato che non sempre le ipotesi alla base dei modelli utilizzati sono verificate (ad esempio, i procedimenti usuali di selezione delle variabili tramite criteri informativi richiedono l'omoschedasticità e la normalità degli errori, ma non abbiamo motivo di ritenere che queste assunzioni siano sempre accettabili). Ancora, non sono definiti criteri univoci per effettuare la selezione delle variabili e spesso vengono imposte arbitrariamente ulteriori limitazioni all'algoritmo.

A causa delle diverse imprecisioni e approssimazioni effettuate, non abbiamo in realtà alcuna garanzia che la funzione costruita sia effettivamente efficace nel descrivere il valore della perdita in ogni scenario e quindi non possiamo dire che il quantile al 99.5% del modello proxy sia sufficientemente vicino al quantile 99.5% del modello stocastico. Un altro aspetto che necessita di verifica è la capacità di previsione della funzione trovata, ossia la sua abilità nel descrivere con opportuna precisione la perdita anche in scenari diversi da quelli utilizzati per la calibrazione. Risulta allora necessario un sistema di validazione che verifichi l'adeguatezza della funzione proxy prima di procedere al suo utilizzo per la fase di calcolo dell'SCR.

Una prima fase del processo di validazione può essere un'ispezione manuale grafica della funzione creata, in modo da verificare che non siano evidenti problemi di under- o over-fitting. Inoltre, un'analisi dell' R^2 e del valore assunto dai criteri informativi durante le ultime iterazioni dell'algoritmo di selezione dei termini può aiutare a verificare la stabilità della funzione selezionata.

Tuttavia, un metodo più efficace, oggettivo e automatizzabile per verificare la bontà della funzione è il confronto diretto tra il valore previsto dalla proxy e il valore ottenuto tramite simulazioni stocastiche annidate in un nuovo set di scenari (out-of-sample scenarios), non utilizzati per la calibrazione. Questa volta, vorremo ottenere una valutazione precisa tramite le simulazioni annidate, quindi utilizzeremo un numero elevato (ad es. 1000) scenari interni per ogni scenario esterno, e non più 2 come per la calibrazione. Naturalmente, anche con tanti scenari neutrali al rischio, sarà ancora presente un errore di simulazione (seppur ridotto).

Per scegliere gli scenari di validazione, potremmo simulare dalla distribuzione real world dei fattori di rischio, oppure, come nella calibrazione, dalla distribuzione uniforme. Più spesso però, considerato il limitato numero di scenari in cui possiamo effettuare una valutazione precisa tramite simulazioni stocastiche annidate senza pesare troppo sul budget computazionale complessivo (che comunque in gran parte dovrà andare alla fase di calibrazione), si adotta una logica reminiscente il curve fitting, ossia vengono selezionati deterministicamente un numero limitato di scenari giudicati particolarmente significativi. Ad esempio, si potrebbe ritenere interessante verificare in particolare la bontà della proxy in scenari vicini al percentile 99.5% della distribuzione congiunta, o in scenari in cui viene considerato un valore significativamente elevato per solo uno o due risk factor. Non esistono metodi formali per la selezione del numero e della tipologia degli scenari da utilizzare per la validazione, alcuni suggerimenti pratici sono contenuti in The IFoA Aggregation and Simulation Working Party (2016).

Una volta calcolata in maniera precisa la perdita tramite simulazioni stocastiche annidate in ogni scenario di validazione, la confronteremo con quella prevista dalla proxy. Consideriamo in particolare K scenari di validazione e indichiamo con \hat{y}_k e y_k rispettivamente i valori previsti dalla proxy e i valori calcolati tramite simulazioni stocastiche annidate per $k = 1, \dots, K$. Possiamo allora definire le seguenti metriche, che utilizziamo per misurare la bontà di previsione:

$$\begin{aligned}\mathcal{M}_1 &= \max_{k=1, \dots, K} \left| \frac{\hat{y}_k}{y_k} - 1 \right| \\ \mathcal{M}_2 &= \frac{1}{K} \sum_{k=1}^K \left| \frac{\hat{y}_k}{y_k} - 1 \right| \\ \mathcal{M}_3 &= \frac{1}{K} \sum_{k=1}^K \frac{|y_k|}{\sum_{h=1}^K |y_h|} \left| \frac{\hat{y}_k}{y_k} - 1 \right| \\ \mathcal{M}_4 &= \max_{k=1, \dots, K} |\hat{y}_k - y_k| \\ \mathcal{M}_5 &= \frac{1}{K} \sum_{k=1}^K |\hat{y}_k - y_k| \\ \mathcal{M}_6 &= 1 - \frac{\sum_{k=1}^K (\hat{y}_k - y_k)^2}{\sum_{k=1}^K y_k^2}\end{aligned}$$

In particolare, \mathcal{M}_1 misura il massimo errore, in termini relativi, su tutti gli scenari di validazione e \mathcal{M}_2 l'errore assoluto medio (MAE); \mathcal{M}_3 generalizza \mathcal{M}_2 introducendo un sistema di pesi in modo da tenere conto del peso diverso delle differenze relative a seconda del valore nello scenario. \mathcal{M}_4 e \mathcal{M}_5 rappresentano le misure assolute corrispondenti a \mathcal{M}_1 e \mathcal{M}_2 . Infine, \mathcal{M}_6 rappresenta il coefficiente di determinazione R^2 calcolato sugli scenari di validazione (out-of-sample R^2), che abbiamo introdotto nella sezione precedente.

Utilizzeremo le metriche considerate sopra anche per il confronto tra modelli diversi. Oltre a questi indicatori, tuttavia, nella comparazione dei modelli è fondamentale tenere in considerazione la performance in termini di tempo di esecuzione: considerato infatti che l'obiettivo principale per l'esistenza del modello LSMC e dei modelli proxy in generale è ridurre lo sforzo computazionale necessario rispetto alle simulazioni stocastiche annidate, la velocità di esecuzione dell'algoritmo risulta un fattore molto significativo, che potrebbe anche far propendere per modelli meno precisi, ma più rapidi.

Infine, osserviamo che finora abbiamo concentrato l'attenzione sull'analisi di una singola funzione proxy, la cui sfera di applicazione dipenderà dalla granularità alla quale si è deciso di implementare il modello. Una scelta comune, come avevamo detto, è di considerare una funzione per ogni diverso portafoglio di attività e di passività della stessa tipologia detenuto dalla compagnia assicurativa. Oltre a validare la singola funzione, sarà opportuno stimare ex-post l'impatto complessivo del modello proxy sull'SCR della compagnia. Un possibile approccio in tal senso consiste nel considerare un set di scenari "vicini" al quantile 99.5% della distribuzione real-world (SCR Capital Region) e utilizzare un modello basato sulle simulazioni stocastiche annidate anziché sulle proxy limitatamente a quegli scenari, poi confrontare il risultato con quello derivante dal modello proxy. Così facendo non si considera l'errore dovuto al fatto che la Capital Region selezionata con il modello proxy non corrisponderà a quella ottenibile con una piena applicazione del modello annidato, ma si otterrà nondimeno una stima dell'errore di approssimazione commesso dalle tecniche proxy.

2.2 Estensioni della forma strutturale del modello

Nella sezione precedente abbiamo studiato il modello LSMC nella forma in cui viene applicato nella pratica in un contesto assicurativo. Si è visto che il passo maggiormente critico nella calibrazione del modello riguarda la procedura di determinazione della struttura del polinomio, per la quale linee guida precise spesso non sono disponibili e per cui le ipotesi statistiche alla base spesso non sono verificate o lo sono solo approssimativamente. Inoltre, la struttura polinomiale può richiedere un numero molto elevato di termini per ottenere un fit adeguato e, in generale, non riesce comunque a catturare adeguatamente curve che presentano una significativa non-linearità. L'obiettivo di questa sezione è discostarsi dalla forma strutturale tradizionalmente utilizzata dal modello LSMC nel contesto assicurativo, per esplorare proposte alternative presenti in letteratura.

2.2.1 Polinomi ortogonali

Date due funzioni a valori reali, f e g , definiamo il loro prodotto interno su un insieme di punti (x_1, \dots, x_n) tramite la relazione:

$$\langle f, g \rangle = \sum_{i=1}^n f(x_i)g(x_i)w_i \quad (2.15)$$

ove w_i sono pesi positivi assegnati. Definiamo inoltre la norma di una funzione come $\|f\| = (\langle f, f \rangle)^{\frac{1}{2}}$. Due funzioni sono ortogonali rispetto ad un insieme di punti se il loro prodotto interno sull'insieme è nullo.

Una sequenza di funzioni $\varphi_0, \dots, \varphi_n$ costituisce un sistema ortogonale rispetto a (x_1, \dots, x_n) se $\langle \varphi_i, \varphi_j \rangle = 0 \forall i \neq j$ e se $\|\varphi_i\| \neq 0 \forall i$. Se inoltre $\|\varphi_i\| = 1 \forall i$ si parla di sistemi ortonormali.

Volendo estendere la definizione di prodotto interno al caso continuo, consideriamo un intervallo $[a, b]$. Diremo che la famiglia di funzioni $\varphi_0, \dots, \varphi_n$ costituisce un sistema ortogonale rispetto ad una data funzione peso $w(x)$ se:

$$\int_a^b w(x) \varphi_i(x) \varphi_j(x) dx = \begin{cases} 0 & \text{se } i \neq j \\ C_i \neq 0 & \text{se } i = j \end{cases} \quad (2.16)$$

Una famiglia di polinomi p_0, \dots, p_n si dice triangolare se il grado del polinomio p_i è proprio pari ad i . Una famiglia triangolare di polinomi è detta ortogonale se costituisce un sistema ortogonale per un dato vettore di pesi.

Le principali famiglie di polinomi ortogonali definite nel continuo sono:

- Polinomi di Chebyshev: $p_n(\cos(\theta)) = \cos(n\theta)$, ortogonali rispetto alla funzione peso $w(x) = \frac{1}{\sqrt{1-x^2}}$ su $] -1, 1[$.
- Polinomi di Legendre: $p_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2 - 1)^n$, ortogonali rispetto alla funzione peso $w(x) = 1$ su $] -1, 1[$.
- Polinomi di Laguerre: $p_n(x) = \frac{1}{n!} e^x \frac{d^n}{dx^n} (e^{-x} x^n)$, ortogonali rispetto alla funzione peso $w(x) = e^{-x}$ su $[0, +\infty[$.
- Polinomi di Hermite: $p_n(x) = (-1)^n e^{(x^2/2)} \frac{d^n}{dx^n} e^{-(x^2/2)}$, ortogonali rispetto alla funzione peso $w(x) = e^{-(x^2/2)}$ su \mathbb{R} .

Le famiglie di polinomi ortogonali godono, generalmente, di buone proprietà. Alcuni approfondimenti teorici e i principali risultati riguardanti i polinomi ortogonali sono reperibili, ad esempio, in Björck (1996). Inoltre, nel modello LSMC formale introdotto nella sezione 2.1.2, qualora si consideri una famiglia di polinomi ortogonali come funzioni base nell'equazione (2.6), si riescono a dimostrare versioni più generali dei teoremi relativi alla convergenza dell'algoritmo LSMC presentati nella sezione 2.1.3. Rimandiamo a Ha & Bauer (2018) per tali risultati.

In particolare, nel contesto della stima secondo il metodo dei minimi quadrati, consideriamo il caso di una singola variabile esplicativa, della quale vogliamo inserire nel modello le prime n potenze. Consideriamo la struttura di regressione:

$$Y_i = \sum_{j=0}^n \beta_j p_j(x_i) + \varepsilon_i$$

Osserviamo che nell'approccio di regressione polinomiale adottato, avremo significativi problemi di multicollinearità, come osservato già nella sezione 2.1.1. Ciò comporta che la matrice design del modello sarà potenzialmente malcondizionata e l'inversione di $X^T X$ comporterà risultati fortemente instabili. Grazie alle proprietà

dei polinomi ortogonali, invece, questa limitazione del modello viene risolta: posto $X = (p_j(x_i))_{i,j}$ e considerata la stima dei minimi quadrati $\hat{\beta} = (X^T X)^{-1} X^T \mathbf{y}$, se p_0, \dots, p_n costituiscono un sistema ortogonale, allora $X^T X$ si riduce ad una matrice diagonale. Essendo l'inversa di una matrice diagonale pari alla matrice contenente l'inverso dei singoli elementi, ciò rende triviale il computo di $(X^T X)^{-1}$. In particolare, sotto questa impostazione avremo:

$$\hat{\beta}_j = \frac{\sum_{i=1}^N Y_i p_j(x_i)}{\sum_{i=1}^N (p_j(x_i))^2} \quad \forall j = 1, \dots, n$$

Se l'impostazione alla base dei polinomi ortogonali risulta semplice nel caso di un solo regressore, nel caso multidimensionale la costruzione di un sistema ortogonale in via analitica è significativamente più pesante dal punto di vista matematico e va oltre gli scopi di questa tesi.

Il problema del malcondizionamento della matrice design viene risolto in via numerica ortogonalizzando la matrice tramite la decomposizione QR, già presentata nella sezione 2.1.1. Si è verificato empiricamente che l'utilizzo di metodi alternativi di ortogonalizzazione non conduce, in generale, a sensibili miglioramenti nella calibrazione della funzione proxy. Per analisi in tal senso si veda Cathcart (2012).

2.2.2 Modelli additivi e spline

I modelli additivi sono estensioni del modello di regressione lineare, ottenuti supponendo che la funzione di regressione di Y rispetto ai predittori X_1, \dots, X_p sia costituita da una somma di funzioni lisce (cioè derivabili in ogni punto) dei regressori, permettendo quindi di catturare effetti di non-linearità pur mantenendo l'additività del modello. Osserviamo come la struttura teorica alla base del modello LSMC, delineata, ad es., nell'equazione (2.6), rientri perfettamente dentro la definizione strutturale dei modelli additivi. In questo framework generale è allora possibile considerare forme funzionali più complesse di quanto non sia stato fatto finora nell'approccio di regressione polinomiale.

Un caso di particolare interesse, nonché un'estensione naturale dell'approccio finora delineato, risulta essere quello considerante come funzioni base funzioni polinomiali a tratti (spline). In generale, una spline polinomiale di grado m è costruita

in modo che sia un polinomio di grado m negli intervalli e che sia continua e abbia derivate continue fino all'ordine $m - 1$ nei nodi. Vediamo nel seguito le principali tipologie di funzioni base utilizzabili, seguendo Wood (2017):

- B-splines: considerata una variabile x e $k+m+1$ nodi $x_1 < x_2 < \dots < x_{k+m+1}$ nel suo spazio di appartenenza, chiamiamo B-spline (o basis spline) di ordine m e rango k una funzione $f(x) = \sum_{i=1}^k \beta_i B_i^{(m)}(x)$ ove la B-spline $B_i^{(m)}$ è definita tramite la relazione ricorsiva di Cox-de Boor:

$$B_i^{(m)} = \frac{x - x_i}{x_{i+m-1} - x_i} B_i^{(m-1)}(x) + \frac{x_{i+m} - x}{x_{i+m} - x_{i+1}} B_{i+1}^{(m-1)}(x) \quad \forall i = 1, \dots, k$$

ove $B_i^{(1)} = \mathbb{1}_{[x_i, x_{i+1}]}$. Le B-splines sono rappresentazioni stabili di funzioni polinomiali a tratti.

- P-splines: considerata sempre una variabile x e $k + m + 1$ nodi $x_1 < x_2 < \dots < x_{k+m+1}$ nel suo spazio di appartenenza, la P-spline o B-spline penalizzata è una spline con rappresentazione in forma di B-spline in cui ai coefficienti $(\beta_1, \dots, \beta_k)$ viene applicata una penalità in modo da controllare il grado di tortuosità della funzione. Ciò infatti permette di ridurre il pericolo di overfitting senza dover effettuare scelte riguardanti la posizione e il numero di nodi, che è, solitamente, un problema di complessità molto maggiore. L'idea dietro alla penalizzazione è che quanto più i coefficienti sono vicini tra loro, tanto più la spline sarà liscia. La penalizzazione dipenderà da un parametro di lisciamiento λ . Valori elevati di λ comportano un peso maggiore alla necessità di lisciamiento rispetto a quella di adattamento ai dati.
- Prodotti tensoriali: l'obiettivo dei prodotti tensoriali è quello di introdurre effetti di interazione nel modello. Per semplicità, introdurremo questa tecnica nel caso bivariato: date due variabili x , y e due P-splines corrispondenti, $f_x(x) = \sum_{i=1}^k \beta_i B_i(x)$ e $f_y(y) = \sum_{j=1}^l \gamma_j C_j(y)$, l'idea alla base dei prodotti tensoriali è quella di far variare i coefficienti di una P-spline in maniera liscia rispetto all'altra variabile. Un effetto di questo tipo è ottenibile utilizzando le funzioni base già disponibili. Ad esempio, possiamo porre $\beta_i = \sum_{j=1}^l \gamma_{i,j} C_j(y)$, da cui si ottiene $f_{x,y}(x, y) = \sum_{i=1}^k \sum_{j=1}^l \gamma_{i,j} B_i(x) C_j(y)$.

Si rimanda a Wood (2017) per maggiori dettagli sulla costruzione dei prodotti tensoriali e per l'estensione a questo caso della scelta dei parametri di liscio.

Abbiamo visto che nelle P-splines dovremo selezionare un parametro di liscio λ che regoli la penalizzazione da applicare per la tortuosità. Un primo modo per la selezione del parametro è la scelta del valore di λ che minimizzi la funzione di cross-validation (calcolata sul campione di numerosità n):

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i^{(-i)}(\lambda)) \quad (2.17)$$

ove $\hat{y}_i^{(-i)}(\lambda)$ è il valore previsto dal modello stimato senza l' i -esima osservazione e tenendo fisso λ . La funzione di cross-validation è computazionalmente onerosa e non soddisfa proprietà di invarianza rispetto alle rotazioni, per cui problemi di ottimizzazione che conducono agli stessi risultati inferenziali possono portare a valori diversi di cross-validation. Per ovviare a queste limitazioni, è stata proposta una modifica che prende il nome di cross-validation generalizzata (*GCV*). Il valore di λ scelto sarà quello che minimizzerà la *CV* o la *GCV* del modello.

Rimandiamo a Wood (2017) per ulteriori dettagli sull'implementazione dei modelli additivi. L'approccio basato sulle spline è stato implementato nel problema di calcolo del capitale economico da Larcher (2015), che in un contesto controllato tramite spline ottiene risultati migliori che con l'impostazione polinomiale. Tuttavia, nell'impostazione di Larcher sono considerati solamente un numero esiguo di fattori di rischio e la selezione dei termini è stata eseguita manualmente. Nonostante l'ottima capacità del modello nell'adattarsi ai dati, l'applicazione del metodo in un contesto più generale risulta molto difficile: non è infatti definito in maniera chiara un procedimento automatico per la selezione dei termini rilevanti e della forma funzionale da attribuire loro, in particolare per quanto concerne i termini di interazione. Una costruzione manuale di una funzione proxy basata sulle spline, oltre a introdurre arbitrarietà nel modello, risulta infatti difficilmente fattibile in un contesto assicurativo pratico. Un metodo basato sulle spline, ma che tiene conto automaticamente anche della selezione dei termini, limitandosi però a funzioni lineari a tratti, è dato dal modello MARS, cui accenniamo in seguito.

2.2.3 Altre proposte in letteratura

Nella letteratura statistica sono stati sviluppati molti metodi per modellare la relazione tra variabili, spesso utilizzabili anche nel contesto del modello Least-Squares Monte Carlo al posto della regressione polinomiale. Considerata tuttavia la necessità di costruire un modello sufficientemente generale da essere in grado di produrre in maniera automatica risultati significativi in un'ampia gamma di portafogli diversi e considerata anche la necessità di ottenere un modello semplice e ampiamente utilizzato, in modo da poter essere giudicato trasparente e approvato dagli organismi di vigilanza, non sono frequenti, nella pratica, le applicazioni di tecniche più avanzate. Ci proponiamo di fornire unicamente dei brevi cenni a quelle tecniche che sono state proposte in letteratura nell'ambito del modello LSMC, rimandando il lettore ad altri articoli concernenti maggiori dettagli teorici e applicazioni pratiche di tali tecniche ai modelli proxy.

- Kernel regression: si tratta di una tecnica basata sullo stimatore di Nadaraya e Watson per la funzione di regressione $E(Y | X) = f(X)$:

$$\hat{f}_h(x) = \frac{\sum_{i=1}^n K_h(x - x_i) y_i}{\sum_{i=1}^n K_h(x - x_i)}$$

in cui n rappresenta il numero di osservazioni, K_h una funzione Kernel (integrabile e non negativa) e h un parametro di liscio. Tipicamente vengono utilizzati Kernel gaussiani o di Epanechnikov. Lo stimatore di Nadaraya e Watson si estende inoltre in via immediata al caso multivariato. Per maggiori dettagli sulla costruzione dello stimatore e sul problema di scelta del parametro h si veda Hastie et al. (2009); un tentativo di applicazione ad un modello di stampo LSMC è contenuto in Chilla (2014).

- Robust regression: osserviamo che il problema di stima dei minimi quadrati delinato nell'equazione (2.8) è in realtà un caso particolare del problema di minimo:

$$\tilde{\beta} = \arg \min_{\beta \in \mathbb{R}^H} \left\{ \sum_{i=1}^N l \left(y_i - \sum_{h=1}^H \beta_h g_h(\mathbf{x}_i) \right) \right\}$$

ove si era considerato $l(r) = r^2$. Allora si potrebbe pensare di stimare il vettore dei coefficienti tramite funzioni di perdita $l(\cdot)$ diverse da quella quadratica, in modo da garantire maggiore robustezza al metodo, pesando eventuali valori anomali meno di quanto accadrebbe con il metodo dei minimi quadrati. In particolare, le funzioni comunemente utilizzate nel metodo di regressione robusta considerano $l(r) = r^2$ per valori ridotti di r , e funzioni diverse, meno penalizzanti, per valori grandi (ad esempio considerando la funzione di perdita l come costante da un certo r in poi).

Uno studio delle caratteristiche formali necessarie per una buona funzione di perdita, nonché un'applicazione al modello LSMC per il prezzamento di opzioni americane è reperibile in Jonen (2011). Il metodo di regressione robusta presentato in Jonen (2011) viene successivamente esteso al contesto del modello LSMC per il calcolo dei requisiti di capitale di un'impresa di assicurazioni in Nikolić et al. (2017), ove si riesce ad ottenere un miglioramento rispetto al metodo OLS nella capacità di previsione della funzione proxy. Inoltre, ivi si propone un innovativo approccio di regressione robusta con funzione di perdita asimmetrica (quindi con soglie e penalizzazioni diverse per outlier positivi e negativi, in modo da tener conto dell'asimmetria della distribuzione empirica dei residui).

Le principali limitazioni dell'approccio sono la difficoltà nel definire un criterio per la selezione delle soglie e delle penalizzazioni da applicare alla funzione di perdita; la difficoltà nell'automatizzazione del processo (la selezione dell'approccio da utilizzare dipenderà, ad esempio, dalla simmetria della funzione); l'utilizzo dei classici metodi OLS per la selezione della struttura polinomiale del modello.

- Modelli MARS (Multivariate Adaptive Regression Splines): si tratta di una tecnica di regressione non-parametrica che consente in maniera automatica di tenere conto degli effetti di non-linearità e interazione. Come nel modello di regressione polinomiale finora considerato, avremo una funzione di regressione con struttura di tipo $\sum_{h=1}^H \beta_h g_h(\mathbf{x})$. La particolarità dei modelli MARS consiste nel considerare come funzioni base funzioni lineari a tratti (o loro prodotti) anziché monomi.

Più nel dettaglio, la generica funzione base $g_h(\mathbf{x})$ potrà essere costante (l'intercetta), una funzione di tipo $(x - t)_+$ o $(t - x)_+$, oppure ancora prodotto di più funzioni di tipo $(x - t)_+$ o $(t - x)_+$ con diversi regressori, in modo da tener conto degli effetti di interazione⁵. I valori t rappresentano i nodi della funzione. In particolare, nel modello MARS sono candidabili come nodi tutti i valori osservati del regressore in considerazione: dati p regressori e N osservazioni, il numero di funzioni base possibili, senza tenere conto degli effetti di interazione, è di $2Np + 1$.

Considerando una singola variabile esplicativa, la struttura di regressione creata dal modello MARS quindi è del tipo:

$$Y = \beta_0 + \beta_1(x - t_1)_+ + \beta_2(x - t_2)_+ + \beta_3(t_3 - x)_+ + \beta_4(t_4 - x)_+ + \dots + \varepsilon$$

Nei modelli MARS è implementata anche in via automatica la procedura di selezione del modello tramite un algoritmo di selezione stepwise basato sul criterio della cross-validation generalizzata (GCV), di cui abbiamo parlato con riferimento ai modelli additivi.

Un'introduzione più completa ai modelli MARS è reperibile in Hastie et al. (2009). I modelli MARS godono di una forte semplicità e interpretabilità, e contrariamente alla regressione polinomiale sono già presenti loro implementazioni efficienti nei moderni software statistici. Inoltre, rispetto ai modelli additivi, di cui costituiscono un caso particolare, i modelli MARS consentono in maniera più semplice di selezionare la struttura del modello e catturano in maniera automatica gli effetti di interazione rilevanti. A conoscenza di chi scrive, tuttavia, non sono presenti in letteratura applicazioni del modello MARS al contesto delle tecniche proxy per il calcolo dell'SCR. Nella sezione 2.3 forniremo pertanto un tentativo di applicazione al nostro modello.

Un'interpretazione alternativa dei modelli di tipo MARS è invece data nell'ambito dei modelli ad albero nella sezione 3.1.1.

- Reti neurali: si tratta di modelli di apprendimento statistico la cui struttura è ispirata a quella di un sistema nervoso biologico. Esse sono costituite da

⁵Come da convenzione, indichiamo con $(f)_+$ la parte positiva di f . Si pone cioè $(x - t)_+ = \max(x - t, 0)$.

più strati di nodi (neuroni), collegati tra loro da segmenti (sinapsi). Ogni neurone riceve segnali dal mondo esterno o da altri neuroni e li combina tramite una somma pesata, per poi processarli tramite una funzione di attivazione e mandarli ai neuroni dello strato successivo.

Non ci proponiamo, in questa sede, di fornire dettagli sull'implementazione pratica di questi modelli, che presentano diversi gradi di complessità e, benché spesso siano in grado di modellare in maniera robusta, efficiente e parsimoniosa una vasta gamma di forme funzionali diverse, risultano complessi dal punto di vista dell'interpretazione dei parametri, venendo usualmente considerati come tecniche "black box", in cui il funzionamento interno del sistema risulta ignoto all'utente. Si rimanda sempre a Hastie et al. (2009) per un'introduzione ai meccanismi di funzionamento delle reti neurali, a Hejazi & Jackson (2017) per una prima applicazione ai modelli stocastici per il calcolo del capitale economico. Recenti applicazioni di confronto con il metodo LSMC sono date da Born (2018) e Nilsson & Sandberg (2018), nelle quali si ottengono risultati generalmente in linea con l'approccio polinomiale e talvolta leggeri miglioramenti in previsione, a scapito però dell'interpretabilità del modello e del tempo di esecuzione necessario.

Osserviamo infine che restando nello schema teorico su cui fonda il modello Least-Squares Monte Carlo - ossia considerando sempre un'approssimazione di $PFVP_1$ come combinazione lineare di funzioni base dei fattori di rischio - potremmo considerare come funzioni base, anziché solo monomi dei fattori di rischio, funzioni che "ricordano" le formule chiuse di valutazione di opportuni strumenti finanziari. In questo modo ci riconduciamo ad un modello ibrido tra Least-Squares Monte Carlo e portafogli replicanti, in cui i coefficienti, stimati con un modello di regressione, sono interpretabili come quote di composizione del portafoglio replicante.

Inoltre, uno studio formale sull'ottimalità delle funzioni base da utilizzare per velocizzare la convergenza dell'algoritmo LSMC è contenuto in Ha & Bauer (2018).

2.3 Implementazione del modello LSMC

Nel corso di questa sezione ci proponiamo di sottoporre ad un'applicazione pratica il modello LSMC di cui finora si sono esaminate le caratteristiche teoriche. L'implementazione sarà eseguita con un algoritmo indipendentemente sviluppato per la costruzione della funzione proxy (presentato nell'Appendice B), mentre tratteremo come dati in input sia le simulazioni dei fattori di rischio considerati, ottenute mediante sequenze di Sobol, sia il valore dei portafogli considerati in ciascuna simulazione. Possiamo pensare, infatti, che le prime ci siano date da un opportuno generatore di scenari stocastici e i secondi siano calcolati con opportuni modelli attuariali di proiezione dei flussi di cassa. In un contesto pratico, solitamente, tali strumenti saranno già a disposizione di un'impresa che intenda avvalersi di un modello di questo tipo per il calcolo dell'SCR, pertanto si ritiene ridondante, oltre che fuori dalla portata di questa tesi, lo sviluppo autonomo anche di queste componenti, pur fondamentali, del modello.

In tal modo, inoltre, si potrà testare il modello LSMC e studiarne le diverse configurazioni in un contesto complesso, considerando quindi una pluralità di fattori di rischio diversi, in modo da verificare l'adeguatezza dell'approccio anche in problemi ad alta dimensionalità.

In questa sezione limiteremo le nostre analisi in larga parte alla struttura polinomiale del modello LSMC, mantenendoci quindi fedeli all'approccio proposto in letteratura. Verificheremo, in particolare, il funzionamento del modello sulla base delle metriche di errore presentate nella sezione 2.1.5.

2.3.1 I dati e il contesto lavorativo

Ai fini della nostra analisi, considereremo tre compagnie di assicurazioni vita, che si differenziano per collocazione geografica, per tipologie di attività e passività detenute, per composizione dei propri portafogli, per fattori di rischio osservati. In tutti i casi la quantità modellata sarà il Market Value dei portafogli di attività e le Best Estimates of Liabilities (BEL) al lordo della riassicurazione per i portafogli di passività.

La valutazione separata dei portafogli di attività e delle BEL costituisce un metodo più raffinato di stressare gli elementi del bilancio rispetto a quello basato direttamente sul PVFP, che abbiamo delineato nell'approccio teorico nelle sezioni 1.1.2, 1.2.2., 2.1.2. Ciò accade in quanto il valore delle attività nei modelli di proiezione attuariale solitamente è considerato in maniera relativamente poco granulare, e quindi considerando funzioni proxy separate per le diverse categorie di attivi e passivi si ottiene una rappresentazione più fedele dei movimenti degli elementi del bilancio dell'impresa, a costo di dover procedere allo sviluppo di un processo per l'aggregazione dei risultati. Pertanto, forniremo prima degli esempi di applicazione a singoli portafogli di attivi e passivi, mentre nella sezione 2.3.3 utilizzeremo lo stesso approccio per modellare i PVFP complessivi dell'impresa. Ciò ci permetterà di ricavare in via semplificata una stima per il valore dell'SCR. Per modellare le attività adotteremo un approccio semplice: la loro valutazione infatti è effettuabile in maniera accurata dai tool di asset pricing, senza necessitare di scenari interni neutrali al rischio; quindi, disponendo di dati precisi, saranno sufficienti molti meno scenari esterni per calibrare la funzione proxy. Utilizzeremo pertanto solamente un numero limitato di scenari per il fitting, generati tramite sequenze di Sobol.

Per le passività adotteremo invece il pieno modello LSMC con un budget complessivo per il fitting pari a 50'000 scenari, come anche nelle applicazioni in Manukian (2012), Larcher (2015), Vounzi (2017). Considereremo la configurazione classica adottata negli esempi proposti nella letteratura del modello LSMC, ossia 25'000 scenari esterni generati tramite sequenze di Sobol e 2 scenari neutrali al rischio antitetici per ogni scenario esterno. Per i dettagli teorici relativi alle sequenze di Sobol e al metodo delle variabili antitetiche si rimanda alla sezione 1.2.3. Considereremo inoltre un certo numero di scenari di validazione non utilizzati per il fitting, per i quali disponiamo del valore delle passività calcolato con 1'000 scenari neutrali al rischio. Il numero e la tipologia di scenari utilizzati per validare la funzione varierà da compagnia a compagnia. Oltre a questo primo set di scenari di validazione, considereremo un secondo set di scenari sulla coda della distribuzione, con l'obiettivo di verificare non più la bontà globale della funzione, ma la sua abilità di descrivere in maniera adeguata la perdita in scenari estremamente sfavorevoli (vicini, in un certo senso, agli scenari real-world che generano l'SCR). Anche in questo

secondo set di scenari avremo a disposizione il valore del portafoglio calcolato con un approccio stocastico e mille scenari interni.

Ricordiamo che il costo computazionale per la generazione degli scenari di validazione è molto elevato: la creazione di soli 50 scenari di validazione comporta infatti lo stesso costo computazionale dell'intero budget utilizzato per la calibrazione del modello, data la molto più elevata precisione richiesta per il valore in questo scenario. Ne consegue che il confronto tra modelli sarà basato su metriche calcolate sulla base di un numero ridotto di scenari out-of-sample, forse non sufficienti ad ottenere risultati robusti. Tuttavia, aumentare il numero di scenari out-of-sample toglierebbe preziose risorse alla calibrazione del modello.

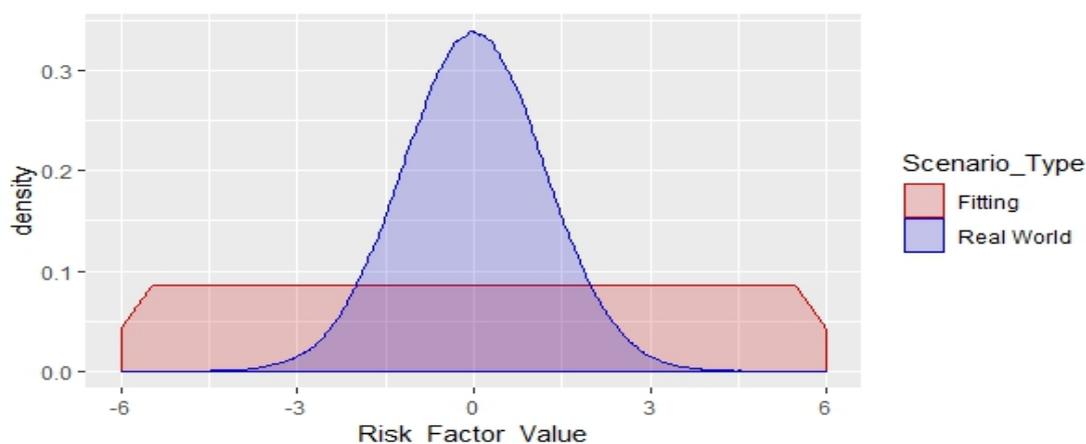
Si potrebbero considerare altri metodi per valutare la bontà di previsione del modello, come ad esempio le tecniche di cross-validation, che non richiedono scenari out-of-sample. Tuttavia, il loro utilizzo in questo contesto non sarebbe del tutto appropriato: ricordiamo infatti che i dati utilizzati per il fitting, essendo calcolati come media di soli due scenari neutrali al rischio, detengono una componente di errore anche molto significativa, che non si vuole catturare. L'unico modo per verificare che il modello sia in grado di prevedere bene dati affidabili è quello basato sugli scenari out-of-sample calcolati sempre con il modello stocastico, ma in maniera molto più precisa.

I rischi considerati saranno:

- A. Rischio di tasso d'interesse;
- B. Rischio azionario;
- C. Rischio di volatilità del mercato;
- D. Rischio di credito;
- E. Rischio di cambio;
- F. Rischio relativo al valore degli attivi a copertura delle riserve;
- G. Rischio legato alle spese di gestione;
- H. Rischio di morbilità;

- I. Rischio di riscatto (contenente anche il rischio legato ad ogni altra opzionalità offerta al contraente, tra cui, in particolare, il rischio di abbandono);
- J. Rischio di mortalità;
- K. Rischio di mortalità catastrofica;
- L. Rischio di longevità.

Molti di questi rischi sono poi modellati in maniera più granulare, considerandone separatamente le diverse componenti, per ciascuna delle quali è stata calibrata una distribuzione real-world tramite metodi analitici o empirici. Gli scenari utilizzati per il fitting della funzione proxy sono stati ricavati coprendo uniformemente - tramite sequenze di Sobol - il range individuato della distribuzione real-world (troncato a quantili particolarmente estremi). A titolo d'esempio, nella figura seguente riportiamo un confronto tra le due distribuzioni per un generico fattore di rischio.



Confronto tra distribuzione real-world per un fattore di rischio e distribuzione degli scenari utilizzati per il fitting della funzione proxy.

Le passività detenute dalle compagnie saranno suddivise in quattro grandi categorie (Lines of Business o LoB), sia per la valutazione delle BEL, sia per l'approccio basato sul PVFP che adotteremo nella sezione 2.3.3. Definiremo tali macro-aree in maniera volutamente vaga, osservato in particolare che le compagnie prese in

considerazione avranno collocazione geografica e quindi anche prodotti diversi nei loro portafogli. In tal modo si punta a verificare la validità complessiva del modello su una pluralità di mercati e di composizioni di portafoglio differenti, senza focalizzarsi su uno specifico prodotto. Chiaramente più è granulare la suddivisione in LoB, più la funzione proxy calibrata sarà in grado di cogliere le specificità delle diverse tipologie di prodotti, ma allo stesso tempo sarà anche più complesso il successivo processo di aggregazione.

Le categorie da noi considerate sono le seguenti:

- **Protezione:** contiene tutti i contratti con beneficio principale in caso di morte (o eventualmente invalidità o malattia), oppure le componenti di beneficio caso morte derivanti dallo spacchettamento di prodotti più complessi. La line of business si suppone affetta dai rischi A, G, H, I, J, K e, eventualmente, dai rischi C, F qualora le specificità del portafoglio lo richiedano.
- **Risparmio:** contiene tutti i contratti con beneficio garantito o beneficio a scadenza se l'assicurato è in vita. Per alcune compagnie, sono incluse anche alcune forme di rendita vitalizia oppure di unit-linked con garanzie finanziarie. La line of business si suppone affetta dai rischi A, C, F, G, H, I, J, L.
- **Unit-linked:** contiene i contratti di tipo unit-linked senza garanzie finanziarie - quindi con beneficio direttamente proporzionale al valore di un fondo d'investimento. La line of business si suppone affetta dai rischi A, C, F, G, I, J.
- **Pensioni:** contiene piani pensionistici e forme di rendita vitalizia differita ad essi assimilabili. La line of business si suppone affetta dai rischi A, G, I, J, M e, eventualmente, dai rischi C, F qualora le specificità del portafoglio lo richiedano.

Per una maggiore caratterizzazione dei possibili contratti assicurativi di ogni categoria e per una descrizione delle specificità di ogni tipologia si veda, ad esempio, Olivieri & Pitacco (2015).

Gli attivi considerati saranno invece suddivisi nelle tre categorie seguenti:

- **Obbligazioni:** si considera l'intero portafoglio obbligazionario dell'impresa, qualsiasi sia la valuta e la provenienza dei titoli. Sono ivi incluse sia obbligazioni societarie che obbligazioni sovrane. La categoria si suppone soggetta ai rischi A, B, D, E.
- **Azioni:** si considera l'intero portafoglio azionario dell'impresa e lo si suppone soggetto ai rischi B, E e, eventualmente, al rischio A qualora le specificità del portafoglio lo richiedano.
- **Derivati:** si considerano i derivati nel portafoglio dell'impresa. Solitamente le imprese di assicurazione detengono un portafoglio con esposizione relativamente ridotta a questa categoria di strumenti finanziari, in particolare è molto bassa l'esposizione a derivati esotici. Pertanto anche questa categoria, nonostante le maggiori difficoltà nel prezzaggio, sarà da noi modellata con l'approccio semplificato e si suppone affetta dai rischi A, B, E.

Osserviamo che il modello viene mantenuto semplice, in questo contesto, a fini illustrativi, supponendo che le compagnie di riferimento detengano attività sufficientemente omogenee da poter essere inquadrate agevolmente in queste categorie. Per ottenere un modello più granulare bisognerebbe considerare separatamente, ad es., portafogli di obbligazioni suddivisi per valuta e per tipologia: risulta ragionevole, ad esempio, che obbligazioni relative a titoli di stato europei abbiano valori diversi dei fattori di rischio rispetto a obbligazioni societarie in dollari. Una più fine ripartizione delle proprie attività comporterà naturalmente l'esigenza di considerare un numero maggiore di fattori di rischio diversi e richiederà necessariamente un processo di aggregazione più complesso per ottenere risultati a livello di compagnia.

Per ogni categoria di attivi e passivi considerata sarà ricercata una funzione proxy polinomiale specifica tramite le tecniche precedentemente esposte. Ove non specificato altrimenti, l'algoritmo utilizzato per calibrare la funzione sarà l'adaptive forward basato sull'AIC, secondo quanto abbiamo esposto nella sezione 2.1.4. L'algoritmo è stato sviluppato in linguaggio R ed è presentato nell'Appendice B di questa tesi.

2.3.2 Applicazione a portafogli di attività e passività

Procediamo applicando l'algoritmo sviluppato alle tre compagnie in esame. Nella prima parte della sezione, il modello semplificato descritto nella sezione precedente sarà applicato ai portafogli di attività, nella seconda parte il modello LSMC pieno sarà applicato ai portafogli di passività. Le statistiche out-of-sample di bontà dell'approccio saranno poi confrontate per diverse configurazioni dell'algoritmo di selezione. In particolare ci concentreremo su:

- La metrica $\mathcal{M}_1 = \max_{k=1, \dots, K} \left| \frac{\hat{y}_k}{y_k} - 1 \right|$, che rappresenta il valore assoluto del massimo errore commesso dalla funzione proxy, in termini relativi rispetto al valore osservato.
- La metrica $\mathcal{M}_2 = \frac{1}{K} \sum_{k=1}^K \left| \frac{\hat{y}_k}{y_k} - 1 \right|$, che rappresenta l'errore assoluto medio (MAE), sempre in termini relativi rispetto al valore osservato.
- La metrica $\mathcal{M}_3 = \frac{1}{K} \sum_{k=1}^K w_k \left| \frac{\hat{y}_k}{y_k} - 1 \right|$ con $w_k = \frac{|y_k|}{\sum_{h=1}^K |y_h|}$, che rappresenta sempre l'errore assoluto medio, ma ponderato per il valore osservato in quello scenario. In tal modo, vengono penalizzati maggiormente gli errori relativi negli scenari in cui il valore osservato è particolarmente elevato. L'obiettivo è quello di "correggere" la metrica \mathcal{M}_2 per tenere conto dell'estrema variabilità che può avere il valore del portafoglio in diversi scenari.

A titolo di esempio, per comprendere la differenza tra le metriche \mathcal{M}_2 e \mathcal{M}_3 , si pensi al caso estremo in cui in uno scenario il portafoglio valga 100€, in un altro 1€. Supponendo che in entrambe le situazioni l'errore commesso dal metodo LSMC sia di 1€, la metrica \mathcal{M}_2 considererebbe i due errori sullo stesso piano, associando al modello un errore relativo medio di circa 50%, mentre la metrica \mathcal{M}_3 tiene conto del diverso impatto monetario dei due errori, associandovi quindi un errore relativo medio pesato di circa 1%.

Ci concentreremo maggiormente sulle metriche \mathcal{M}_1 e \mathcal{M}_2 per le attività e sulle metriche \mathcal{M}_1 e \mathcal{M}_3 per le passività. Ciò in quanto per le passività gli scenari con associato un valore maggiore sono anche quelli più sulla coda d'interesse, mentre per le attività la metrica \mathcal{M}_3 , per quanto utile a fini interpretativi, rischia di pesare maggiormente scenari troppo "benevoli", dando peso ridotto agli scenari sulla coda della distribuzione.

a) Modello LSMC semplificato per portafogli di attivi

Sulla base di quanto esaminato nella sezione 2.3.1, ai fini di valutare le attività, per le quali supponiamo di poter disporre di un numero limitato di dati sufficientemente precisi, applichiamo il modello LSMC e valutiamo la bontà dell'approssimazione su scenari out-of-sample.

Nel seguito sono presentati i risultati relativi alla valutazione delle funzioni calibrate negli scenari di validazione. Osserviamo che per limitare il costo computazionale e il pericolo di overfitting abbiamo considerato soltanto termini fino al quarto grado. Inoltre, come ulteriore vincolo computazionale abbiamo forzato l'arresto dell'algoritmo quando il numero massimo di termini raggiunge 150, in modo da ottenere risultati in tempi ragionevoli. L'impatto di tale limitazione sarà nullo tranne nei casi in cui l'algoritmo di selezione fatica a convergere. Analizzeremo successivamente l'effetto di questo vincolo sulla funzione calibrata.

Compagnia 1					
Categoria	N. termini	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Obbligazioni	148	6m27s	0.44%	0.03%	0.03%
Azioni	150	7m22s	2.35%	0.28%	0.24%

Tabella 2.1: *Statistiche di bontà di previsione per gli attivi della Compagnia 1.*

Osserviamo come il modello abbia raggiunto o quasi raggiunto il limite di 150 termini in tutti i casi, segno che il vincolo imposto a priori alla funzione proxy avrà potenzialmente un impatto rilevante. Il numero elevato di termini della funzione polinomiale è dovuto, anche, alla scarsa granularità che abbiamo adottato nella suddivisione delle attività in portafogli: se, ad esempio, avessimo diviso le esposizioni in maniera più fine, ad esempio separando le principali valute o le tipologie di obbligazioni, allora avremmo potuto ottenere funzioni proxy più contenute, nonché, forse, più interpretabili.

Notiamo poi come l'errore associato al modello per le obbligazioni sia estremamente ridotto, segno che l'adattamento ottenuto dalla funzione polinomiale è praticamente perfetto. L'errore associato al portafoglio azionario invece risulta superiore, ma

comunque di entità minima. I risultati ottenuti in previsione dal modello proxy dunque sembrano molto affidabili.

Ripetiamo l'analisi per la seconda compagnia in esame:

Compagnia 2					
Categoria	N. termini	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Obbligazioni	150	6m2s	0.12%	0.01%	0.01%
Azioni	150	7m33s	2.42%	0.82%	0.75%

Tabella 2.2: *Statistiche di bontà di previsione per gli attivi della Compagnia 2.*

Come per la Compagnia 1, il modello presenta ottime capacità previsive, in particolare per il portafoglio obbligazionario, ove sia l'errore massimo che quello medio sono quasi nulli.

Infine, nella tabella seguente sono riportate le analisi relative alla terza e ultima compagnia assicurativa in esame.

Compagnia 3					
Categoria	N. termini	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Obbligazioni	150	5m42s	0.14%	0.01%	0.01%
Azioni	150	6m40s	0.76%	0.12%	0.10%
Derivati	150	4m21s	1.51%	0.03%	0.01%

Tabella 2.3: *Statistiche di bontà di previsione per gli attivi della Compagnia 3.*

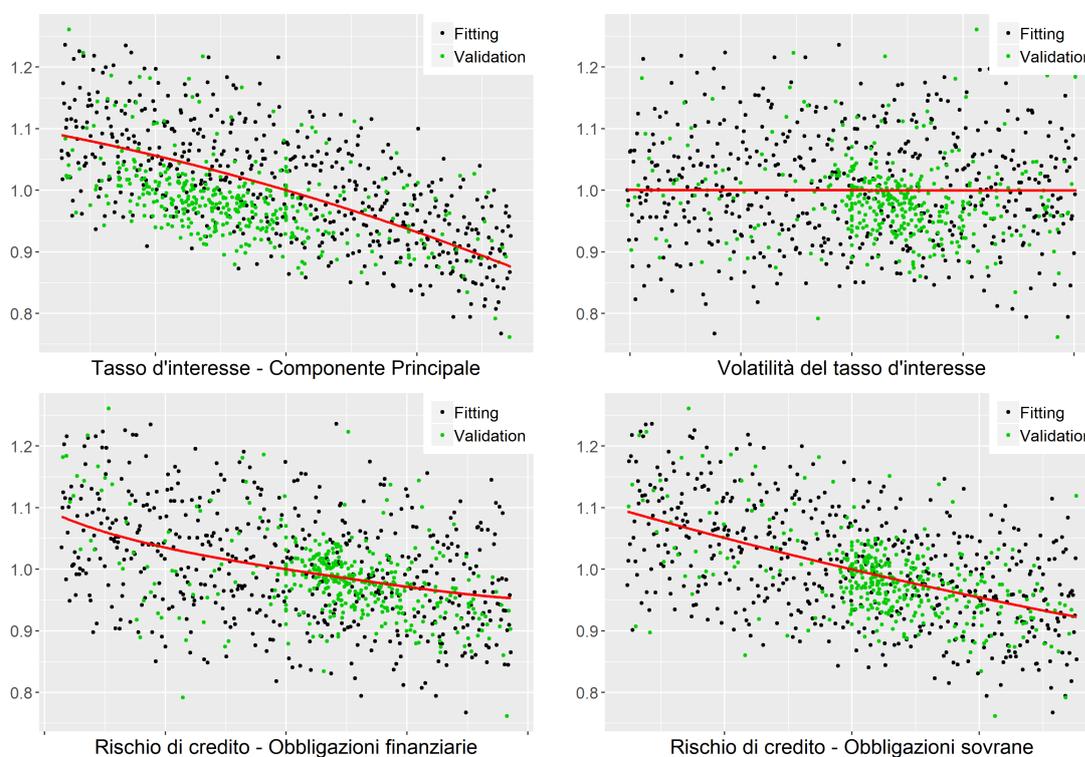
Il modello non evidenzia problemi per nessuno dei tre portafogli: l'errore medio è estremamente basso, l'errore massimo non raggiunge i 2% del valore del portafoglio nello scenario. Osserviamo tuttavia come anche qui l'algoritmo di selezione abbia sempre raggiunto il vincolo di 150 termini.

Procediamo con un'analisi grafica che mira a studiare come si comportano le marginali della funzione trovata rispetto ai fattori di rischio. In tal modo possiamo osservare l'andamento della funzione proxy al variare di un singolo fattore di rischio, tenendo fermi gli altri. Un'analisi grafica di questo tipo tuttavia trascura

l'effetto dei termini d'interazione, che giocano un ruolo molto importante nel modello (si veda, ad esempio, la Tabella 2.7). È necessaria quindi molta cautela nel confrontare l'andamento marginale della funzione in maniera diretta con i punti di fitting e di validazione riportati.

Brevitatis causa, riportiamo i risultati soltanto per alcuni dei portafogli considerati. Inoltre, solo i grafici relativi agli andamenti marginali più evidenti e interpretabili saranno riportati.

Compagnia 1 - Obbligazioni

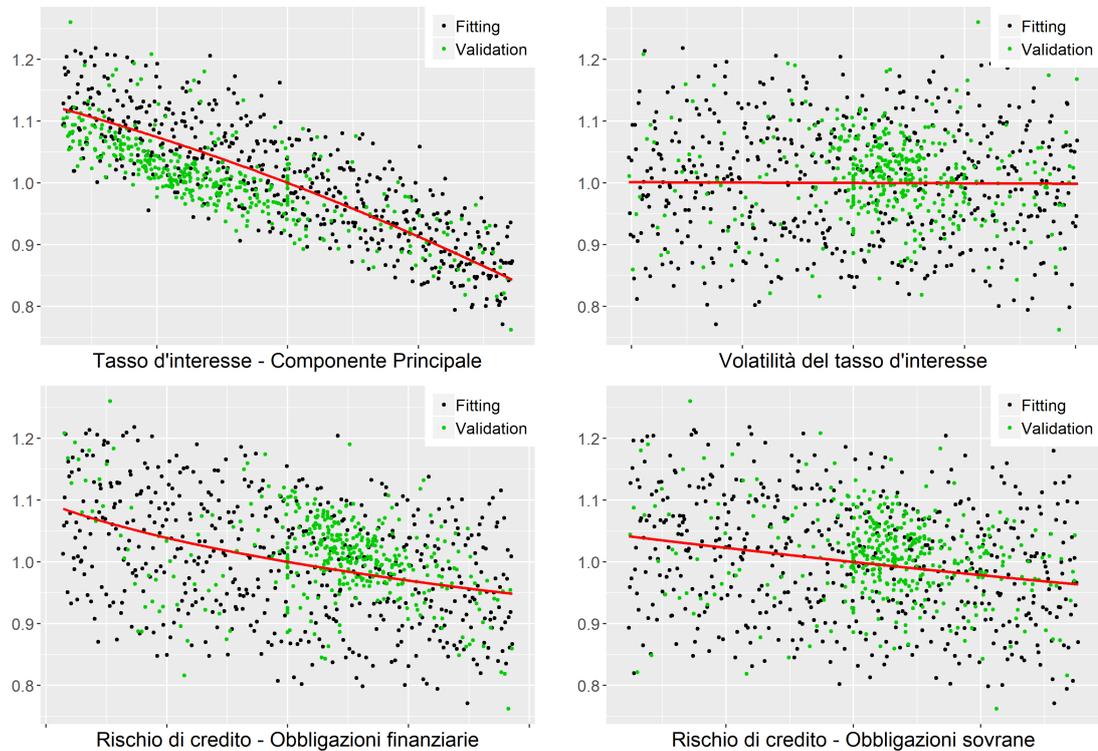


Andamento marginale della proxy rispetto ad alcuni dei fattori di rischio considerati; punti neri rappresentano gli scenari utilizzati per calibrare la funzione, punti verdi quelli su cui sono calcolate le metriche out-of-sample.

Come ci aspettavamo, al crescere della componente principale del tasso d'interesse e al crescere del fattore legato al rischio di credito, diminuisce il valore del portafoglio obbligazionario della compagnia. Tutti gli altri fattori inclusi nel modello - delineati con maggiore dettaglio nella sezione 2.3.1 - sembrano avere un effetto

piuttosto ridotto se considerati marginalmente (per brevità si è riportato il grafico soltanto per l'andamento rispetto alla volatilità del tasso d'interesse).

Compagnia 3 - Obbligazioni

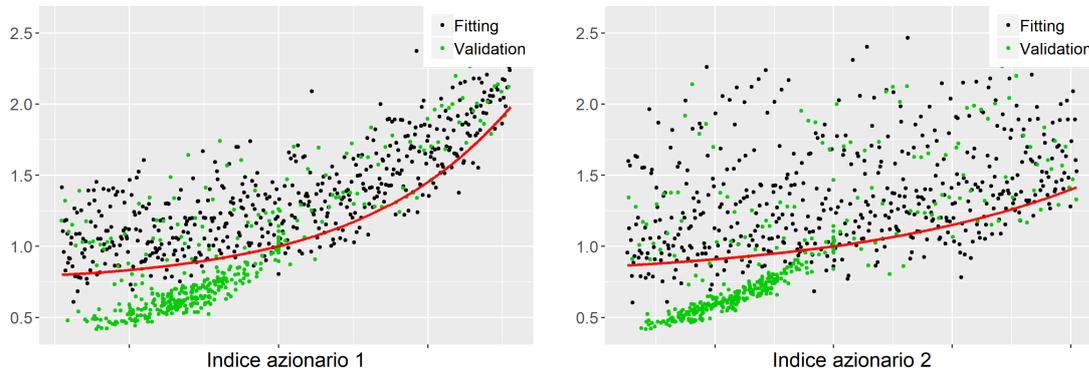


Andamento marginale della proxy rispetto ad alcuni dei fattori di rischio considerati; punti neri rappresentano gli scenari utilizzati per calibrare la funzione, punti verdi quelli su cui sono calcolate le metriche out-of-sample.

Come per la Compagnia 1, osserviamo un andamento decrescente del valore del portafoglio rispetto alla componente principale del tasso d'interesse e rispetto al rischio di credito, sebbene quest'ultima componente sia molto meno marcata, in particolare in relazione al rischio relativo alle obbligazioni sovrane. Anche qui le altre componenti marginalmente hanno un effetto molto ridotto.

Nella figura seguente andiamo invece ad analizzare l'andamento del portafoglio azionario della compagnia. In questo caso, la quasi totalità dei fattori di rischio è data da diversi indici azionari.

Compagnia 3 - Azioni



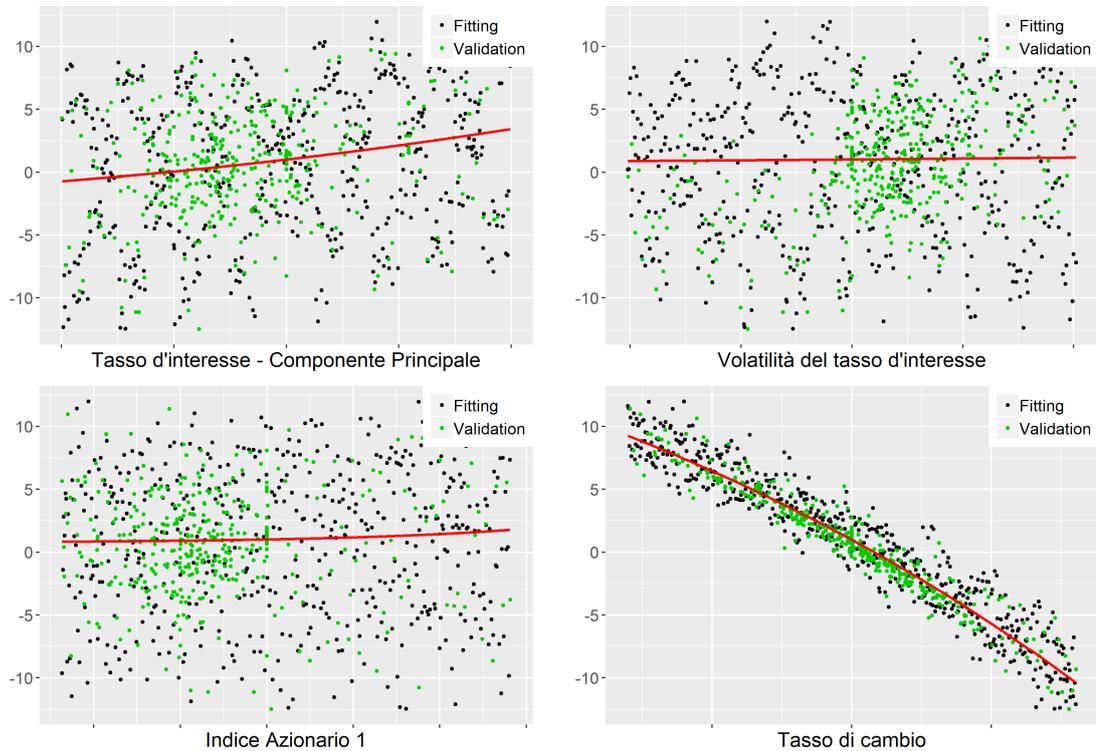
Andamento marginale della proxy rispetto ad alcuni dei fattori di rischio considerati; punti neri rappresentano gli scenari utilizzati per calibrare la funzione, punti verdi quelli su cui sono calcolate le metriche out-of-sample.

L'andamento marginale risulta intuitivo: al crescere del valore assunto dagli indici azionari aumenta il valore del portafoglio. La variabilità nella forma dell'andamento rispetto ai diversi indici è legata alla composizione del portafoglio azionario.

È particolarmente visibile una nuvola di scenari di validazione con valore basso degli indici e, anche, valore del portafoglio significativamente minore degli scenari utilizzati per il fitting. Ciò è dovuto al fatto che, mentre parte degli scenari di validazione sono stati generati tramite sequenze di Sobol con gli stessi principi degli scenari utilizzati per il fitting, una parte considerevole è stata selezionata a partire dalla coda della distribuzione multivariata.

Il fatto che la curva marginale di regressione fornisca valori minori di quelli che ci si aspetterebbe da una funzione di regressione univariata in un modello considerando i diversi regressori singolarmente è dovuto a più cause, tra cui, principalmente, l'interazione tra i diversi fattori di rischio e il fatto che lo stress nullo associato agli altri fattori di rischio nella curva marginale non sempre coincide con il valore medio sullo spazio di fitting.

Compagnia 3 - Derivati



Andamento marginale della proxy rispetto ad alcuni dei fattori di rischio considerati; punti neri rappresentano gli scenari utilizzati per calibrare la funzione, punti verdi quelli su cui sono calcolate le metriche out-of-sample.

Osserviamo un andamento crescente rispetto alla componente principale del tasso d'interesse, una funzione proxy marginale poco accentuata rispetto alle altre componenti. Di particolare nota l'andamento molto forte rispetto al tasso di cambio.

Infine, effettuiamo alcune analisi di sensitività sul modello così costruito. In particolare, abbiamo provveduto alla selezione del modello tramite un criterio basato sull'AIC. Ci si potrebbe chiedere quanto la funzione costruita sia sensibile a tale scelta arbitraria. Effettuiamo quindi una nuova calibrazione basata sul BIC per vedere come variano le statistiche considerate.

Compagnia 1					
Categoria	N. termini	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Obbligazioni	137	4m27s	0.42%	0.02%	0.02%
Azioni	113	3m4s	1.59%	0.26%	0.23%
Compagnia 2					
Categoria	N. termini	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Obbligazioni	150	6m26s	0.12%	0.01%	0.01%
Azioni	150	7m40s	2.42%	0.82%	0.75%
Compagnia 3					
Categoria	N. termini	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Obbligazioni	150	5m31s	1.51%	0.03%	0.01%
Azioni	150	6m34s	0.76%	0.12%	0.10%
Derivati	102	59s	2.15%	0.03%	0.01%

Tabella 2.4: *Statistiche di bontà di previsione per una calibrazione basata sul BIC.*

Come ci aspettavamo, essendo il BIC un criterio più penalizzante dell'AIC (come avevamo osservato nella sezione 2.1.4), l'algoritmo di selezione in questo caso tende ad arrestarsi prima, ottenendo un modello più parsimonioso. I valori delle metriche out-of-sample tendono ad essere leggermente migliorati, segno di un potenziale pericolo di overfitting nel modello basato sull'AIC. In tal senso, probabilmente, il BIC potrebbe essere, in questo contesto, un criterio più adeguato. In ogni caso, le differenze tra i risultati delle due calibrazioni non sono particolarmente accentuate.

Ricordiamo che per facilitare la convergenza dell'algoritmo di selezione dei regressori e evitare il rischio di overfitting, abbiamo imposto delle limitazioni al modello. Nel seguito viene studiato l'effetto di tali vincoli rispetto a modelli più completi, considerando in particolare la Compagnia 2, per la quale entrambi i portafogli hanno raggiunto il limite al numero di termini sia nella calibrazione con l'AIC che in quella con il BIC. Anche nella configurazione più completa abbiamo limitato il numero di termini a 500, in modo da forzare l'arresto dell'algoritmo qualora esso ancora non sia in grado di convergere autonomamente.

Compagnia 2					
Categoria	N. termini	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Obbligazioni	500	3h8m26s	0.10%	0.01%	0.01%
Azioni	169	9m46s	2.57%	0.82%	0.75%

Tabella 2.5: *Statistiche di bontà di previsione per una calibrazione considerante fino a 500 termini.*

Rispetto alla Tabella 2.2, per il portafoglio obbligazionario della compagnia notiamo un leggero miglioramento nelle metriche di validazione: come ci aspettavamo, infatti, il limite arbitrariamente fissato a 150 termini non contribuisce a migliorare la capacità previsiva del modello; il tempo di esecuzione, in questa configurazione, risulta tuttavia molto superiore. Nonostante il vincolo molto elevato a 500 termini, l'algoritmo di selezione forward continua a non arrestarsi. Per il portafoglio azionario invece l'arresto non vincolato avviene poco dopo il limite imposto manualmente. Le statistiche considerate sono in linea.

Nella Tabella 2.6 proviamo a togliere anche il vincolo al grado del polinomio. Tuttavia, visto il notevole sforzo computazionale, tale analisi sarà limitata al portafoglio obbligazionario della compagnia.

Compagnia 2					
Categoria	N. termini	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Obbligazioni	500	19h30m42s	0.07%	0.01%	0.01%

Tabella 2.6: *Statistiche di bontà di previsione per una calibrazione priva di vincoli al grado e considerante fino a 500 termini.*

Non limitando il grado del polinomio, le metriche di validazione migliorano ancora: nel modello LSMC semplificato che stiamo analizzando, il valore del portafoglio in ogni scenario esterno è calcolato in maniera precisa, rendendo il rischio di overfitting meno gravoso che nell'applicazione piena che vedremo per le passività. Il tempo d'esecuzione per un modello del genere tuttavia risulta molto pesante.

Nelle tabelle 2.7 e 2.8 verifichiamo anche l'effetto di alcuni vincoli più stringenti

che potrebbero essere scelti per mantenere il modello semplice e facilmente interpretabile. Una prima strada per semplificare il modello potrebbe essere quella di non includere effetti di interazione:

Compagnia 1					
Categoria	N. termini	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Obbligazioni	22	1s	2.26%	0.35%	0.35%
Azioni	46	4s	11.72%	1.94%	1.83%
Compagnia 2					
Categoria	N. termini	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Obbligazioni	32	1s	2.30%	0.25%	0.25%
Azioni	51	5s	10.74%	1.97%	1.90%
Compagnia 3					
Categoria	N. termini	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Obbligazioni	19	1s	2.36%	0.31%	0.31%
Azioni	59	7s	3.15%	0.64%	0.61%
Derivati	19	1s	28.15%	0.83%	0.32%

Tabella 2.7: *Statistiche di bontà di previsione per una calibrazione trascurante gli effetti di interazione.*

Trascurare gli effetti di interazione conduce a modelli molto più semplici e parsimoniosi, e il tempo di esecuzione diventa minimo. Tuttavia, i modelli così costruiti non sono sufficienti a spiegare la complessità del fenomeno: l'errore di previsione in media risulta infatti più di 10 volte superiore rispetto a quello riscontrato nelle tabelle 2.1, 2.2, 2.3. Ciò conferma l'intuizione che le interazioni tra i fattori di rischio sono una componente fondamentale del modello per il valore del fondo.

Nella Tabella 2.8 invece ammettiamo effetti di interazione, ma limitiamo i termini ammissibili ai monomi di primo e secondo grado. In altre parole, si suppone che un andamento al più quadratico rispetto ai fattori di rischio sia sufficiente a descrivere il fenomeno. Come la limitazione agli effetti di interazione, anche la limitazione al grado del polinomio conduce a modelli semplici, con tempi di calibrazione molto più bassi. Pure nella Tabella 2.8, però, gli indicatori di bontà del modello sono drasticamente peggiorati rispetto alla calibrazione di base, segno che l'utilizzo di

un polinomio di grado più elevato risulta importante per lo studio del fenomeno in esame.

Compagnia 1					
Categoria	N. termini	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Obbligazioni	51	9s	0.77%	0.12%	0.12%
Azioni	87	48s	15.88%	4.02%	3.05%
Compagnia 2					
Categoria	N. termini	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Obbligazioni	80	29s	0.87%	0.06%	0.06%
Azioni	137	2m41s	10.42%	2.34%	2.08%
Compagnia 3					
Categoria	N. termini	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Obbligazioni	70	19s	0.46%	0.07%	0.07%
Azioni	131	2m3s	16.01%	5.08%	4.29%
Derivati	32	2s	38.28%	0.68%	0.23%

Tabella 2.8: *Statistiche di bontà di previsione per una calibrazione limitata ad un polinomio di secondo grado.*

Dato il notevole aumento nell'errore di previsione, non considereremo oltre questi esempi semplificati.

Infine, procediamo ad un'implementazione del metodo MARS esposto nella sezione 2.2.3 tramite il pacchetto R *mda*. Mentre per il modello polinomiale si è dovuto procedere alla costruzione di un algoritmo apposito, un vantaggio distinto del modello MARS è la sua implementazione nella maggior parte dei moderni pacchetti di calcolo statistico. Si riportano gli esiti dell'esperimento esclusivamente per la configurazione considerante effetti di interazione al massimo fino al quarto grado, quindi coerente con la limitazione imposta alla funzione polinomiale.

Compagnia 1				
Categoria	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Obbligazioni	1s	1.43%	0.13%	0.13%
Azioni	2s	7.58%	3.53%	3.36%
Compagnia 2				
Categoria	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Obbligazioni	1s	1.06%	0.15%	0.15%
Azioni	2s	7.22%	1.79%	1.81%
Compagnia 3				
Categoria	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Obbligazioni	1s	0.70%	0.07%	0.07%
Azioni	2s	5.58%	1.64%	1.47%
Derivati	1s	14.62%	0.53%	0.30%

Tabella 2.9: *Statistiche di bontà di previsione per una calibrazione basata sul metodo MARS.*

Il modello MARS rappresenta un'alternativa rapida al modello polinomiale, in grado lo stesso di ottenere buone metriche sugli scenari di validazione (con l'eccezione dell'errore massimo sul portafoglio di derivati della Compagnia 1, che abbiamo già commentato relativamente alla Tabella 2.1, e che qui è ancora significativamente peggiorato). Rispetto a configurazioni altrettanto rapide dell'algorithmo di selezione polinomiale (Tabelle 2.7 e 2.8), il modello conduce a risultati migliori. Tuttavia, rispetto ad un giro più completo dell'algorithmo di selezione (tabelle 2.1, 2.2, 2.3), le statistiche considerate sono peggiorate. Ciò si suppone dovuto al fatto che alla base del modello MARS vi è comunque l'ipotesi di una relazione localmente lineare tra i regressori e la variabile dipendente, che nel nostro contesto spesso è violata. Notiamo che un processo più attento di tuning dei parametri del modello MARS può portare ad un miglioramento nelle statistiche riportate nella Tabella 2.9, tuttavia la performance del modello così ottenuto risulta in ogni caso sempre peggiore di quella del più tradizionale modello polinomiale. Uno spunto interessante di sviluppo potrebbe riguardare un modello ispirato al MARS, ma considerante spline quadratiche o cubiche, anziché lineari.

b) Modello LSMC per portafogli di passività assicurative vita

Procediamo ora all'implementazione piena del modello LSMC su portafogli di passività vita. Come per le attività, confronteremo i risultati del modello con quelli ottenuti in maniera precisa in scenari out-of-sample per le tre compagnie sotto esame. Questo giro consideriamo 25'000 dati imprecisi (ottenuti tramite 2 scenari interni antitetici) per la calibrazione, e due set di scenari di validazione, nei quali ogni dato è stato calcolato con 1'000 scenari interni. Come anticipato nella sezione 2.3.1, il primo set studia la validità globale della funzione, il secondo la validità in scenari estremi. Si noti tuttavia che mentre il primo set di validazione è determinabile fin da subito, e quindi valori particolarmente elevati delle relative metriche d'errore possono portare ad indagare oltre e ridefinire la funzione prima di procedere con i passi successivi del modello interno, il secondo set di validazione è determinabile solo una volta conclusa l'implementazione del modello e determinata la distribuzione della perdita: solo a quel punto, infatti, si possono conoscere gli scenari estremi che comportano una perdita vicina al quantile 99.5% della distribuzione. Valori elevati nel secondo set di validazione pertanto non possono essere corretti se non procedendo ad un nuovo giro completo del modello. Ricordiamo in ogni caso che anche 1'000 scenari interni potrebbero non essere sufficienti affinché il valore calcolato con il modello stocastico sia effettivamente interpretabile come il "vero" valore nello scenario: per effettuare un esperimento, si è scelto uno specifico scenario esterno (corrispondente allo stress nullo di tutti i fattori di rischio), nel quale oltre ad una valutazione con 1'000 scenari interni è stata fatta una valutazione con 16'000 scenari interni. Nella Tabella 2.10 è riportata la differenza relativa tra le BEL calcolate nello scenario.

Differenza tra BEL - 1'000 e 16'000 scenari	
	Compagnia 1
Pensioni	-0.10%
Protezione	-0.02%
Risparmio	-0.01%
Unit-Linked	0.00%

Tabella 2.10: *Differenza relativa tra BEL stimate con 1'000 e 16'000 scenari interni, per le diverse LoB della Compagnia 1.*

Siamo complessivamente soddisfatti del grado di precisione ottenuto con 1'000 simulazioni interne, considerato in particolare il già elevatissimo costo computazionale associato agli scenari di validazione. Allo stesso tempo, però, il potenziale termine d'errore ancora insito nella valutazione stocastica impone cautela nell'effettuare affermazioni conclusive riguardo alla validità del modello proxy, in particolare per quanto riguarda gli scenari del secondo set, in cui, essendo sulla coda, la variabilità nella stima del modello stocastico è più alta, e le differenze nella precisione potenzialmente maggiori di quelle riportate nella Tabella 2.10.

Nel seguito sono presentati i risultati relativi alla valutazione delle funzioni calibrate negli scenari di validazione. Come nell'esempio relativo ai portafogli di attività, per limitare il costo computazionale abbiamo considerato soltanto termini fino al quarto grado e abbiamo forzato l'arresto dell'algoritmo quando il numero di termini raggiunge 150. Anche qui, l'impatto sarà nullo tranne nei casi in cui l'algoritmo di selezione fatica a convergere, e l'effetto di questo vincolo sulla funzione calibrata sarà successivamente analizzato.

Compagnia 1								
Categoria	N. termini	T. esecuzione	Set di validazione 1			Set di validazione 2		
			\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Pensioni	39	1m3s	1.59%	0.59%	0.55%	4.11%	1.48%	1.50%
Protezione	130	28m44s	0.08%	0.02%	0.02%	4.42%	1.34%	1.34%
Risparmio	147	58m24s	0.44%	0.16%	0.15%	7.27%	1.91%	1.88%
Unit-Linked	13	4s	2.98%	0.70%	0.61%	2.98%	0.38%	0.34%

Tabella 2.11: *Statistiche di bontà di previsione per i passivi della Compagnia 1.*

Nella Tabella 2.11 osserviamo come il modello conduce a risultati relativamente buoni per tutte e quattro le LoB considerate. L'errore massimo commesso nel set di valutazione della bontà complessiva è sempre inferiore al 5%, con il MAE che non raggiunge nemmeno l'1%. Dal secondo set di validazione si desume che la funzione polinomiale riesce a catturare in maniera piuttosto precisa anche valori calcolati in scenari particolarmente estremi, con un errore assoluto medio che si aggira tra 1% e 2%. Notiamo ancora come vi sia una forte divergenza tra il numero di termini scelto dall'algoritmo di selezione per i vari portafogli: mentre per alcune LoB

pochi termini sono stati sufficienti a catturare l’andamento della BEL, per altre il numero di termini selezionato raggiunge quasi il limite di 150 che ci siamo imposti all’inizio. Si noti poi come l’algoritmo sviluppato (esposto nell’appendice B) risulta estremamente efficiente per un numero limitato di termini (<100), mentre si ha una crescita significativa nel tempo di esecuzione quando l’algoritmo di selezione fatica ad arrestarsi.

Compagnia 2								
			Set di validazione 1			Set di validazione 2		
Categoria	N. termini	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Pensioni	31	25s	1.31%	0.36%	0.36%	/	/	/
Protezione	77	6m33s	1.67%	0.27%	0.24%	/	/	/
Unit-Linked	43	1m12s	3.28%	0.59%	0.59%	/	/	/

Tabella 2.12: *Statistiche di bontà di previsione per i passivi della Compagnia 2.*

Per la seconda compagnia in esame, i dati relativi al secondo set di validazione non sono disponibili. Studieremo pertanto esclusivamente i risultati relativi al set sulla validità generale della funzione proxy, non quello relativo ai valori estremi. Osserviamo come qui gli algoritmi di selezione non faticino ad arrestarsi, completando l’esecuzione in pochi minuti. Inoltre, le statistiche di valutazione sono molto buone, con un errore assoluto medio sempre significativamente minore dell’1%.

Compagnia 3								
			Set di validazione 1			Set di validazione 2		
Categoria	N. termini	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Protezione	150	1h14m41s	235.94%	2.67%	7.45%	311.87%	18.53%	24.89%
Risparmio	150	1h5m23s	0.56%	0.30%	0.30%	12.35%	1.46%	1.47%

Tabella 2.13: *Statistiche di bontà di previsione per i passivi della Compagnia 3.*

Rispetto alle compagnie precedenti, osserviamo innanzitutto come gli algoritmi di selezione faticino a convergere per entrambe le LoB considerate, raggiungendo il limite imposto manualmente di 150 termini. Per la LoB “Protezione” inoltre, os-

serviamo come il modello si riveli incapace di cogliere adeguatamente l'andamento della BEL. Vi sono diversi fattori che spiegano un risultato così estremo:

- La BEL relativa alla LoB ha valore prossimo allo 0, per cui l'errore relativo risulta enorme anche quando l'errore effettivamente commesso è minimo in termini monetari, il che comporta la necessità, per un'implementazione pratica del modello, di considerare anche le metriche assolute esposte nella sezione 2.1.5, che noi abbiamo trascurato in questa analisi.
- In entrambi i set di validazione è stato inserito uno scenario particolarmente estremo, con valore del rischio di mortalità catastrofica significativamente superiore al massimo valore utilizzato per il fitting. Si rileva quindi la potenziale incapacità della funzione polinomiale di cogliere adeguatamente scenari al di fuori del range utilizzato per calibrarla, un risultato ben noto in letteratura. Rimosso lo scenario anomalo dai set di validazione, le statistiche tornano in linea: le metriche \mathcal{M}_1 e \mathcal{M}_3 sul primo set, ad esempio, assumono rispettivamente i valori 5.92% e 0.47%. Nelle analisi seguenti, continueremo a considerare anche l'outlier nei set, in modo da verificare l'andamento dei modelli in una condizione così estrema.

Al di là della LoB Protezione della Compagnia 3, il modello ha fornito risultati estremamente positivi, con un errore massimo di 1-3% su un singolo scenario e un errore assoluto medio globale sempre significativamente minore dell'1%. Su blocchi estremi della distribuzione congiunta, l'errore assoluto medio rimane comunque sempre inferiore al 2%. Anche per le polizze di protezione della Compagnia 3, le uniche in cui gli errori considerati risultano significativi, abbiamo individuato il problema e, come vedremo, possiamo intervenire manualmente per migliorare l'adattamento. Per quanto infatti l'utilizzo di algoritmi automatici di selezione si mostri generalmente molto efficace, e permetta in maniera rapida di modellare anche un gran numero di portafogli, con un'analisi statistica specifica si possono sempre ottenere risultati migliori. Tale è infatti anche l'utilità del dedicare una parte considerevole del budget computazionale agli scenari di validazione: eventuali problematiche riscontrate possono essere repentinamente corrette.

Ricordiamo che per verificare la bontà della funzione stiamo considerando metriche in valore assoluto; in realtà però è plausibile che vi siano degli effetti di compensazione tra scenari con sovrastima e scenari con sottostima della perdita. Nella Tabella 2.14 consideriamo la percentuale di scenari in cui vi è sottostima e l'errore medio relativo sul set di validazione 2, quindi sulla coda della distribuzione.

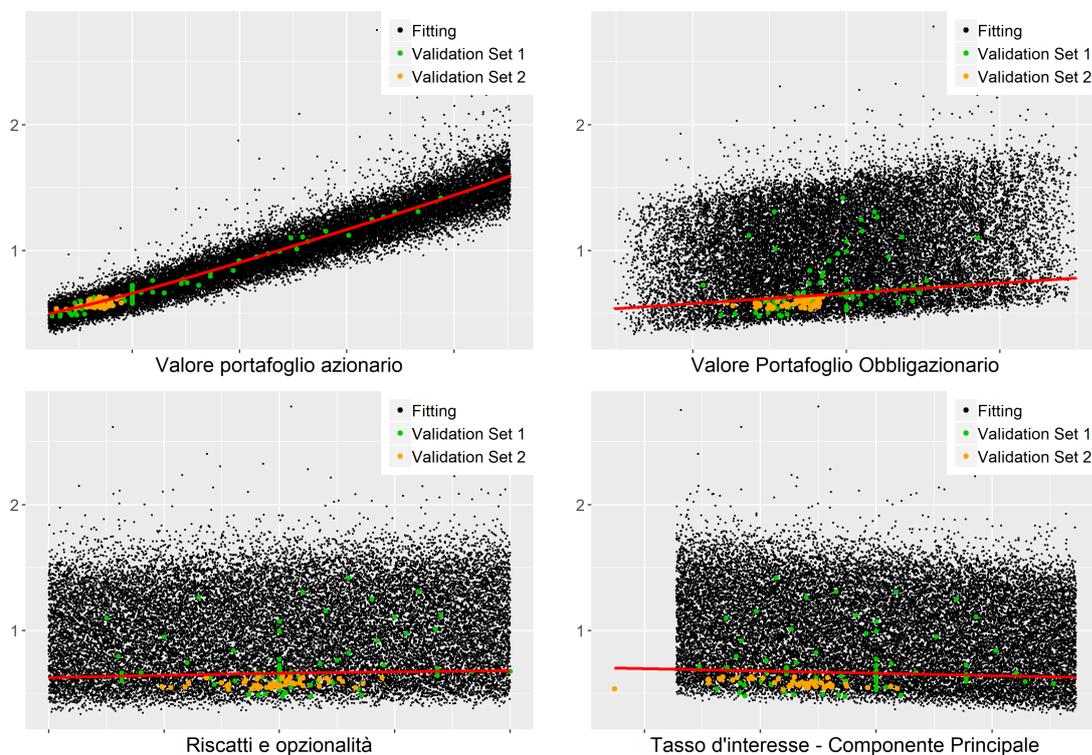
Compagnia 3 - Set di validazione 2		
Categoria	Quota di scenari con sottostima	Errore medio relativo
Protezione	60%	-1.52%
Risparmio	46%	0.06%

Tabella 2.14: *Percentuale di scenari con sottostima e errore medio relativo sulla coda della distribuzione, calcolate sul set di validazione 2 per le LoB della compagnia 3.*

Osserviamo come la quota di scenari di validazione in cui la proxy sottostima il valore è relativamente vicina al 50% per entrambe le LoB. Inoltre, l'errore medio è significativamente minore della corrispondente metrica assoluta \mathcal{M}_2 , per cui il modello non sembra attuare una drastica sottostima o sovrastima della coda.

Nelle figure seguenti, l'andamento marginale rispetto ai principali fattori di rischio è analizzato per ogni portafoglio delle compagnie 1 e 3. In tal modo possiamo osservare l'andamento della funzione proxy al variare di un singolo fattore di rischio, tenendo fermi gli altri. Come avevamo osservato già per le attività, un'analisi grafica di questo tipo trascura l'effetto dei termini d'interazione, che giocano un ruolo molto importante nel modello (si confronti la Tabella 2.17 con le Tabelle 2.11, 2.12, 2.13). La curva marginale della funzione non potrà quindi essere direttamente comparata né con i punti di fitting, né con quelli di validazione, in particolare in relazione all'andamento rispetto al valore del portafoglio obbligazionario: la curva marginale infatti cattura l'andamento puro rispetto al valore del portafoglio, mentre i punti di fitting avranno andamenti legati alla loro correlazione con gli altri fattori di rischio. I grafici quindi sono riportati al fine di evidenziare quali possono essere gli andamenti delle BEL, quali le regioni più critiche nella distribuzione multivariata, quali gli aspetti colti dalla funzione calibrata, ma da essi non si possono davvero trarre indicazioni sulla bontà del modello complessivo.

Compagnia 1 - Pensioni



Andamento marginale della proxy costruita per il valore della BEL per le pensioni della prima compagnia, rispetto ad alcuni dei fattori di rischio considerati; punti neri rappresentano gli scenari utilizzati per calibrare la funzione, i punti verdi e arancioni quelli su cui sono calcolate le metriche di validazione.

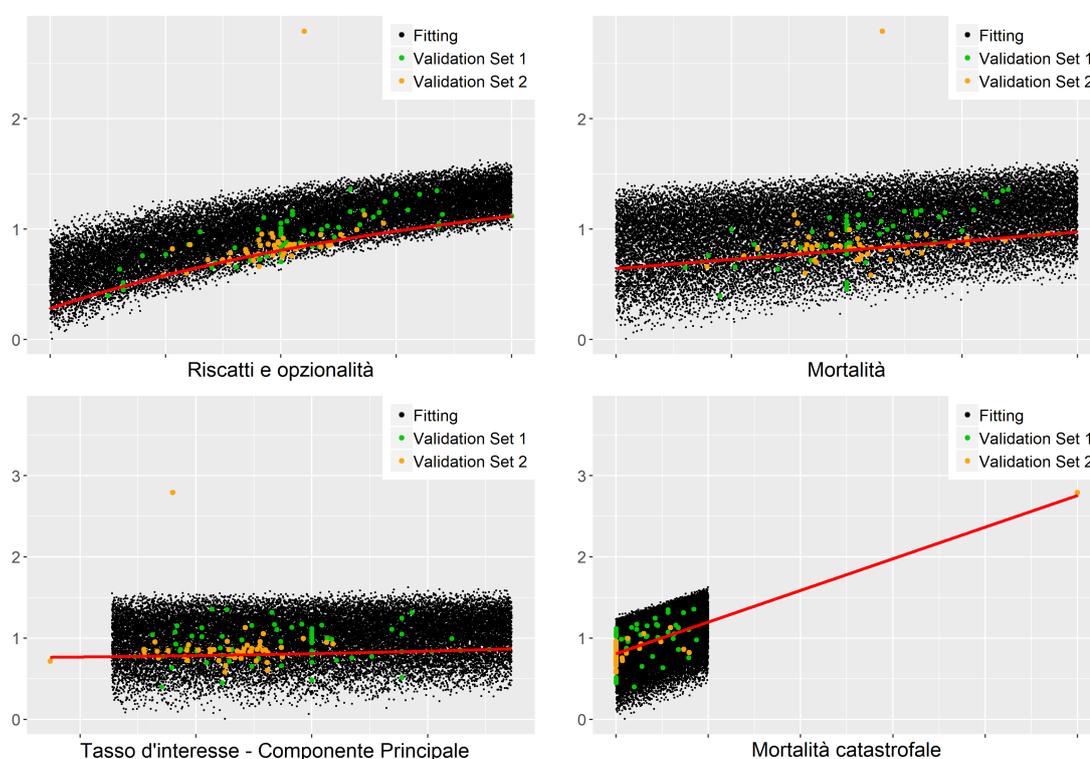
Marginalmente, vi sembra essere una forte dipendenza dal valore del portafoglio azionario della compagnia, mentre le altre variabili considerate hanno un effetto molto ridotto. Osserviamo poi come la curva marginale del modello sembra vicina ai punti di validazione e come essa sembra fornire una stima conservativa rispetto alla massa degli scenari di fitting, e quindi rispetto a quella che sarebbe la funzione di regressione univariata che potrebbe essere calibrata marginalmente esclusivamente rispetto allo specifico fattore di rischio.

Come già osservato per gli attivi, questa discrepanza tra curva marginale e funzione di regressione univariata è legata a più fattori: gli effetti di interazione tra i fattori di rischio e il fatto che la curva marginale per un regressore è ottenuta associando valore zero alle altre variabili inserite nel modello, valore che non sempre coincide con la loro media sul range di fitting.

Osserviamo che, in questo modo, la curva marginale risulta vicina agli scenari estremi del set di validazione 2, rappresentante realizzazioni vicine al quantile 99.5% della distribuzione multivariata.

In ultimo, osserviamo come per la componente principale del rischio di tasso d'interesse uno degli scenari di validazione cada al di fuori del range utilizzato per il fitting della proxy. Come avevamo già notato, ciò può risultare critico per gli indicatori di bontà del modello: nulla garantisce infatti che l'andamento estrapolato su un dato range di dati sia valido in un punto lontano al di fuori del range selezionato. La forma delle funzioni polinomiali, in particolare di grado più elevato, accentua questo rischio.

Compagnia 1 - Protezione

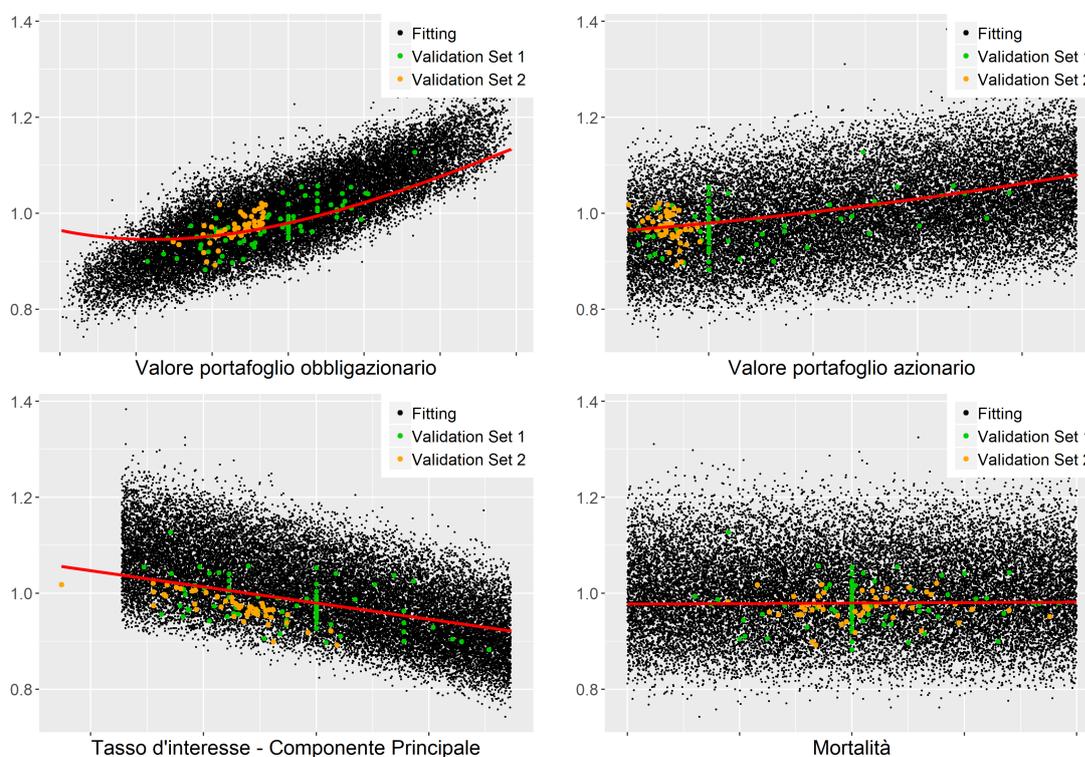


Andamento marginale della proxy costruita per il valore della BEL per la LoB Protezione della prima compagnia, rispetto ad alcuni dei fattori di rischio considerati; punti neri rappresentano gli scenari utilizzati per calibrare la funzione, i punti verdi e arancioni quelli su cui sono calcolate le metriche di validazione.

Per il portafoglio relativo a polizze di protezione, valgono le stesse considerazioni effettuate per la LoB precedente. Ricordato che la LoB protezione detiene

principalmente contratti con beneficio primario in caso di morte dell'assicurato, riteniamo coerente l'andamento crescente della BEL al crescere della mortalità. In particolare, relativamente alla mortalità catastrofica, osserviamo la presenza di un punto di validazione estremamente lontano dallo spazio di fitting. Una situazione così estrema è dovuta alla peculiarità della distribuzione real-world utilizzata per questo fattore di rischio, che non è continua e per quantili estremamente elevati assume valori catastrofici. Dal grafico della curva marginale corrispondente si vede come l'andamento lineare coglie comunque bene anche lo scenario estremo. In questo caso infatti gli indicatori di bontà nella Tabella 2.11 non presentano anomalie.

Compagnia 1 - Risparmio



Andamento marginale della proxy costruita per il valore della BEL per la LoB Risparmio della prima compagnia, rispetto ad alcuni dei fattori di rischio considerati; punti neri rappresentano gli scenari utilizzati per calibrare la funzione, i punti verdi e arancioni quelli su cui sono calcolate le metriche di validazione.

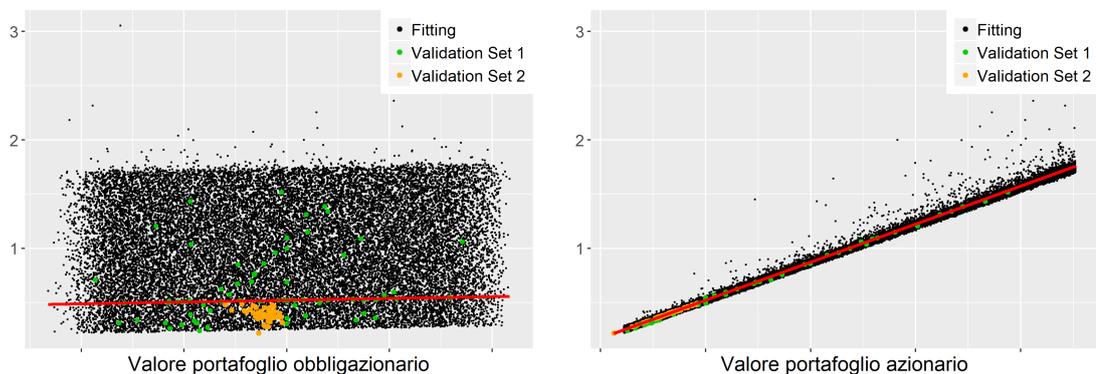
Rispetto al portafoglio di polizze di protezione, quello legato alle polizze di risparmio risulta di nuovo principalmente esposto ai fattori di rischio di natura finanziaria

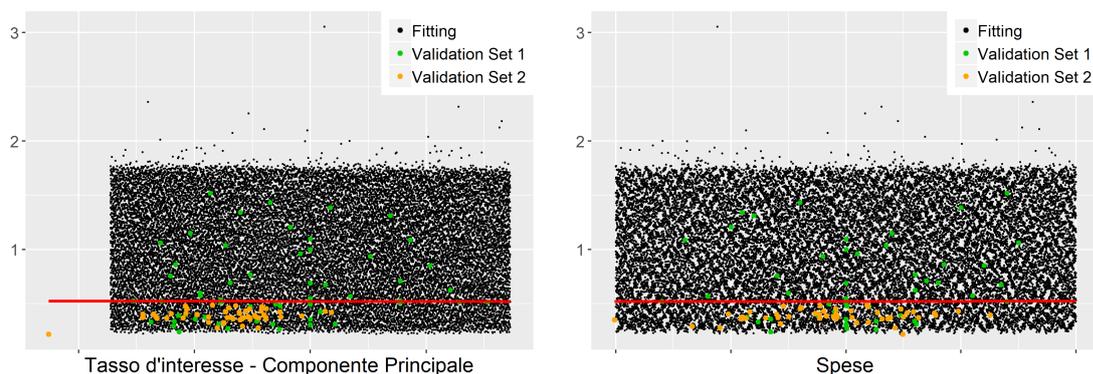
e legati all'andamento delle attività. Fattori di rischio biometrici, quali ad esempio la mortalità, hanno un impatto del tutto marginale rispetto a quello relativo all'andamento degli attivi e ai rischi finanziari. Trattandosi spesso di contratti con beneficio garantito, in cui la componente di mortalità provoca semplicemente una sua accelerazione temporale, ciò non dovrebbe stupire.

Rispetto a quanto osservato per la LoB relativa alle pensioni della Compagnia 1, notiamo qui una dipendenza molto più forte dal portafoglio obbligazionario piuttosto che da quello azionario dell'impresa, dovuta alla diversa strategia di investimento degli attivi corrispondenti. Osserviamo poi come la curva marginale assuma un andamento non propriamente in linea rispetto ai valori di fitting del portafoglio obbligazionario. Ciò è dovuto alla correlazione presente tra l'andamento del tasso d'interesse e il rischio di credito, entrambi influenti sul valore del portafoglio obbligazionario dell'impresa.

Osserviamo ancora che al crescere del fattore legato al tasso d'interesse decrescono sia il valore del portafoglio obbligazionario (come abbiamo visto nell'esempio relativo ai portafogli di attività), sia il valore della BEL (gli importi futuri della BEL infatti vengono scontati maggiormente). Al crescere del valore del portafoglio obbligazionario, aumenta tendenzialmente anche il valore della BEL a causa dei meccanismi di partecipazione agli utili solitamente insiti in contratti di questo tipo. La presenza di garanzie finanziarie per gli assicurati porta l'andamento a discostarsi dalla linearità: per valori bassi del portafoglio d'investimento dell'assicuratore, entra in vigore la garanzia e il valore della BEL non scende più proporzionalmente.

Compagnia 1 - Unit-Linked

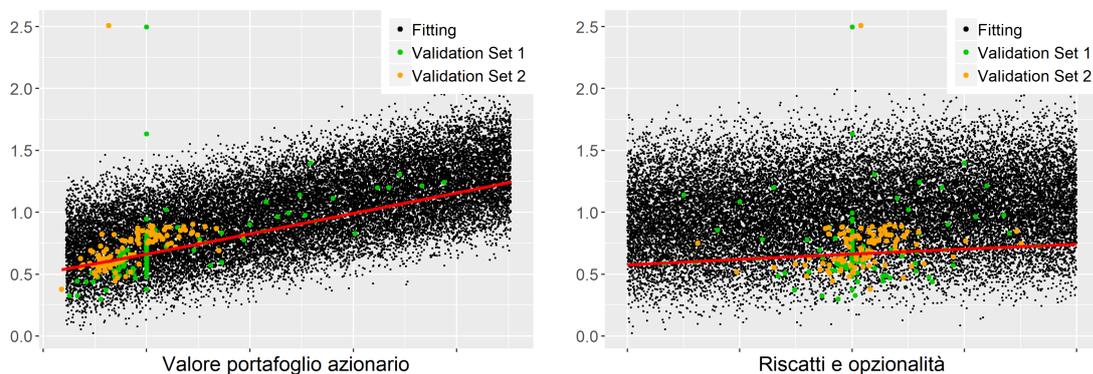


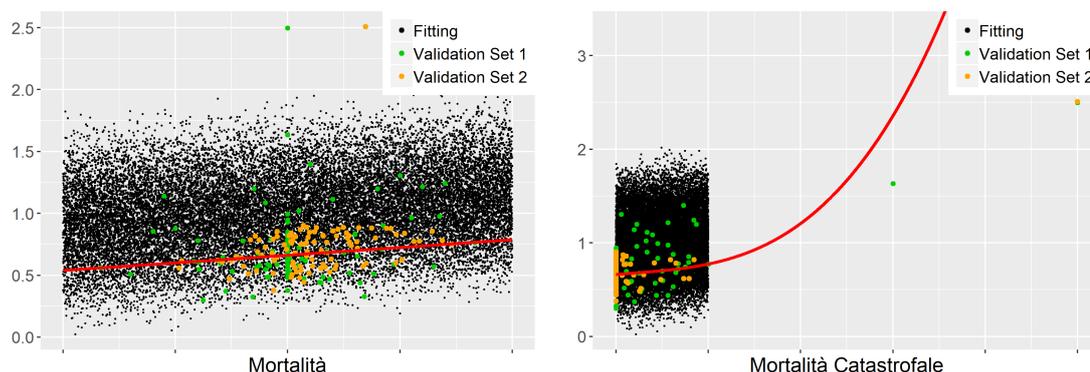


Andamento marginale della proxy costruita per il valore della BEL per la LoB Unit-Linked della prima compagnia, rispetto ad alcuni dei fattori di rischio considerati; punti neri rappresentano gli scenari utilizzati per calibrare la funzione, i punti verdi e arancioni quelli su cui sono calcolate le metriche di validazione.

Per le Unit-Linked, osserviamo come il valore abbia una dipendenza lineare fortissima dal valore del portafoglio azionario, mentre gli altri fattori risultano poco significativi. Ciò non dovrebbe stupire, data la natura dei contratti di tipo unit-linked: il rendimento corrisposto agli assicurati è direttamente legato al rendimento di un fondo d'investimento. Il rischio d'investimento qui è pienamente a carico degli assicurati, non sono nemmeno presenti effetti di non-linearità legati a garanzie e opzionalità (si tratta, infatti, di polizze unit-linked senza garanzie finanziarie). La maggiore semplicità nello spiegare l'andamento di un portafoglio di questo tipo si evince anche dalla Tabella 2.11, ove si vede che l'algoritmo ha ritenuto sufficienti solo 13 termini per sviluppare la funzione proxy.

Compagnia 3 - Protezione

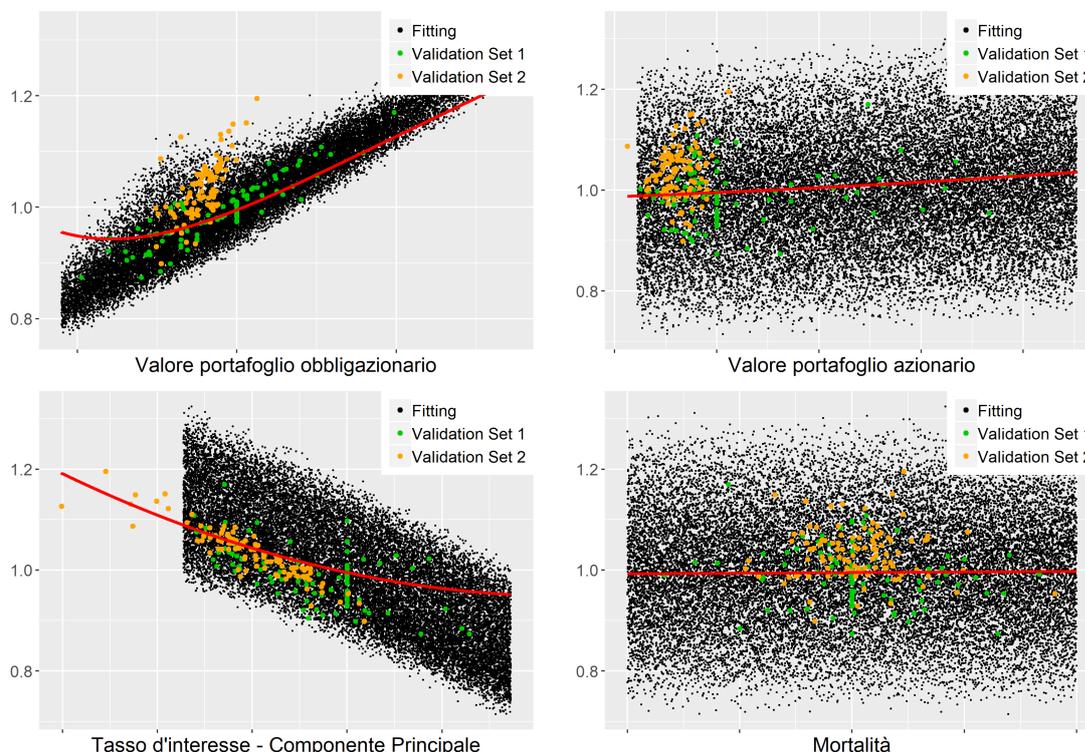




Andamento marginale della proxy costruita per il valore della BEL per la LoB Protezione della terza compagnia, rispetto ad alcuni dei fattori di rischio considerati; punti neri rappresentano gli scenari utilizzati per calibrare la funzione, i punti verdi e arancioni quelli su cui sono calcolate le metriche di validazione.

Rispetto alle polizze di protezione della Compagnia 1, osserviamo una tendenza meno netta rispetto all'evoluzione dei riscatti, mentre assume rilevanza l'andamento del portafoglio azionario dell'impresa, segno che per questa compagnia sono presenti meccanismi di partecipazione agli utili anche per contratti di questo tipo. Risulta interessante osservare come l'algoritmo di selezione dei regressori ha scelto una funzione non lineare per l'andamento marginale rispetto alla mortalità catastrofica e tale funzione risulta significativamente diversa dal valore osservato al di fuori del range utilizzato per il fitting: qui nello scenario estremo l'errore commesso dalla funzione proxy è considerevole, come avevamo già commentato relativamente alla Tabella 2.13. Il grafico della curva marginale potrebbe anche darci un'indicazione su come migliorare le statistiche della capacità di previsione riportate nella Tabella 2.13: si osserva infatti che l'andamento rispetto alla mortalità catastrofica sembra proseguire linearmente al di fuori dello spazio utilizzato per il fitting. Si potrebbe allora pensare di forzare l'algoritmo a non considerare le potenze del fattore di rischio legato alla mortalità catastrofica, limitandosi quindi ad un andamento lineare. Una scelta di questo tipo riuscirebbe a ridurre significativamente l'errore di previsione commesso, riportandolo in linea con i risultati ottenuti per le altre compagnie. In questo modo, con un aggiustamento manuale si possono quindi facilmente correggere eventuali problemi nell'algoritmo di selezione.

Compagnia 3 - Risparmio



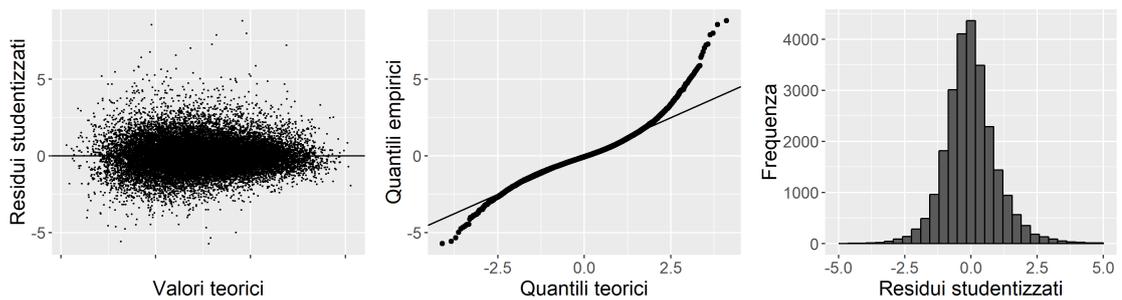
Andamento marginale della proxy costruita per il valore della BEL per la LoB Protezione della terza compagnia, rispetto ad alcuni dei fattori di rischio considerati; punti neri rappresentano gli scenari utilizzati per calibrare la funzione, i punti verdi e arancioni quelli su cui sono calcolate le metriche di validazione.

Le BEL delle polizze di risparmio della Compagnia 3, rispetto a quelle della Compagnia 1, risultano molto più legate all'andamento del portafoglio obbligazionario e molto meno a quello del portafoglio azionario. Anche qui si evidenzia la particolarità dell'andamento della curva marginale relativa all'andamento del valore della BEL rispetto al valore del portafoglio obbligazionario, del quale abbiamo già parlato in merito al portafoglio della Compagnia 1. Le garanzie finanziarie presenti nelle polizze si traducono in un andamento convesso del valore della BEL.

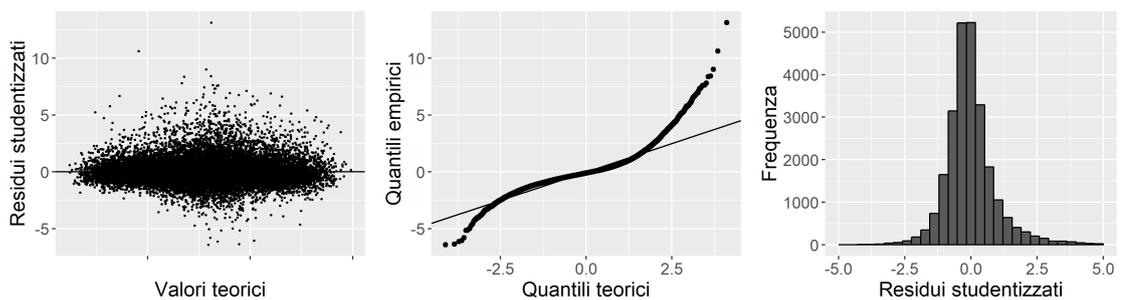
In ultimo, si ritiene doveroso un commento relativo al motivo per il quale la dipendenza dai fattori di rischio biometrici risulti così poco marcata: un fatto forse a prima vista controintuitivo, in particolare per le BEL che ad essi dovrebbero essere maggiormente legate, quali quelle delle polizze di protezione, ove, tendenzialmente, non è presente o è presente in misura molto ridotta l'accumulo della riserva

matematica. Tale fenomeno si spiega osservando che il fattore di rischio considerato rappresenta, generalmente, la variazione ad un anno del rischio sottostante: fenomeni biometrici quali la mortalità risultano, generalmente, molto più stabili nel breve termine rispetto al mondo della finanza, per cui le variazioni considerate nel modello rimangono più contenute. Un discorso leggermente diverso vale per il fattore relativo alla mortalità catastrofica, che rappresenta una variazione estrema della mortalità osservata nell'anno e pertanto conduce ad una dipendenza osservata molto più netta rispetto ad altri fattori biometrici. In particolare per questo fattore di rischio abbiamo osservato più volte come gli scenari di validazione possano cadere molto al di fuori del range utilizzato per il fitting, il che evidenzia una fortissima asimmetria nella distribuzione real-world del fattore di rischio. Nel seguito, si forniscono alcune analisi relative ai residui del modello. Per brevità, esse saranno riportate solo per le LoB della Compagnia 3:

Compagnia 3 - Protezione



Compagnia 3 - Risparmio



Alcuni grafici diagnostici del modello per le LoB della Compagnia 3.

Il modello presenta residui simmetrici rispetto allo 0, con un andamento piuttosto casuale e di entità non estrema, anche se leggermente superiore a quanto ci

si aspetterebbe in caso di normalità. Le code osservate infatti sono più pesanti rispetto a quelle della normale; leggere forme di eteroschedasticità possono essere osservate nel grafico dei residui contro i valori teorici. Ricordato che, come abbiamo osservato nella sezione 2.1.1, le ipotesi di normalità e omoschedasticità non sono fondamentali per la stima OLS - che rimane consistente anche in loro assenza - siamo comunque soddisfatti dell'adattamento ottenuto dal modello. Siamo tuttavia consapevoli che le deviazioni dalla normalità o dall'omoschedasticità possono influenzare alcuni aspetti importanti quali il procedimento di selezione dei regressori.

Come per i portafogli di attività, procediamo ora ad analizzare l'effetto di un criterio alternativo di arresto per l'algoritmo di selezione. In particolare, nella Tabella 2.15 la calibrazione sarà effettuata con il BIC anziché con l'AIC.

Compagnia 1								
Categoria	N. termini	T. esecuzione	Set di validazione 1			Set di validazione 2		
			\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Pensioni	21	13s	1.30%	0.53%	0.51%	4.62%	2.02%	2.02%
Protezione	85	7m39s	0.07%	0.02%	0.02%	4.41%	1.33%	1.33%
Risparmio	60	4m11s	0.68%	0.18%	0.18%	7.17%	1.83%	1.80%
Unit-Linked	8	1s	2.81%	0.67%	0.58%	2.49%	0.44%	0.40%

Compagnia 2								
Categoria	N. termini	T. esecuzione	Set di validazione 1			Set di validazione 2		
			\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Pensioni	16	5s	1.37%	0.39%	0.40%	/	/	/
Protezione	40	1m4s	0.78%	0.23%	0.22%	/	/	/
Unit-Linked	22	13s	3.50%	0.57%	0.57%	/	/	/

Compagnia 3								
Categoria	N. termini	T. esecuzione	Set di validazione 1			Set di validazione 2		
			\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Protezione	83	13m6s	223.16%	2.56%	7.04%	298.44%	18.40%	24.40%
Risparmio	107	30m1s	0.67%	0.33%	0.33%	11.31%	1.43%	1.43%

Tabella 2.15: *Statistiche di bontà di previsione per le passività con una calibrazione basata sul BIC.*

Come attendibile, il criterio basato sul BIC seleziona un numero significativamente minore di regressori: avendo un numero elevato di dati, la penalizzazione alla verosimiglianza effettuata dal BIC per ogni nuovo regressore è significativamente maggiore della corrispondente limitazione dell'AIC (circa 5 volte tanto), portando l'algoritmo di selezione ad arrestarsi molto prima. L'utilizzo del BIC riduce il rischio di overfitting, ma potrebbe condurre a modelli troppo semplici.

I risultati ottenuti nella Tabella 2.15 non sono conclusivi: in alcuni casi una calibrazione basata sul BIC fornisce risultati leggermente migliori, in altri leggermente peggiori. La difficoltà nell'effettuare in maniera generale una scelta ottimale relativamente al criterio da utilizzare per la selezione dei regressori rappresenta una delle limitazioni del modello.

Nella tabella seguente, invece, si procede con un'analisi relativa ai vincoli imposti all'algoritmo di selezione dei regressori. In particolare, si rimuoverà il vincolo al grado del polinomio (prima il grado massimo considerato era il quarto) e si aumenterà il numero massimo di termini della funzione da 150 a 250. Dato l'onere computazionale di un'analisi di questo tipo, essa sarà effettuata soltanto per le LoB della Compagnia 3, che nella Tabella 2.13 avevano entrambe raggiunto il limite di 150 termini.

Compagnia 3								
			Set di validazione 1			Set di validazione 2		
Categoria	N. termini	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Protezione	199	4h13m31s	294%	3.14%	9.21%	510%	20.69%	31.82%
Risparmio	211	4h48m15s	0.56%	0.25%	0.26%	12.05%	1.45%	1.45%

Tabella 2.16: *Statistiche di bontà di previsione per una calibrazione più completa per la Compagnia 3.*

Rispetto alla corrispondente analisi relativa ai portafogli di attività, ove avevamo notato un miglioramento riducendo i vincoli, qui nelle statistiche utilizzate per misurare la bontà del modello osserviamo un peggioramento per i contratti di protezione, un miglioramento estremamente ridotto per quelle di risparmio. Come avevamo preannunciato, infatti, nella forma classica del modello LSMC che stiamo

utilizzando per le passività, i dati disponibili sono calcolati in maniera poco precisa, per cui il rischio di overfitting è notevole: l’algoritmo di selezione quindi tende a catturare anche una parte dell’errore statistico insito nei dati. In questo caso, la limitazione arbitraria cui avevamo soggetto il modello ha dunque condotto ad un miglioramento nelle metriche considerate, ma non si ritiene che i risultati ottenuti nei due casi considerati siano sufficienti a presumere che la cosa sia valida anche per altri portafogli. Purtroppo, non vi sono regole precise relative all’ottimalità dei vincoli, la cui bontà dipenderà da portafoglio a portafoglio. Come nella nostra applicazione, essi saranno principalmente dettati da questioni relative alla velocità richiesta all’esecuzione dell’algoritmo.

In maniera analoga a quanto svolto per le attività, nelle tabelle seguenti vediamo anche l’effetto di calibrazioni con vincoli più restrittivi all’algoritmo di selezione.

Compagnia 1								
			Set di validazione 1			Set di validazione 2		
Categoria	N. termini	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Pensioni	16	2s	5.20%	1.31%	1.17%	8.73%	3.71%	3.71%
Protezione	16	2s	4.24%	0.61%	0.56%	4.35%	1.51%	1.46%
Risparmio	25	6s	3.60%	0.84%	0.81%	4.37%	1.41%	1.42%
Unit-Linked	9	1s	4.41%	0.98%	0.82%	1.32%	0.68%	0.66%

Compagnia 2								
			Set di validazione 1			Set di validazione 2		
Categoria	N. termini	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Pensioni	12	1s	2.38%	0.64%	0.64%	/	/	/
Protezione	17	2s	9.58%	1.11%	0.96%	/	/	/
Unit-Linked	16	3s	4.85%	1.39%	1.37%	/	/	/

Compagnia 3								
			Set di validazione 1			Set di validazione 2		
Categoria	N. termini	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Protezione	27	7s	64.16%	2.43%	3.41%	57.69%	14.59%	15.06%
Risparmio	25	6s	4.09%	1.09%	1.09%	9.12%	2.83%	2.91%

Tabella 2.17: *Statistiche di bontà di previsione per le passività con una calibrazione trascurante gli effetti di interazione.*

Trascurare gli effetti di interazione conduce a modelli estremamente semplici ed efficienti, ma con un significativo peggioramento nelle metriche considerate rispetto ai modelli più completi presentati nelle tabelle 2.11, 2.12, 2.13, il che conferma che l’inserimento dei termini di interazione tra i termini candidabili conduce ad un miglioramento nella capacità previsiva del modello.

Compagnia 1								
Categoria	N. termini	T. esecuzione	Set di validazione 1			Set di validazione 2		
			\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Pensioni	33	28s	2.33%	1.08%	1.00%	4.78%	1.50%	1.50%
Protezione	42	26s	1.07%	0.21%	0.20%	4.71%	1.57%	1.58%
Risparmio	63	2m35s	1.65%	0.46%	0.46%	6.43%	1.38%	1.36%
Unit-Linked	13	4s	2.98%	0.70%	0.61%	2.98%	0.38%	0.34%

Compagnia 2								
Categoria	N. termini	T. esecuzione	Set di validazione 1			Set di validazione 2		
			\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Pensioni	26	9s	1.17%	0.34%	0.35%	/	/	/
Protezione	46	4m25s	1.27%	0.29%	0.25%	/	/	/
Unit-Linked	36	25s	3.61%	0.59%	0.59%	/	/	/

Compagnia 3								
Categoria	N. termini	T. esecuzione	Set di validazione 1			Set di validazione 2		
			\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Protezione	82	5m14s	27.12%	1.05%	1.65%	64.27%	15.63%	15.11%
Risparmio	74	4m23s	1.77%	0.63%	0.63%	5.35%	1.10%	1.09%

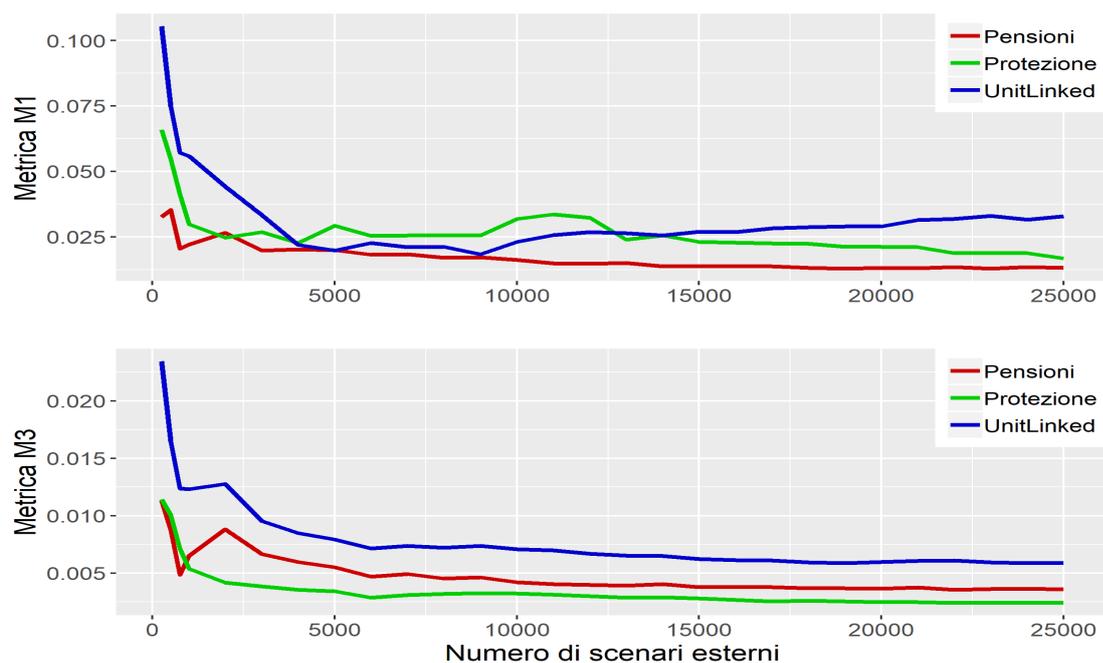
Tabella 2.18: *Statistiche di bontà di previsione per le passività con una calibrazione limitata ai termini di grado minore o uguale a 2.*

Anche la limitazione al grado del polinomio proposta nella Tabella 2.18 peggiora, tendenzialmente, le metriche di validazione. Il modello così costruito risulta quindi troppo semplice per catturare l’andamento della BEL. Osserviamo tuttavia che per le polizze di protezione della terza compagnia, ove la proxy analizzata nella Tabella 2.13 mostrava un pessimo adattamento, qui vi è un sensibile miglioramento: esso è dovuto al fatto che, impedendo all’algoritmo di selezione di includere nel modello

potenze elevate del fattore di mortalità catastrofale - che avevamo identificato come causa del cattivo adattamento, essendoci punti di validazione significativamente al di fuori del range di stima - la proxy risultante si avvicina maggiormente al trend lineare che, come intuibile dall'analisi grafica degli andamenti marginali, sembra caratterizzare questo fattore di rischio.

Nel seguito, ci accingeremo a studiare la bontà della funzione calibrata al variare del numero di scenari esterni. In particolare, ci limiteremo alla Compagnia 2 e alle metriche \mathcal{M}_1 e \mathcal{M}_3 (la metrica \mathcal{M}_2 infatti è sufficientemente vicina alla \mathcal{M}_3 da presentare un andamento simile - nel caso della compagnia 2, in particolare, la differenza tra le due metriche è estremamente ridotta, quindi il suo inserimento sarebbe ridondante). Consideriamo dapprima invariata la struttura polinomiale selezionata con il modello completo, e vediamo, a parità di struttura, come variano le due metriche sotto esame:

Compagnia 2 - Convergenza metriche

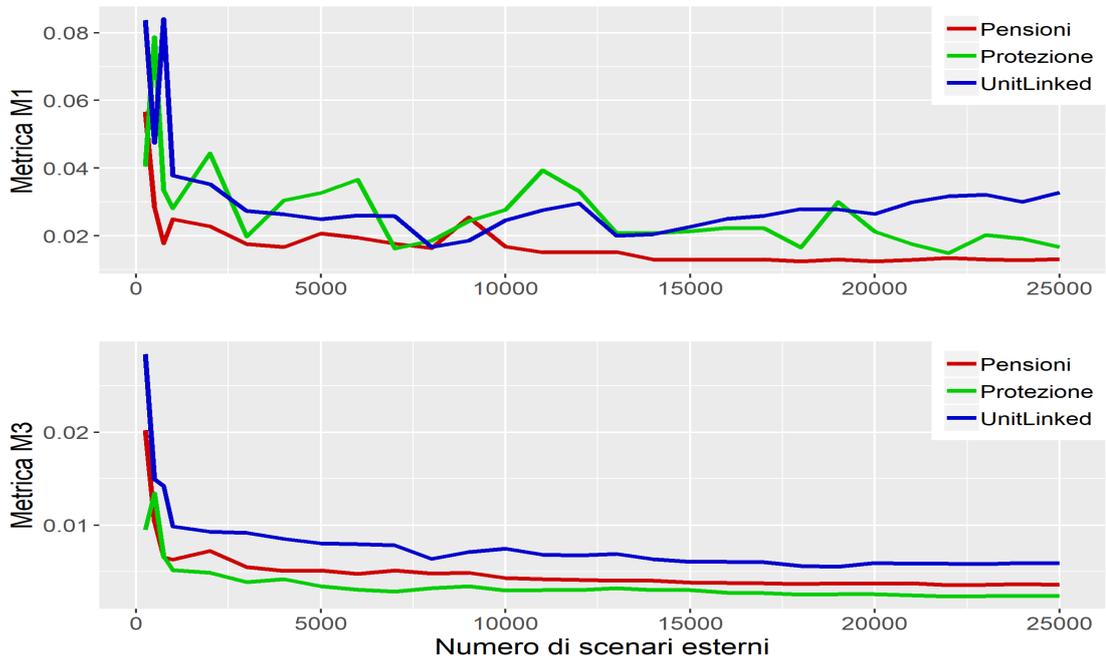


Andamento delle metriche \mathcal{M}_1 e \mathcal{M}_3 al variare del numero di scenari, mantenuta costante la struttura di regressione.

Non ci stupisce la maggiore variabilità del massimo (metrica \mathcal{M}_1) rispetto alla me-

dia (metrica \mathcal{M}_3). Riteniamo in particolare che in questa analisi sia quest'ultima la metrica più importante da osservare: \mathcal{M}_1 infatti, per quanto fondamentale per capire quale sia, al massimo, l'errore di approssimazione commesso dal modello proxy su un singolo scenario, risulta troppo variabile per essere considerata l'indicatore primario di bontà del modello. Per quanto riguarda la metrica \mathcal{M}_3 , invece, mantenuta fissa la struttura di regressione, già 10'000-15'000 scenari esterni sembrano sufficienti a raggiungere valori stabili. Simili risultati erano stati ottenuti anche da Vounzi (2017). Tuttavia, nell'analisi finora effettuata abbiamo supposto che la struttura di regressione sia fissata pari a quella calcolata con 25'000 scenari esterni. Un'ipotesi di questo tipo risulta fortemente riduttiva, in quanto un numero diverso di scenari porterà l'algoritmo automatico di selezione, che già abbiamo commentato essere potenzialmente soggetto a variabilità relativamente alle configurazioni utilizzate, a includere nel modello termini diversi. Ripetiamo l'esercizio sopra ricalcolando l'intera struttura polinomiale (anziché solo i coefficienti) tramite il numero ridotto di scenari.

Compagnia 2 - Convergenza metriche



Andamento delle metriche \mathcal{M}_1 e \mathcal{M}_3 al variare del numero di scenari, ricalcolando anche la struttura di regressione.

Variando anche la struttura di regressione, la metrica \mathcal{M}_1 risulta significativamente più instabile: essa infatti dipende fortemente dai termini polinomiali scelti. La metrica \mathcal{M}_3 risulta invece molto stabile per tutte le LoB considerate, nonostante eventuali variazioni nella struttura polinomiale: già con 10'000-15'000 scenari si ottengono ottimi risultati. L'algoritmo di costruzione della funzione proxy sembrerebbe garantire risultati molto stabili al variare del numero di scenari. L'utilizzo di 25'000 effettuato da noi e dalla maggior parte delle applicazioni in letteratura dovrebbe quindi essere sufficientemente elevato da garantire la stabilità e l'affidabilità delle previsioni effettuate dal modello. Ricordiamo in ogni caso che tale risultato è sicuramente in parte dovuto anche all'utilizzo delle sequenze di Sobol per la determinazione degli scenari di fitting; come avevamo osservato nella sezione 1.2.3, la copertura dello spazio ottenuta tramite la sequenza a bassa discrepanza consente di stabilizzare i risultati utilizzando un numero di simulazioni molto più basso di quello che sarebbe necessario qualora ci si basasse sui numeri pseudo-casuali.

Nella tabella seguente riportiamo, a titolo d'esempio, i risultati relativi alla calibrazione effettuata considerando solo 12'500 scenari esterni (sempre generati con sequenze di Sobol), dimezzando quindi il budget computazionale utilizzato per il fitting:

Compagnia 2 - 12'500 scenari esterni					
Categoria	N. termini	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Pensioni	27	11s	1.31%	0.36%	0.36%
Protezione	68	2m36s	2.45%	0.35%	0.31%
Unit-Linked	39	36s	2.59%	0.66%	0.66%

Tabella 2.19: *Statistiche di bontà di previsione per i passivi della Compagnia 2 considerando solo 12'500 scenari.*

Come ci aspettavamo, le statistiche nella Tabella 2.19 sono leggermente peggiorate rispetto a quelle della Tabella 2.12, ma la differenza è relativamente ridotta; anche con un dimezzamento del budget computazionale, il modello quindi sembra funzionare molto bene.

Passiamo ora ad analizzare la bontà del modello al variare del numero di scenari

interni. Come abbiamo riportato nella sezione 1.3.3, nelle analisi effettuate da Cathcart (2012) la configurazione ottimale sembra prevedere due scenari interni antitetici: tale via è quella solitamente seguita nelle applicazioni pratiche del modello. Tale approccio è diverso da quello utilizzato da Bauer et al. (2009), Ha & Bauer (2018) e molti altri in letteratura, ove viene invece considerato un singolo scenario interno: si dimostra infatti che non ricorrendo alle variabili antitetiche un singolo scenario risulta ottimale.

Proveremo a confrontare queste due impostazioni per la Compagnia 2, fissato il budget computazionale complessivo a 25'000 scenari. Utilizzando due scenari interni antitetici, come abbiamo fatto finora, otteniamo dunque 12'500 scenari esterni per la calibrazione della funzione, e i relativi risultati sono già riportati nella Tabella 2.19. Nella Tabella 2.20, invece, riportiamo i risultati relativi a 25'000 scenari esterni con un singolo scenario interno. Confrontando le due tabelle, osserviamo come - a parità di scenari - l'utilizzo delle variabili antitetiche provochi un significativo miglioramento nella bontà della funzione proxy. I risultati ottenuti sono in linea con quelli di Cathcart (2012) che abbiamo presentato nella sezione 1.3.3 e confermano la bontà dell'approccio basato sugli scenari antetici che abbiamo adottato finora.

Compagnia 2 - 1 scenario interno					
Categoria	N. termini	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Pensioni	23	10s	2.36%	0.81%	0.81%
Protezione	42	1m10s	1.24%	0.42%	0.42%
Unit-Linked	37	50s	4.42%	1.11%	1.13%

Tabella 2.20: *Statistiche di bontà di previsione per i passivi della Compagnia 2 considerando un solo scenario interno per ciascuno dei 25'000 scenari esterni.*

In ultimo, come per le attività effettuiamo un'implementazione del metodo MARS tramite il pacchetto R *mda*. Nella Tabella 2.21 sono riportati i risultati dell'esperimento.

Compagnia 1							
		Set di validazione 1			Set di validazione 2		
Categoria	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Pensioni	5s	2.16%	0.73%	0.67%	12.90%	4.32%	4.31%
Protezione	2s	2.83%	0.52%	0.50%	72.31%	2.47%	2.00%
Risparmio	8s	1.71%	0.40%	0.39%	9.76%	4.56%	4.58%
Unit-Linked	1s	3.52%	0.69%	0.60%	1.99%	0.67%	0.68%

Compagnia 2							
		Set di validazione 1			Set di validazione 2		
Categoria	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Pensioni	2s	1.81%	0.50%	0.51%	/	/	/
Protezione	2s	3.98%	1.56%	1.57%	/	/	/
Unit-Linked	7s	2.46%	0.79%	0.79%	/	/	/

Compagnia 3							
		Set di validazione 1			Set di validazione 2		
Categoria	T. esecuzione	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Protezione	13s	44.12%	1.89%	2.81%	1635.58%	41.45%	44.69%
Risparmio	10s	1.19%	0.28%	0.27%	17.34%	4.23%	4.31%

Tabella 2.21: *Statistiche di bontà di previsione per le passività con una calibrazione basata sul modello MARS.*

Come abbiamo constatato relativamente ai portafogli di attivi, il modello così costruito ottiene delle statistiche in linea con le configurazioni altrettanto rapide del modello polinomiale riportate nelle tabelle 2.17 e 2.18, ma risulta significativamente peggiore del modello polinomiale completo sviluppato nelle tabelle 2.11, 2.12, 2.13. Osserviamo in particolare un peggioramento significativo nel caso di punti al di fuori del range di fitting, anche nei casi in cui il modello polinomiale non evidenziava problemi (LoB Protezione della Compagnia 1). Riteniamo quindi che il modello MARS non rappresenti, nella sua forma classica, una soluzione adeguata al problema. Ribadiamo tuttavia le potenzialità di un modello MARS con funzioni base non localmente lineari.

2.3.3 Applicazione ai PVFP e calcolo dell'SCR

Nella sezione precedente abbiamo applicato il metodo LSMC a portafogli di attività e passività assicurative. Come abbiamo già detto, tale metodo necessita di un successivo processo di aggregazione per ottenere una distribuzione complessiva per i fondi propri della compagnia e quindi per ottenere l'SCR. Lo sviluppo di un modello di aggregazione del genere va oltre gli scopi di questa tesi.

Per giungere ad una misura dell'SCR seguiremo invece una strada diversa, applicando il modello LSMC direttamente ai PVFP della compagnia, come abbiamo delineato nell'approccio seguito nelle sezioni 1.1.2, 1.2.2, 2.1.2. Considereremo successivamente in maniera semplificata i PVFP come proxy del capitale disponibile dell'impresa. Così facendo, possiamo calcolare la perdita in ogni scenario real-world direttamente sulla base del modello LSMC per i PVFP, senza necessitare di un processo di aggregazione.

Tale approccio è anche quello seguito nella letteratura relativa al modello LSMC per il calcolo dei requisiti di capitale, ove esso per semplicità viene solitamente applicato ai PVFP, anziché separatamente a portafogli di attività e passività. Si veda, ad esempio, Bauer et al. (2009) o DAV (2015).

L'approccio che seguiremo in questa sezione ci permette quindi di ottenere in maniera semplice un valore per l'SCR della compagnia e quindi di studiare anche un possibile impatto sui requisiti di capitale delle diverse configurazioni del modello. Per quanto corretto da un punto di vista teorico e utile per le nostre analisi, tale approccio non è consigliato per un'implementazione pratica del modello: modellando direttamente i fondi propri della compagnia con un'unica funzione proxy, infatti, l'effetto dei fattori di rischio risulta molto meno intuitivo e dipenderà fortemente dalla composizione dei propri portafogli. A titolo di esempio, si pensi ad una compagnia con un portafoglio consistente di polizze con beneficio principale in caso di morte dell'assicurato e un portafoglio altrettanto consistente di polizze con beneficio principale in caso di vita dell'assicurato. Il fattore di rischio legato all'aumento nella mortalità avrà ovviamente un effetto negativo sul primo portafoglio, positivo sul secondo. Considerando un'unica funzione per i PVFP dell'impresa, l'effetto complessivo del fattore non sarà più evidente e dipenderà dalla composizione del portafoglio. In tal modo, quindi, non si potranno più effettuare valutazioni relati-

ve all'esposizione dei singoli portafogli ai diversi rischi e la funzione generata sarà molto meno robusta rispetto a variazioni nei portafogli.

Procediamo in ogni caso con la costruzione del modello LSMC per i PVFP dell'impresa, utilizzando lo stesso approccio che abbiamo delineato per la valutazione della BEL dei portafogli di passività in ambito vita, quindi considerando sempre 25'000 scenari esterni e 2 scenari interni per ogni scenario esterno. Questa volta però non consideriamo più separatamente le diverse LoB, modellando direttamente i PVFP complessivi dell'impresa. Ai fini di questa analisi considereremo soltanto la compagnia 3 dell'esempio sviluppato in precedenza.

Di seguito le statistiche relative alla qualità del modello risultante (la configurazione dell'algoritmo è sempre data dal forward AIC con grado massimo 4 e vincolo computazionale di 150 termini). L'errore si riferisce soltanto al set relativo alla validazione globale della funzione proxy, in quanto per questa applicazione non siamo in possesso delle informazioni relative agli altri scenari di validazione.

Modello per il Present Value of Future Profits			
	N. termini	T. esecuzione	\mathcal{M}_3
Compagnia 3	150	59m23s	4.31%

Tabella 2.22: *Statistiche di bontà di previsione per i PVFP della Compagnia 3.*

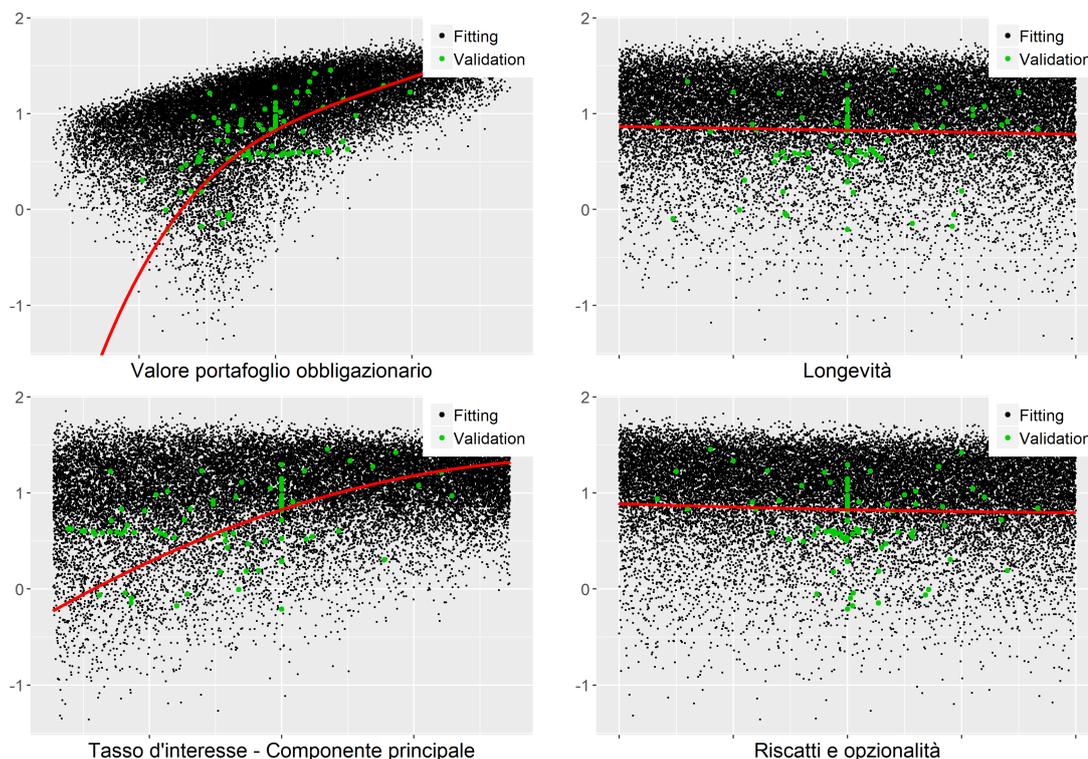
Osserviamo subito che l'errore di previsione risulta, in questo contesto, significativamente più elevato di quello negli esempi precedenti. Ciò è dovuto a due cause:

- La granularità di applicazione del modello LSMC è, in questo esempio, significativamente ridotta, dato che non stiamo più considerando la suddivisione in LoB. La funzione proxy quindi dovrà cogliere andamenti meno interpretabili, in quanto le componenti legate alle diverse tipologie di contratti nel portafoglio di polizze della compagnia si compensano o interagiscono tra di loro.
- Il valore dei PVFP è, in termini relativi, più instabile rispetto a quello delle BEL, e la sua variazione rispetto ai fattori di rischio è più pronunciata.

Inoltre, gli importi relativi ai PVFP sono di entità significativamente minore rispetto alle BEL, per cui errori anche ridotti su importi quasi nulli portano ad un forte incremento dell'errore relativo.

Per visualizzare al meglio l'andamento marginale della funzione così calibrata, consideriamo come negli esempi precedenti un'analisi grafica della curva marginale rispetto ai dati considerati. Ricordiamo sempre che la curva non è direttamente confrontabile con i punti a causa del fatto che lo stress nullo applicato agli altri fattori di rischio non coincide sempre con la loro media e che la curva marginale cattura l'effetto puro di ogni fattore di rischio, mentre gli scenari utilizzati possono includere effetti di interazione.

Compagnia 3 - PVFP



Andamento marginale della proxy costruita per il valore dei PVFP della compagnia, rispetto ad alcuni dei fattori di rischio considerati; punti neri rappresentano gli scenari utilizzati per calibrare la funzione, punti verdi quelli su cui è calcolato l'errore out-of-sample.

Osserviamo subito la maggiore variabilità dei dati relativi ai PVFP complessivi della compagnia rispetto ai modelli più granulari e più stabili che avevamo considerato nella sezione precedente. Anche qui, il ruolo preponderante è dato dai rischi finanziari, mentre altre componenti, quali i riscatti e o i rischi biometrici, hanno un effetto significativamente più ridotto, con una curva marginale molto più stabile.

Non considereremo, per questo esempio, le diverse configurazioni alternative che abbiamo sviluppato nella sezione precedente, in quanto i risultati sono pressoché analoghi: l'imposizione di vincoli più stringenti riduce significativamente la bontà del modello, l'utilizzo del BIC non provoca chiari miglioramenti o peggioramenti, il metodo MARS risulta un'alternativa potenzialmente interessante per la sua rapidità d'esecuzione, ma porta a risultati comunque peggiori del modello polinomiale.

Ricordato che nella sezione 1.1.2 avevamo definito $L = AC_0 - \frac{AC_1}{1+r(0,1)}$ assumiamo in via semplificata che:

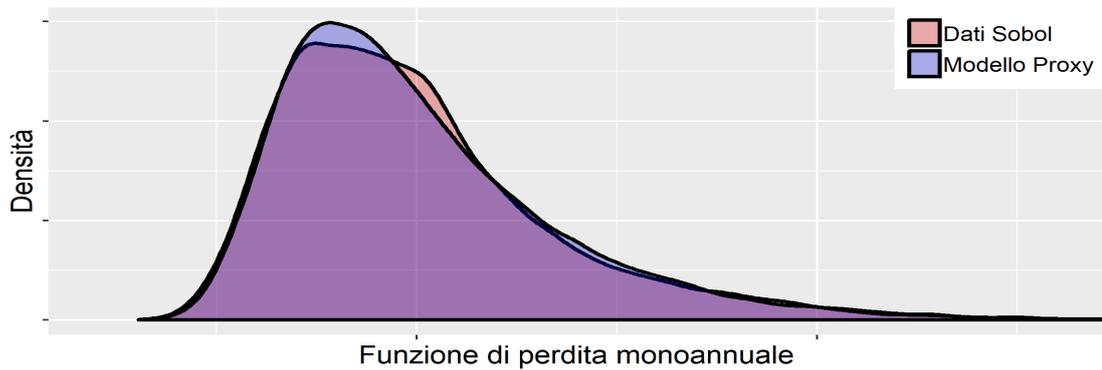
- La variazione dei fondi propri ammissibili coincida con la variazione dei PVFP, ossia trascuriamo eventuali cambiamenti nelle altre componenti dei fondi propri;
- $\frac{AC_1}{1+r(0,1)}$ corrisponda proprio al $PVFP_1$ calibrato dalla funzione proxy, il che è lecito in quanto, nella nostra applicazione, gli scenari tengono già conto in maniera implicita dell'attualizzazione;
- AC_0 sia pari al valore atteso di $PVFP_1$ sotto la probabilità real-world e stimato tramite la media empirica sulle simulazioni effettuate.

In tal modo, dati N scenari real-world, possiamo utilizzare la funzione proxy calibrata sui PVFP della compagnia per ottenere direttamente il valore della perdita monoannuale L in ogni scenario, e quindi ricavare una distribuzione simulata della perdita della compagnia in base alla quale calcolare il VaR. Naturalmente, l'ipotesi relativa ad AC_0 risulta piuttosto lontana dalla realtà ed è effettuata unicamente per avere un parametro di confronto nel nostro modello semplificato; nella pratica, tale quantità sarà nota all'impresa, in quanto stimata in maniera consistente

tramite il modello stocastico di proiezione dei flussi. In ogni caso la forma della distribuzione considerata per la perdita - a meno della sua posizione - non dipenderà da AC_0 , che, per quanto detto, supponiamo non aleatorio, per cui la componente stocastica per il nostro modello è data unicamente da AC_1 .

Nella figura seguente confrontiamo le distribuzioni empiriche simulate della perdita, calcolata sui dati originali e su quelli previsti dalla funzione proxy.

Compagnia 3 - Distribuzione della perdita sullo spazio di fitting

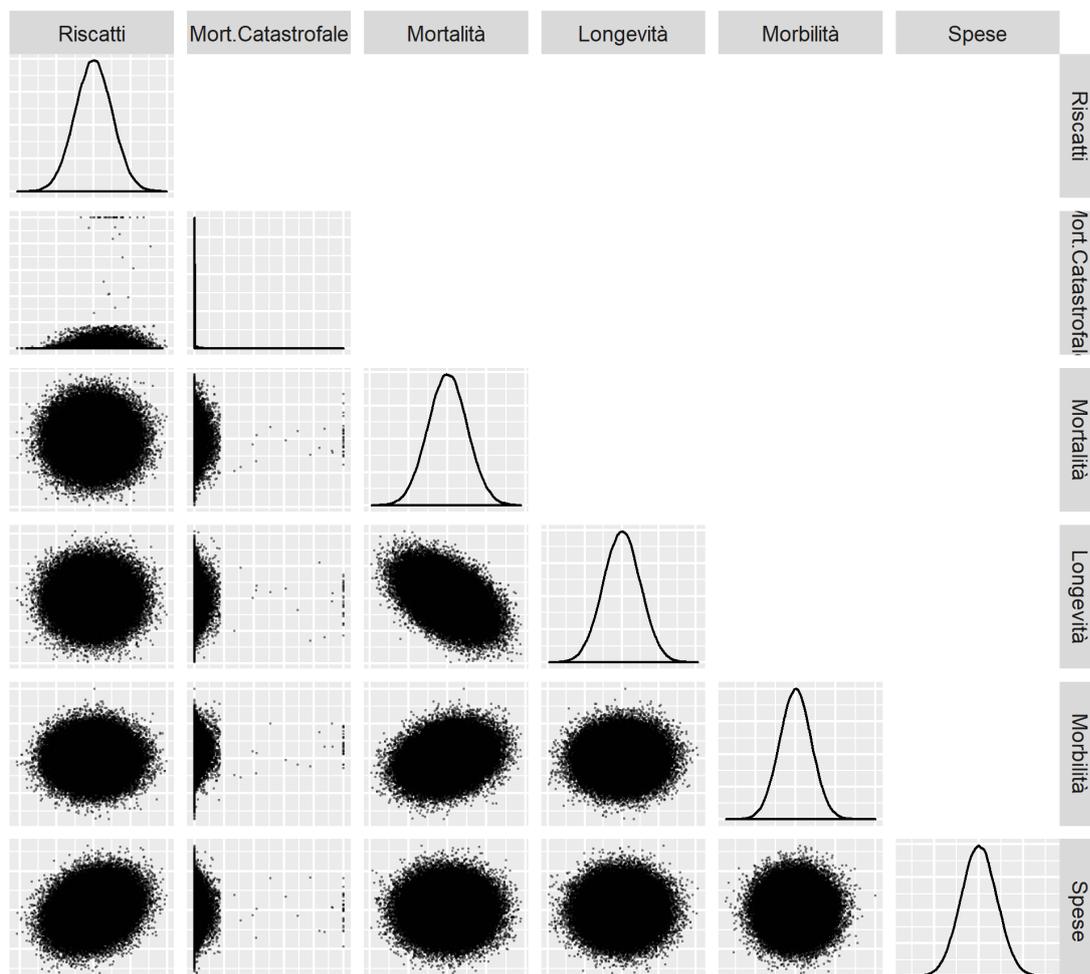


Distribuzione empirica della perdita sullo spazio di fitting; valori veri e valori previsti dalla proxy.

Pur non essendo perfetto, anche a causa della scarsa granularità a cui abbiamo applicato questo modello, la forma della distribuzione risulta preservata passando alla proxy, in particolare sulla coda. Ricordiamo in ogni caso che, essendo i dati di fitting calcolati come medie su due soli valori, non riteniamo la distribuzione da essi ottenuta per forza affidabile, e che l'analisi grafica sopra, per quanto interessante da un punto di vista interpretativo, non ci fornisce davvero informazioni sulla capacità della funzione proxy di approssimare perdite calcolate in maniera precisa con l'approccio stocastico.

Passiamo ora alla distribuzione real-world dei fattori di rischio. A tal fine, si suppongono calibrate le distribuzioni marginali per ogni fattore di rischio e le correlazioni tra i fattori, in modo da poter ottenere la distribuzione congiunta da cui effettuare le simulazioni. A titolo d'esempio, nella figura seguente sono raffigurati gli andamenti congiunti delle coppie dei fattori di rischio tipici del settore

vita. Analogamente, si possono raffigurare le coppie relative ai rischi finanziari e, qualora presenti, le interazioni tra singoli fattori finanziari e non finanziari.



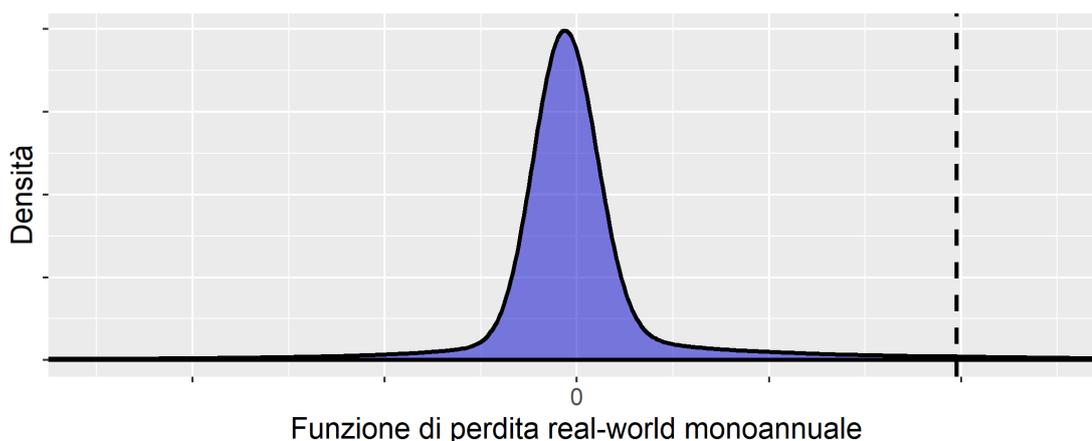
Andamento congiunto delle coppie di fattori considerati relativamente ai rischi tipici del settore vita.

Abbiamo quindi supposto una correlazione positiva tra riscatti e spese e tra mortalità e morbilità; una correlazione negativa relativamente forte tra mortalità e longevità; correlazioni di entità meno evidente tra le altre coppie. Per quanto riguarda la correlazione tra mortalità e longevità, si osservi che se a livello di singola polizza si può assumere massima diversificazione tra i due rischi, ciò non è più valido globalmente sul portafoglio, infatti gli assicurati con beneficio caso vita

e quelli con beneficio caso morte possono avere caratteristiche significativamente diverse: in particolare, essi possono appartenere a fasce d'età diverse, per cui le variazioni nella mortalità/longevità possono impattare sul portafoglio in modi per cui le perdite su alcune tipologie di contratti non sono compensate da profitti su altre. Ne segue che, per quanto sia ragionevole assumere una correlazione negativa, essa non dovrà essere troppo marcata (nella standard formula, ad esempio, si considera -25%). Di particolare rilievo, inoltre, la forma fortemente asimmetrica della distribuzione della mortalità catastrofica, in linea con il significato attribuito al fattore di rischio.

Alla luce di quanto detto, considereremo ora 100'000 simulazioni dalla distribuzione congiunta real-world dei fattori di rischio inseriti nel modello, ad ognuna delle quali applicheremo la funzione proxy in modo da stimare il valore della perdita nello scenario. Nella figura seguente riportiamo la distribuzione real-world della perdita così ricavata e il suo VaR al 99.5%.

Compagnia 3 - Distribuzione della perdita real-world



Distribuzione empirica real-world della perdita; la linea tratteggiata indica il VaR al 99.5%.

Abbiamo quindi ottenuto la funzione di perdita per la compagnia e un valore per l'SCR.

2.4 Conclusioni

Nel corso di questo capitolo abbiamo concentrato la nostra attenzione sul modello Least-Squares Monte Carlo e sulla sua applicazione nel contesto della stima dei requisiti di capitale nel contesto di un modello interno coerente con i dettami di Solvency II, con particolare riferimento ad un'impresa di assicurazioni vita.

Più nel dettaglio, abbiamo proposto uno schema pratico di implementazione del modello e delineato le possibili scelte riguardanti la selezione dei termini polinomiali, nonché fornito il quadro teorico di riferimento in cui si colloca il metodo, esaminandone il comportamento asintotico seguendo Ha & Bauer (2018).

Successivamente, abbiamo tracciato le basi per un processo di validazione del modello e fornito brevi cenni alle diverse proposte nella recente letteratura mirate a migliorare la struttura di regressione su cui fonda il modello.

Infine, abbiamo lavorato ad un'implementazione del modello tramite il linguaggio R (nell'Appendice B sono esposte le principali funzioni sviluppate) e abbiamo provato diverse configurazioni su veri portafogli assicurativi, mostrando come in un contesto pratico l'applicazione delle tecniche LSMC possa portare, generalmente, a una buona approssimazione dei risultati ottenibili con un modello stocastico annidato, mantenendo però basso lo sforzo computazionale.

Risulta impossibile stabilire con precisione quale sia l'impatto dell'utilizzo del modello LSMC rispetto ad un modello stocastico annidato puro. Tuttavia, analizzando singoli scenari - anche estremi - si è osservato che l'errore di previsione del modello polinomiale risulta, a meno di casi particolari, sempre minore del 5%, con un errore assoluto medio solitamente minore dell'1%, a fronte di un risparmio computazionale immenso. L'errore medio di previsione risulta ancora minore a causa delle compensazioni tra leggere sovra- e sottostime. Si ritiene quindi che la tecnica LSMC sia, in generale, molto valida per la costruzione di un modello interno nel contesto del risk management assicurativo.

Ritenendo che il punto più delicato del modello consista nel meccanismo di costruzione della funzione proxy polinomiale, abbiamo poi testato diverse configurazioni dell'algoritmo di selezione, mostrandone l'impatto sulle metriche usualmente utilizzate per la validazione del modello. Abbiamo successivamente verificato l'an-

damento delle marginali del modello polinomiale, verificandone la consistenza con quanto ci si aspettava intuitivamente, e testato un'applicazione del metodo MARS in alternativa alla regressione OLS: esso rappresenta una soluzione più veloce e parsimoniosa rispetto al modello polinomiale. Tuttavia, nella sua configurazione standard, il modello MARS solitamente non fornisce risultati altrettanto buoni.

Infine, abbiamo proposto un'applicazione del modello direttamente ai PVFP complessivi di una compagnia assicurativa, in modo da mostrare in via semplificata come, una volta calibrata la funzione proxy, si possa procedere con la determinazione della distribuzione della perdita e il successivo calcolo dell'SCR. Il modello semplificato per i PVFP risulta più instabile ed è da considerarsi a puri fini illustrativi, non molto adeguato, quantomeno al livello estremamente poco granulare da noi considerato, per un'effettiva implementazione pratica in un modello interno per una compagnia con una molteplicità di prodotti diversi nel proprio portafoglio di polizze.

Nonostante il giudizio complessivamente molto positivo, il modello LSMC non è esente da alcune limitazioni, che riassumiamo brevemente:

- La validità del modello per scenari al di fuori dello spazio considerato per la calibrazione può essere potenzialmente compromessa, anche in maniera significativa, come abbiamo constatato nella Tabella 2.13.
- L'algoritmo di selezione può richiedere limitazioni arbitrarie per convergere in tempi ragionevoli (si veda la Tabella 2.6); inoltre, la scelta del criterio d'arresto dell'algoritmo non è univoca e non sempre il modello è robusto rispetto alla decisione effettuata. Vi possono poi essere casi in cui l'algoritmo di selezione non ottiene un buon adattamento ed è richiesto un intervento manuale.
- Diverse configurazioni dell'algoritmo di selezione possono portare a risultati diversi, e non esiste un criterio valido globalmente per ricercare formalmente la configurazione ottimale.

Allo stesso tempo, tuttavia, tali limitazioni sono in larga parte comuni anche agli altri modelli proxy: nel curve fitting, il valore ottenuto dipende in maniera ancora

più forte dalle scelte effettuate nel processo di fitting, garantendo meno robustezza, e anche molte proposte di calibrazione automatica dei portafogli replicanti dipendono da un processo di regressione, portando il modello a soffrire di limitazioni analoghe. Inoltre, rispetto alle attuali alternative in letteratura, l'algoritmo LSMC è senza dubbio quello che richiede meno interventi manuali per ottenere un buona capacità di adattamento.

Ricordiamo che la quantità ridotta di scenari di validazione a disposizione per il calcolo delle metriche ci impone cautela nel prendere come assoluti i risultati del confronto tra modelli. L'implementazione del modello sarà sempre legata ad una certa quantità di incertezza sull'effettiva validità dei risultati, dovuta principalmente all'utilizzo di dati imprecisi per la calibrazione della funzione proxy. Considerati però i forti vincoli computazionali alla base di un modello di questo tipo, riteniamo che la generale capacità del modello nel descrivere il fenomeno in esame sia estremamente buona, e che non ci si possa aspettare un errore significativamente minore da un modello soggetto ai vincoli computazionali fissati.

Un'ultima osservazione: il metodo LSMC, come gli altri modelli proxy, ha l'obiettivo di approssimare un altro modello, ossia la proiezione dei cash-flow tramite simulazioni stocastiche annidate. Qualora le distribuzioni dei fattori di rischio o i modelli di proiezione siano inadeguati, i risultati del modello LSMC saranno ovviamente anch'essi compromessi. Il modello proxy, in sostanza, al più è buono quanto il modello alla sua base.

Capitolo 3

Un modello proxy basato su metodi di machine learning

Gli algoritmi di machine learning mirano a costruire in maniera automatica modelli matematici per cogliere gli schemi sottostanti ad un insieme di dati, migliorando progressivamente la capacità di previsione o di classificazione. Gli approcci alla base della costruzione di modelli di machine learning sono molteplici e risulta difficile darne una descrizione comune. In quanto segue, ci limiteremo a considerare tecniche di apprendimento supervisionato, nelle quali la calibrazione del modello avviene sulla base di una sequenza di coppie di input e output inizialmente fornite - nella nostra specifica applicazione, tale sequenza sarà data dai valori dei fattori di rischio nelle simulazioni esterne utilizzate per il fitting, e dai corrispondenti valori degli elementi del bilancio calcolati con il modello di proiezione dei flussi di cassa.

Restriangeremo ulteriormente l'analisi alle tecniche di apprendimento costruite a partire da strutture ad albero. Nella prima parte del capitolo, forniremo un breve inquadramento teorico relativo agli alberi di regressione e ai metodi più complessi che su di essi si basano: in particolare, introdurremo le foreste casuali (random forests) e i più avanzati algoritmi di gradient boosting. Nella seconda parte del capitolo utilizzeremo le tecniche esposte per costruire una proxy per il modello stocastico annidato e confronteremo la bontà di previsione con quella della funzione polinomiale del modello LSMC tradizionale analizzato nel capitolo precedente.

3.1 L'apprendimento con modelli ad albero

Nel modello di regressione lineare, abbiamo modellato gli effetti non-lineari e gli effetti congiunti dei regressori inserendo nel modello potenze e prodotti tra le variabili esplicative, osservando che all'aumentare della complessità del problema in esame, la dimensione del modello cresce significativamente e le funzioni di regressione costruite perdono di interpretabilità. Inoltre, gli algoritmi di selezione possono richiedere tempi molto lunghi per calibrare la funzione, come abbiamo visto nella sezione 2.3. Tali difficoltà sono dovute al fatto che il modello di regressione è un modello globale, in cui un'unica formula predittiva è valida sull'intero spazio campionario.

La logica alla base dei modelli ad albero invece è quella di suddividere lo spazio in regioni più piccole, in modo da poter gestire con maggiore facilità le non-linearità e gli effetti di interazione. Ciò è effettuato tramite partizionamento ricorsivo (*recursive partitioning*) fino a che le regioni individuate sono sufficientemente piccole da potervi applicare un modello semplice. La struttura ad albero non è che la rappresentazione della procedura di partizionamento ricorsivo.

Dopo aver sviluppato il modello degli alberi di regressione, presenteremo alcune tecniche di apprendimento ensemble: l'idea alla loro base è quella di combinare le previsioni di un insieme di alberi di regressione, al fine di costruire un modello unico con una miglior capacità predittiva rispetto a quelli da cui è costituito.

3.1.1 Alberi di regressione

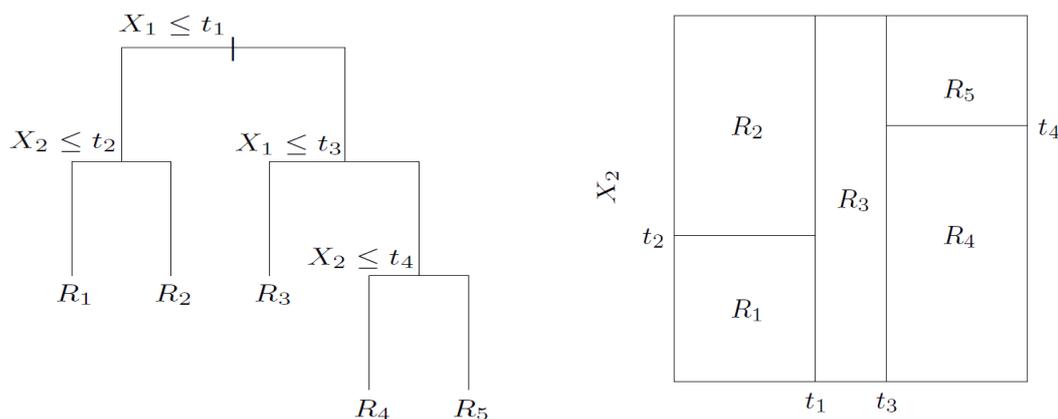
Gli alberi di regressione nascono dall'utilizzo di una struttura decisionale ad albero per la costruzione di un modello idoneo alla previsione. Come preannunciato, l'obiettivo è effettuare un partizionamento ricorsivo dello spazio campionario, applicando poi un modello semplice per descrivere i dati in ciascuna regione così individuata.

Ogni punto di ripartizione dell'albero è detto nodo, il nodo iniziale - in cui lo spazio non ha ancora subito nessuna suddivisione - radice dell'albero. La struttura dell'albero è gerarchica: la radice dell'albero è il livello più alto della gerarchia, dopodiché le relazioni gerarchiche tra nodi si esprimono chiamandoli nodi padri

(o ascendenti) e nodi figli (o discendenti). I nodi senza ulteriori discendenti sono detti nodi terminali o foglie e rappresentano il livello più basso della struttura gerarchica.

Nel seguito, descriveremo la struttura di tipo CART (Classification and Regression Trees), considerata al giorno d'oggi la metodologia standard per la costruzione di un albero di classificazione o di regressione. Ci limiteremo al caso in cui la variabile dipendente Y sia una variabile aleatoria quantitativa continua: infatti, nelle nostre applicazioni essa rappresenterà sempre una perdita o il valore di un portafoglio. In altre parole, ci limiteremo a considerare gli alberi di regressione, non trattando quindi l'utilizzo delle strutture CART per la risoluzione di problemi di classificazione.

Nel modello CART classico, ogni partizione è binaria, ossia provoca una bipartizione dello spazio a seconda del fatto che una variabile sia maggiore o minore di una fissata soglia (partizioni in più di due gruppi rispetto ad una data variabile possono comunque essere ottenute tramite partizioni binarie ripetute). Inoltre, il modello utilizzato in ciascun nodo terminale dell'albero è semplicemente dato da una stima costante della variabile dipendente. L'albero quindi crea una funzione costante a tratti, il che consente di effettuare previsioni in maniera rapida e intuitiva.



Struttura di un albero di regressione (a sinistra) e partizione dello spazio campionario da esso creata (a destra). Immagine tratta da Hastie et al. (2009).

Qualora necessario, osserviamo che in ogni nodo terminale potremmo considerare anche modelli più complessi di quello che associa ad ogni campione del nodo un valore costante. Con una scelta di questo tipo, si passa da alberi di regressione a c.d. “alberi modello”. Un caso particolare di questo approccio è dato proprio dal metodo MARS, che abbiamo brevemente esposto nella sezione 2.2.3 nell’ambito dei modelli additivi. La struttura ottenuta con il metodo MARS può essere, in un certo senso, interpretata come una particolare struttura ad albero in cui in ogni nodo si ha un modello lineare anziché una costante (corrispondente al modello nullo), e sulla quale si è imposto come ulteriore vincolo la continuità nei nodi.

Tornando al caso base, seguiamo Hastie et al. (2009) e consideriamo dapprima una partizione dello spazio campionario in M regioni R_1, \dots, R_M . Il modello sarà:

$$f(\mathbf{x}) = \sum_{m=1}^M c_m \mathbb{1}_{R_m}(\mathbf{x}) \quad (3.1)$$

con c_m costante. Se utilizziamo come criterio di ottimizzazione quello dei minimi quadrati, dato il campione $(y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n)$ con $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,p})$, la stima di c_m sarà data semplicemente dalla media aritmetica dei valori y_i per i quali $\mathbf{x}_i \in R_m$, ossia avremo $\hat{c}_m = \bar{y}^{(m)}$.

Rimane da studiare il metodo con cui effettuare la partizione dello spazio. Risolvere globalmente il problema dei minimi quadrati sull’insieme delle possibili partizioni non è computazionalmente attuabile, per cui l’approccio adottato è quello che prevede la ricerca di un ottimo locale ad ogni split binario. A partire dal nodo radice dell’albero, in cui tutti i dati sono considerati insieme, per ogni variabile esplicativa $j \in 1, \dots, p$ e per ogni possibile punto di suddivisione s consideriamo le due regioni $R_1(j, s) = \{\mathbf{x} : x_j \leq s\}$, $R_2(j, s) = \{\mathbf{x} : x_j > s\}$.¹

Per determinare j e s , procediamo utilizzando nuovamente il criterio dei minimi quadrati e risolviamo il problema di ottimo:

¹Osserviamo che potremmo anche pensare a regole di suddivisione più complesse per effettuare la bipartizione, ad esempio considerando una condizione del tipo $g(x_1, \dots, x_p) \leq s$. La logica del procedimento analizzato è valida anche in questo caso più generale, anche se il problema di ottimo diventerà, potenzialmente, più complesso da risolvere.

$$\min_{j,s} \left(\min_{c_1} \sum_{\mathbf{x}_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{\mathbf{x}_i \in R_2(j,s)} (y_i - c_2)^2 \right) \quad (3.2)$$

Per l'osservazione sopra, fissati j e s , le soluzioni ai due problemi di minimo interni sono dati, rispettivamente, dalle medie campionarie degli y_i sulle regioni individuate da $R_1(j, s)$ e $R_2(j, s)$, ossia $\bar{y}^{(1)}(j, s)$ e $\bar{y}^{(2)}(j, s)$.

Quindi il problema di minimo diventa:

$$\min_{j,s} \left(\sum_{\mathbf{x}_i \in R_1(j,s)} (y_i - \bar{y}^{(1)}(j, s))^2 + \sum_{\mathbf{x}_i \in R_2(j,s)} (y_i - \bar{y}^{(2)}(j, s))^2 \right) \quad (3.3)$$

Una volta risolto tale problema individuando il valori ottimali di j e s , che indichiamo con j^* , s^* , avremo una bipartizione dello spazio campionario in $R_1(j^*, s^*)$ e $R_2(j^*, s^*)$. A questo punto, lo stesso procedimento viene applicato separatamente alle due regioni di spazio individuate, che saranno nuovamente bipartite, suddividendo complessivamente lo spazio campionario in quattro regioni. Il procedimento continua iterativamente fino al raggiungimento di una condizione d'arresto.

Per individuare la condizione d'arresto, il metodo CART prevede dapprima la costruzione di un modello che individua una partizione molto fine dello spazio, continuando con le bipartizioni dello spazio fino a che ogni regione non detiene che un numero esiguo di dati oppure, eventualmente, fino a che non si è raggiunto un fissato numero massimo di iterazioni. Successivamente, al fine di ridurre il pericolo di overfitting, viene effettuata una procedura di potatura (pruning) dell'albero. La potatura di un qualsiasi nodo interno dell'albero comporta la sua eliminazione, oltre che quella di tutti i suoi nodi figli. La procedura effettuata eliminerà i nodi meno significativi in base ad una fissata funzione di costo.

Delineamo ora una delle possibili funzioni di costo utilizzate nella potatura: la funzione di costo-complessità. Dato l'albero T_0 , indichiamo con $T \subset T_0$ ogni sotto-albero ottenibile potando l'albero T_0 . Indichiamo inoltre con $|T|$ il numero di nodi terminali di T .

Possiamo allora definire per il generico albero T l'errore quadratico medio relativo alla regione R_m individuata dall' m -esimo nodo terminale dell'albero:

$$Q_m(T) = \frac{1}{N_m} \sum_{\mathbf{x}_i \in R_m} (y_i - \bar{y}^{(m)})^2 \quad (3.4)$$

ove, come prima, N_m rappresenta il numero di elementi del campione in R_m .

Possiamo infine definire il costo-complessità dell'albero T come:

$$C_\alpha(T) = \sum_{m=1}^{|T|} N_m Q_m(T) + \alpha |T| \quad (3.5)$$

ove il parametro $\alpha \geq 0$ rappresenta il peso dato alla componente di complessità della funzione di costo C ; fissato α , si sceglierà l'albero T che minimizza la funzione di costo $C_\alpha(T)$. Per quanto riguarda la scelta del parametro α , in letteratura sono stati proposti diversi metodi. Si può dimostrare che per ogni valore di α esiste un unico sotto-albero ottimo T_α che minimizza la funzione di costo C_α ed è contenuto in ogni altro sotto-albero con lo stesso valore di C_α . La scelta di α è cruciale per il procedimento di potatura: valori elevati di α porteranno a prediligere alberi di dimensione ridotta, mentre valori bassi porteranno ad alberi più vicini all'albero completo. In particolare, per $\alpha = 0$ si ottiene proprio l'albero iniziale T_0 .

Quando, come nel nostro caso, l'interesse è quello di massimizzare la capacità predittiva del modello, in letteratura si suggerisce di determinare il valore ottimo di α per il problema in esame tramite un procedimento di cross-validation o out-of-sample validation: si sceglierà il valore α^* che conduce all'albero ottimo T_{α^*} che minimizza l'errore quadratico medio di previsione.

I modelli ad albero presentano numerosi vantaggi:

- Trattandosi di un modello non parametrico, non è soggetto a rigide ipotesi sulla funzione di regressione da stimare ed è in grado di cogliere anche relazioni fortemente non lineari.
- La struttura ad albero è immediatamente interpretabile e, data la relativa semplicità del modello, i suoi meccanismi sono facilmente comprensibili anche ai non esperti in materia.
- La selezione delle variabili esplicative rilevanti viene effettuata automaticamente nel procedimento di costruzione dell'albero.

- La struttura gerarchica del modello permette di includere in maniera naturale gli effetti di interazione tra le variabili esplicative.

Allo stesso tempo, però, essi presentano anche diverse limitazioni:

- A causa del bipartizionamento ricorsivo, il modello fatica a catturare strutture puramente additive (prive di effetti di interazione) né è incentivato a cogliere la regolarità nelle relazioni che stima.
- La funzione costante a tratti prodotta dal modello può essere inadeguata in un contesto di regressione, ove si potrebbe voler calibrare una funzione continua.
- Il modello produce, generalmente, risultati poco robusti: una piccola variazione nei dati può portare l'algoritmo ad effettuare una partizione dello spazio anche molto diversa. Ciò è dovuto alla struttura gerarchica del modello: dopo le prime suddivisioni, il numero di dati in ogni regione diminuisce significativamente, rendendo le stime più instabili, e variazioni nelle prime suddivisioni si propagano per tutta la struttura dell'albero.

I primi due punti possono essere risolti considerando strutture più complesse nei nodi, ad esempio passando a modelli di tipo MARS, che abbiamo brevemente introdotto nella sezione precedente. Il terzo, tuttavia, è il punto più critico del modello analizzato. Nel seguito, vedremo alcune tecniche atte a mitigare questo problema.

3.1.2 Foreste casuali

Le foreste casuali (random forests) rappresentano un'evoluzione del tradizionale modello ad albero esposto nella sezione precedente. Esse si basano sulla tecnica di apprendimento ensemble del bagging, o bootstrap aggregating, per la quale si aggregano tanti modelli semplici - nel nostro caso alberi di regressione - stimati ognuno su un campione bootstrap del campione originale. Infine, la previsione del modello complessivo sarà data dalla media aritmetica delle previsioni dei singoli modelli di cui è costituito.

Si verifica (Hastie et al. (2009)) che le procedure di bagging possono ridurre drasticamente la variabilità di strutture singolarmente poco robuste, quali gli alberi di regressione.

Oltre al bagging di numerosi alberi di regressione, l'algoritmo alla base delle foreste casuali, proposto da Breiman (2001), prevede che ogni albero sia sviluppato considerando soltanto m variabili esplicative selezionate casualmente tra le p variabili esplicative da inserire nel modello (ovviamente $m \leq p$).

Per sviluppare il modello da un punto di vista teorico, indichiamo con $\boldsymbol{\theta}$ il campione relativo al vettore p -variato di variabili esplicative nel modello e con $\boldsymbol{\theta}_k$ il campione (ottenuto, nella nostra applicazione, applicando tecniche bootstrap al campione di partenza) relativo al vettore m -variato delle variabili esplicative ammissibili per il k -esimo albero della foresta.

Indichiamo poi con $h(\mathbf{x})$ una generica previsione per una nuova realizzazione \mathbf{x} delle variabili esplicative. Definiamo allora l'errore quadratico medio di generalizzazione²:

$$MSG E(h(\mathbf{X})) = \mathbb{E}_{(Y, \mathbf{X})} ((Y - h(\mathbf{X}))^2) \quad (3.6)$$

Più nel dettaglio, fissato il campione $\boldsymbol{\theta}$, indichiamo con $h(\mathbf{x}; \boldsymbol{\theta})$ la previsione per \mathbf{x} fornita da un albero di regressione e con $h(\mathbf{x}; \boldsymbol{\theta}_k)$ la previsione fornita dal k -esimo albero della foresta. La previsione complessiva di una foresta di K alberi sarà data da:

$$h^{(K)}(\mathbf{x}) = \frac{1}{K} \sum_{k=1}^K h(\mathbf{x}; \boldsymbol{\theta}_k) \quad (3.7)$$

Sfruttando la legge forte dei grandi numeri, Breiman (2001) dimostra che per $K \rightarrow +\infty$, $MSG E(h^{(K)}(\mathbf{X}))$ tende quasi certamente a $MSG E(\mathbb{E}_{\boldsymbol{\Theta}}(h(\mathbf{X}; \boldsymbol{\Theta})))$.

Inoltre, si dimostra anche che:

$$MSG E(h^{(K)}(\mathbf{X})) \leq \bar{\rho} \mathbb{E}_{\boldsymbol{\Theta}}(MSG E(h(\mathbf{X}; \boldsymbol{\Theta}))) \quad (3.8)$$

²Come da convenzione, indicheremo con $\mathbb{E}_{(Y, \mathbf{X})}(\cdot)$ la speranza matematica rispetto al vettore aleatorio (Y, \mathbf{X}) , le cui determinazioni rappresentano nuove realizzazioni della variabile dipendente e delle esplicative e con $\mathbb{E}_{\boldsymbol{\Theta}}(\cdot)$ la speranza matematica considerando come aleatorio il campione utilizzato per calibrare il modello.

ove $\bar{\rho}$ rappresenta, interpretativamente, una media ponderata dei coefficienti di correlazione tra gli errori di previsione $Y - h(\mathbf{X})$ relativi ai diversi alberi della foresta. Per i dettagli formali sulla costruzione di $\bar{\rho}$ e per la dimostrazione dei risultati qui enunciati si rimanda a Breiman (2001).

Interpretativamente, la formula (3.8) ci dice che vi sono due fattori che influiscono sulla limitazione superiore all'errore di generalizzazione commesso dalla foresta: il primo è legato all'errore di generalizzazione atteso commesso dagli alberi da cui è costituita, il secondo alla correlazione tra gli errori di previsione dei diversi alberi.

Il processo di randomizzazione delle variabili esplicative utilizzate per sviluppare ogni albero della foresta mira proprio a ridurre il fattore legato alla correlazione. Naturalmente, l'utilizzo di un numero ridotto di variabili esplicative ad ogni iterazione comporterà un incremento dell'errore commesso singolarmente da ogni albero, per cui vi è un trade-off tra le due componenti e non vi è una regola generale per la determinazione del numero ottimale m di variabili esplicative con cui far crescere ogni albero. Una regola empirica spesso utilizzata nella pratica suggerisce $m = p/3$, ma il valore migliore del parametro per il problema in esame potrà essere determinato solo confrontando diverse possibilità sulla base della bontà previsiva del modello risultante. In questo contesto, il metodo solitamente utilizzato per valutare la capacità previsiva del modello è l'errore out-of-bag, calcolato effettuando una previsione per ogni elemento del campione originale a partire da quegli alberi della foresta per i quali tale elemento non era incluso nel campione bootstrap utilizzato per svilupparli e confrontandola con il valore effettivo. Si sceglierà quindi il valore m che minimizza l'errore out-of-bag.

3.1.3 Tecniche di boosting

Come il bagging, anche il boosting è una tecnica di apprendimento ensemble che mira ad unire le previsioni di tanti modelli di apprendimento “deboli” per ottenere un modello dalle capacità predittive più accurate. Per quanto i suoi principi possano essere applicati a qualunque modello di regressione, il boosting è solitamente utilizzato in congiunzione con gli alberi decisionali. Il modello sarà quindi composto da tanti alberi di regressione, questa volta non più costruiti in maniera indipendente sulla base di tanti campioni bootstrap, ma in maniera sequenzia-

le, per cui ogni albero viene sviluppato su una versione modificata del campione originale, ricavata utilizzando informazioni derivanti dagli alberi precedenti.

Il primo algoritmo di boosting applicato con successo nella pratica, “AdaBoost.M1”, è stato proposto da Freund & Schapire nel 1997 per migliorare la bontà di un classificatore; esso prevedeva che ad ogni iterazione dell’algoritmo, il classificatore (nel nostro caso, un albero) fosse calibrato applicando un sistema di pesi alle osservazioni del campione, con l’obiettivo di pesare maggiormente le osservazioni che erano state classificate in maniera scorretta nelle iterazioni precedenti - in questo modo, le osservazioni più difficili da classificare ottengono un peso via via maggiore, per cui i classificatori successivi saranno forzati a concentrarsi su di loro. Infine, la previsione fornita dal modello complessivo sarà pari alla media aritmetica pesata delle previsioni ottenute dai diversi classificatori, con pesi maggiori assegnati ai classificatori che sono risultati più precisi.

Per una descrizione più dettagliata dell’algoritmo “AdaBoost.M1” e per ulteriori informazioni sulle generalizzazioni immediate di questo algoritmo e sulle sue estensioni al problema di regressione si rimanda a Hastie et al. (2009).

Per introdurre invece da un punto di vista formale l’algoritmo di gradient boosting sviluppato in Friedman (2001) e da noi considerato per la parte applicativa, consideriamo un generico previsore h - che supponiamo dipendere da un vettore \mathbf{p} di parametri - e indichiamo con $h(\mathbf{x}; \mathbf{p})$ la previsione per una generica nuova osservazione del vettore di variabili esplicative \mathbf{X} . In particolare, ci limiteremo ai previsori derivanti da modelli ensemble additivi, in cui la previsione fornita dal modello complessivo sarà una combinazione lineare delle previsioni $h_i(\mathbf{x}; \mathbf{p}_i)$ dei singoli modelli di cui è composto:

$$h(\mathbf{x}; \mathbf{p}) = \sum_{i=1}^n \alpha_i h_i(\mathbf{x}; \mathbf{p}_i) \quad (3.9)$$

Si noti che sia le foreste casuali, sia gli algoritmi di boosting rientrano in questa categoria. In particolare, nelle foreste casuali il peso assegnato a tutti i sotto-modelli è costante (si veda la formula (3.7)), mentre con il boosting esso varia a seconda della bontà di ogni sotto-previsore h_i .

Fissata una funzione di perdita L , la stima globale dei parametri \mathbf{p} solitamente consiste nella minimizzazione della perdita sulle osservazioni del campione utilizzato per il training del modello³:

$$\hat{\mathbf{p}} = \arg \min_{\mathbf{p}} \sum_{j=1}^N L(y_j, h(\mathbf{x}_j; \mathbf{p})) = \arg \min_{\alpha_1, \dots, \alpha_n, \mathbf{p}_1, \dots, \mathbf{p}_n} \sum_{j=1}^N L \left(y_j, \sum_{i=1}^n \alpha_i h_i(\mathbf{x}_j; \mathbf{p}_i) \right)$$

ove abbiamo indicato con $(y_1, \mathbf{x}_1), \dots, (y_N, \mathbf{x}_N)$ il campione utilizzato per il training. Un tale procedimento di ottimizzazione globale risulta spesso computazionalmente infattibile, per cui si rende necessaria una semplificazione. Una proposta, la cui logica è alla base del gradient boosting, è data dal forward stagewise modeling, per cui i diversi sotto-modelli sono stimati in via sequenziale, e i loro parametri aggiunti al modello additivo della formula (3.9), senza modificare i parametri relativi ai modelli inseriti in precedenza: al passo k -esimo vengono determinati i valori ottimi di (α_k, \mathbf{p}_k) da aggiungere al modello con già stimati i parametri (α_i, \mathbf{p}_i) per $i < k$. In formule:

$$(\hat{\alpha}_k, \hat{\mathbf{p}}_k) = \arg \min_{\alpha_k, \mathbf{p}_k} \sum_{j=1}^N L \left(y_j, \sum_{i=1}^{k-1} \hat{\alpha}_i h_i(\mathbf{x}_j; \hat{\mathbf{p}}_i) + \alpha_k h_k(\mathbf{x}_j; \mathbf{p}_k) \right) \quad (3.10)$$

In questo modo, il modello complessivo è stimato risolvendo in sequenza n problemi di ottimizzazione del tipo (3.10).

In particolare, considerata la funzione di perdita quadratica $L(y, x) = (y - x)^2$, si ha che $\forall j$ vale:

$$\begin{aligned} L \left(y_j, \sum_{i=1}^{k-1} \hat{\alpha}_i h_i(\mathbf{x}_j; \hat{\mathbf{p}}_i) + \alpha_k h_k(\mathbf{x}_j; \mathbf{p}_k) \right) &= \left(y_j - \sum_{i=1}^{k-1} \hat{\alpha}_i h_i(\mathbf{x}_j; \hat{\mathbf{p}}_i) - \alpha_k h_k(\mathbf{x}_j; \mathbf{p}_k) \right)^2 \\ &= (r_{j,k-1} - \alpha_k h_k(\mathbf{x}_j; \mathbf{p}_k))^2 \end{aligned}$$

ove abbiamo indicato con $r_{j,k-1} = y_j - \sum_{i=1}^{k-1} \hat{\alpha}_i h_i(\mathbf{x}_j; \hat{\mathbf{p}}_i)$ il residuo del modello costruito fino alla $(k-1)$ -esima iterazione, ossia il modello che la k -esima iterazione si propone di migliorare.

³In questo contesto stiamo considerando $\mathbf{p} = \{\alpha_1, \dots, \alpha_n, \mathbf{p}_1, \dots, \mathbf{p}_n\}$. Non stiamo includendo nella simbologia i parametri d'insieme del modello ensemble, come ad esempio il numero n di sotto-modelli di cui è composto. Tali parametri infatti spesso sono considerati come fissati prima del procedimento di apprendimento statistico (c.d. "iperparametri") e quindi non rientrano nel processo di ottimizzazione studiato in questo paragrafo.

Il problema di minimo della k -esima iterazione allora diventa:

$$(\hat{\alpha}_k, \hat{\mathbf{p}}_k) = \arg \min_{\alpha_k, \mathbf{p}_k} \sum_{j=1}^N (r_{j,k-1} - \alpha_k h_k(\mathbf{x}_j; \mathbf{p}_k))^2 \quad (3.11)$$

In sostanza, quindi, se ci poniamo nel contesto della perdita quadratica, ad ogni iterazione viene sviluppato un modello che considera come variabile dipendente i residui del modello dell'iterazione precedente. Nella nostra applicazione, in particolare, ogni modello sarà dato da un albero di regressione.

Questo approccio può essere generalizzato per una qualunque funzione di perdita differenziabile tramite l'algoritmo di discesa del gradiente, un metodo numerico per la soluzione di (3.10) basato sul gradiente della funzione di perdita rispetto ai valori dello stimatore nel campione utilizzato per il training. Il generico elemento del gradiente da considerare nella k -esima iterazione è:

$$g_{j,k-1} = \left. \frac{\partial L(y_j, h(\mathbf{x}_j; \mathbf{p}))}{\partial h(\mathbf{x}_j; \mathbf{p})} \right|_{h(\mathbf{x}_j; \mathbf{p})=h^{(k-1)}(\mathbf{x}_j; \mathbf{p})} \quad (3.12)$$

ove abbiamo posto $h^{(k-1)}(\mathbf{x}_j; \mathbf{p}) = \sum_{i=1}^{k-1} \alpha_i h_i(\mathbf{x}_j; \hat{\mathbf{p}}_i)$. Infine l'algoritmo di discesa del gradiente prevede che il k -esimo modello sia calibrato sui dati $-g_{j,k-1}$, ossia, interpretativamente, si cerca di minimizzare la perdita muovendosi nella direzione opposta rispetto al suo gradiente, cioè nella direzione di "discesa più ripida" della funzione.

Nel caso della perdita quadratica, si vede immediatamente dalla (3.12) che $-g_{j,k-1} = r_{j,k-1} \forall j$, per cui il vettore dei residui del modello $k-1$ coincide con l'opposto del gradiente⁴; ne segue che la metodologia esposta nel caso della perdita quadratica può essere interpretata come caso particolare di discesa del gradiente. Nella nostra applicazione, ci limiteremo alla perdita quadratica, per cui non approfondiremo ulteriormente questa tematica; una spiegazione più formale relativamente alla struttura generale dell'algoritmo di discesa del gradiente e alla sua applicazione per la soluzione di un problema di ottimo come in (3.10), nonché implementazioni

⁴In realtà affinché valga questo risultato bisogna definire la perdita quadratica come $L(y, x) = (y - x)^2/2$, anziché come $L(y, x) = (y - x)^2$ come abbiamo fatto noi finora; in ogni caso le due formulazioni sono equivalenti e conducono alla stessa soluzione del problema di ottimo.

con funzioni di perdita diverse dalla quadratica sono rinvenibili in Friedman (2001) e Hastie et. al (2009).

Per quanto riguarda il procedimento di costruzione dei singoli alberi di cui è costituito il modello, esso solitamente ricalca a grandi linee quanto esposto nella sezione 3.1.1, ma non si procede allo sviluppo di un albero molto grande e alla successiva potatura, in quanto, in questa applicazione, ogni albero non rappresenta che una piccola parte del modello complessivo; essendovi ulteriori iterazioni, infatti, è sufficiente un modello meno complesso di quello che sarebbe selezionato in una procedura di questo tipo. Pertanto, solitamente, nelle tecniche di boosting si prediligono metodi semplificati per la costruzione degli alberi. In particolare, nella nostra applicazione la struttura di ogni albero sarà regolata da due parametri:

- la profondità i di ogni albero, che rappresenta il numero di split effettuati nella procedura di partizionamento ricorsivo delineata nella sezione 3.1.1; il parametro controlla la complessità del modello utilizzato ed è detto anche grado di interazione, in quanto essendovi i split, i rappresenta il numero massimo di variabili considerabili congiuntamente in ogni iterazione.
- Il numero minimo o di osservazioni nei nodi terminali dell'albero; esso rappresenta un'ulteriore limitazione per evitare l'overfitting, impedendo al meccanismo di costruzione di effettuare split su poche osservazioni.

L'algoritmo di gradient boosting proposto in Friedman (2001) e qui presentato finora è stato successivamente esteso in Friedman (2002) introducendo una componente di aleatorietà: ad ogni iterazione, il modello non viene stimato su tutto il dataset adibito al training, ma solo su una frazione selezionata casualmente (la proposta di Friedman (2002), derivante da analisi empiriche, è di considerare solo il 50% del dataset ad ogni iterazione), con l'obiettivo di decorrelare le stime di ogni iterazione; si è verificato infatti che in molti problemi il miglioramento dovuto alla decorrelazione più che compensa la diminuzione nella precisione di ogni sotto-previsore del modello, dovuta alla riduzione del numero di dati utilizzato. Questo metodo, denominato stochastic gradient boosting, introduce forzatamente nel boosting l'aleatorietà insita in via naturale nelle tecniche di bagging, ivi dovuta alla procedura di bootstrap, e permette di utilizzare anche in questo contesto l'errore out-of-bag come misura della bontà di previsione del modello.

Un'ulteriore generalizzazione del modello consiste nell'introduzione di un parametro di contrazione (shrinkage) $\lambda \in]0, 1]$, per cui alla struttura additiva del modello delineata nella (3.9) ad ogni iterazione il nuovo sotto-predittore aggiunto viene moltiplicato per λ . Essendo solitamente $\lambda < 1$, esso ha lo scopo di rallentare ulteriormente il processo di apprendimento del modello, permettendo ad un numero maggiore di alberi diversi di "aggredire" i residui. A parità del numero n di alberi (cioè di iterazioni del modello), un valore minore di λ porta ad una riduzione della bontà di adattamento al campione utilizzato per il training, ma anche, quindi, ad una riduzione del rischio di overfitting; analogamente, lo stesso effetto si ha riducendo n a parità di λ . Complessivamente, quindi, la coppia di parametri (n, λ) regola il grado di adattamento del modello ai dati. Si è verificato empiricamente che valori bassi di λ possono migliorare la bontà del modello, ma richiedono un numero anche molto maggiore di iterazioni, quindi il costo computazionale aumenta. La scelta ottimale della coppia dipenderà, come sempre, dal problema in esame; i valori solitamente consigliati in letteratura sono tra 0.001 e 0.1 per λ e $n > 100$.

Rispetto ai singoli alberi decisionali da cui è costituito, il gradient boosting può portare ad un miglioramento significativo nelle capacità di adattamento e previsione, tuttavia il suo processo di costruzione risulta meno interpretabile. Per facilitarne l'interpretabilità è stata proposta l'importanza relativa:

$$\mathcal{I}(l) = \frac{\sqrt{\frac{1}{n} \sum_{k=1}^n \sum_{t=1}^{i-1} \hat{\xi}_{t,k} \mathbb{1}_{v(t)=l}}}{\sum_{q=1}^p \sqrt{\frac{1}{n} \sum_{k=1}^n \sum_{t=1}^{i-1} \hat{\xi}_{t,k} \mathbb{1}_{v(t)=q}}} \quad (3.13)$$

ove $\hat{\xi}_{t,k}$ rappresenta una stima della riduzione di errore quadratico medio dovuta al t -esimo split del k -esimo albero, e questa componente viene aggiunta alla somma solo se in quello split la variabile considerata è proprio l . In sostanza, viene effettuata una media (sugli n alberi) della somma dei guadagni in termini di MSE ottenuti in ogni nodo interno di ogni albero nel quale la variabile l è stata selezionata per effettuare la bipartizione; infine, il tutto viene diviso per la somma degli $\mathcal{I}(l)$ su l in modo da avere una misura relativa. $\mathcal{I}(l)$ quindi associa ad ogni variabile esplicativa l una misura di quanto essa ha contribuito alla costruzione del modello rispetto alle altre.

3.2 Applicazioni al contesto assicurativo

Nella parte precedente di questo capitolo, abbiamo introdotto alcuni tra i più importanti metodi di machine learning basati su strutture ad albero. Considereremo ora gli stessi dati introdotti nella sezione 2.3.1 e proveremo ad utilizzare i metodi di machine learning visti per costruire una funzione proxy per il valore dei portafogli di attività e passività. Per confrontare tra loro i diversi modelli, utilizzeremo sempre le metriche \mathcal{M}_1 , \mathcal{M}_2 , \mathcal{M}_3 , dando particolare importanza, per quanto riguarda le passività, alla metrica \mathcal{M}_3 , come anticipato nella sezione 2.3.

3.2.1 Metodi basati su alberi di regressione

Iniziamo con un'applicazione degli alberi di regressione ai portafogli di attivi delle tre compagnie in esame. L'algoritmo utilizzato è quello del pacchetto *rpart* su R. Le impostazioni utilizzate sono dapprima quelle standard dell'algoritmo; successivamente vengono effettuate alcune analisi relativamente a impostazioni alternative dei parametri del modello (in particolare, relativamente al parametro di complessità α , al numero minimo di osservazioni in ogni nodo terminale nel processo di sviluppo dell'albero, alla profondità massima dell'albero). Nella Tabella 3.1 sono riportati i risultati relativi all'impostazione standard e i risultati con i parametri personalizzati. Il tuning è stato effettuato mediante cross-validation utilizzando principalmente il pacchetto *caret* di R; abbiamo confrontato diversi valori dei parametri e scelto quello che minimizzava l'errore quadratico medio. Confrontando i risultati ottenuti nella tabella seguente con quelli riportati nelle tabelle 2.1, 2.2, 2.3, osserviamo una performance nettamente peggiore del modello ad albero rispetto a quello polinomiale.

Compagnia 1								
Categoria	Impostazioni standard				Tuning dei parametri			
	T	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	T	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Obbligazioni	0s	19.15%	2.77%	2.77%	2s	14.61%	2.53%	2.53%
Azioni	0s	124.38%	34.72%	33.38%	2s	91.35%	24.54%	22.84%

Compagnia 2

Categoria	Impostazioni standard				Tuning dei parametri			
	T	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	T	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Obbligazioni	0s	14.50%	3.99%	3.98%	2s	13.72%	3.35%	3.30%
Azioni	0s	64.66%	18.41%	17.93%	2s	44.22%	9.81%	9.91%

Compagnia 3

Categoria	Impostazioni standard				Tuning dei parametri			
	T	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	T	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Obbligazioni	0s	12.49%	2.45%	2.45%	2s	16.47%	2.38%	2.38%
Azioni	0s	127.03%	37.54%	34.52%	2s	78.58%	21.66%	19.59%
Derivati	0s	500.69%	18.07%	8.16%	2s	147.88%	5.44%	2.72%

Tabella 3.1: *Statistiche di bontà di previsione per una proxy degli attivi basata su un albero di regressione.*

Il modello basato sugli alberi di regressione quindi risulta inadeguato nel descrivere il fenomeno in esame. Per quanto la modellazione automatica dei termini d'interazione sia molto interessante, abbiamo osservato graficamente nella sezione 2.3.2 che le relazioni marginali risultano comunque monotone e, solitamente, piuttosto stabili. Un modello estremamente flessibile come un albero di regressione non ha alcun incentivo a cogliere tali regolarità nei dati, motivo per il quale i risultati forniti in previsione dal modello sono, nella nostra applicazione, scadenti. Inoltre, abbiamo osservato nella sezione 2.3 che, data l'assenza del ciclo interno di simulazione stocastica nella valutazione degli attivi in ogni scenario, i dati relativi agli attivi sono precisi, e il modello polinomiale riesce, solitamente, a catturare estremamente bene l'andamento del valore del portafoglio, con un errore medio spesso inferiore allo 0.5%: il modello di regressione si avvicina ad un modello di interpolazione. Un tale risultato non è ottenibile con il modello qui analizzato.

Passiamo ora all'applicazione piena del modello sui dati relativi ai portafogli di passività. Per brevità, nella Tabella 3.2 forniremo i risultati esclusivamente per il modello con i parametri ottimizzati tramite il processo di tuning. Il tempo di esecuzione riportato nella tabella include quello necessario per il tuning dei parametri.

Compagnia 1							
Categoria	T. esecuzione	Set di validazione 1			Set di validazione 2		
		\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Pensioni	14s	9.68%	2.45%	2.47%	10.56%	2.88%	2.85%
Protezione	1m07s	9.96%	2.13%	2.12%	191.71%	6.16%	4.99%
Risparmio	20s	4.45%	0.87%	0.87%	7.86%	1.76%	1.74%
Unit-Linked	16s	7.83%	1.02%	0.90%	16.11%	1.65%	1.48%

Compagnia 2							
Categoria	T. esecuzione	Set di validazione 1			Set di validazione 2		
		\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Pensioni	12s	6.11%	1.86%	1.85%	/	/	/
Protezione	22s	9.35%	2.54%	2.52%	/	/	/
Unit-Linked	11s	15.53%	3.81%	3.82%	/	/	/

Compagnia 3							
Categoria	T. esecuzione	Set di validazione 1			Set di validazione 2		
		\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Protezione	42s	73.48%	12.98%	13.46%	68.65%	18.16%	19.26%
Risparmio	20s	3.65%	0.80%	0.80%	8.63%	2.11%	2.12%

Tabella 3.2: *Statistiche di bontà di previsione per una proxy dei passivi basata su un albero di regressione con i parametri ottimizzati.*

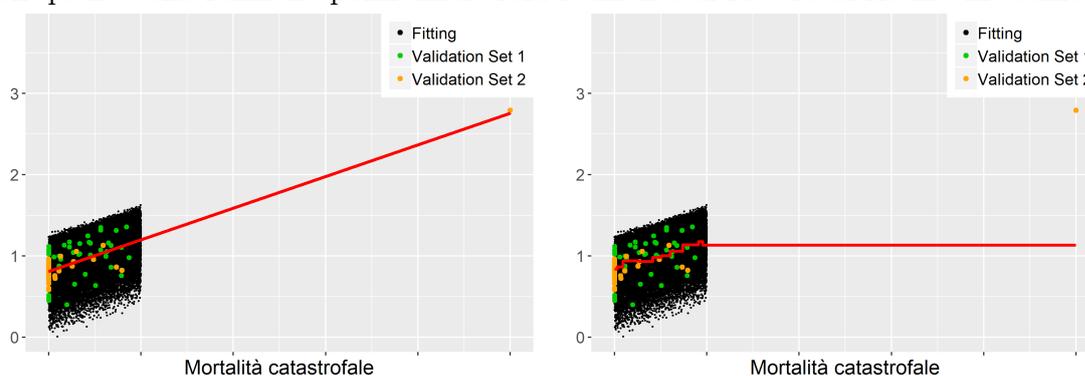
Anche qui, la struttura ad albero è troppo generale e conduce ad un modello eccessivamente semplificato, le cui capacità predittive sono peggiori di quelle del modello polinomiale, analizzate nelle tabelle 2.11, 2.12, 2.13. Le differenze sono meno marcate che nel caso delle attività: data la maggiore variabilità delle stime utilizzate per il fitting in questo contesto, il modello polinomiale ha molta più difficoltà a catturare in maniera perfetta l'andamento; data la significativamente maggiore quantità di dati a disposizione rispetto all'applicazione agli attivi, i modelli di machine learning riescono a svilupparsi in maniera più stabile.

Osserviamo in particolare come il modello ad albero porti ad un errore elevato laddove i fattori di rischio assumono negli scenari di validazione valori significativamente diversi da quelli utilizzati per il fitting: come avevamo già commentato in

merito alle applicazioni nella sezione 2.3.2 e come si vede graficamente nella stessa sezione, ciò accade per il fattore relativo alla mortalità catastrofica nelle LoB di Protezione delle Compagnie 1 e 3. Nella sezione 2.3.2 si era osservato che l'andamento marginale rispetto a tale fattore di rischio sembra continuare linearmente al di fuori dello spazio utilizzato per il fitting e che laddove l'algoritmo di selezione dei termini polinomiali considerava invece anche termini più elevati, le metriche di errore ne risultavano significativamente impattate. Nella stessa sezione, si era osservato anche che il problema poteva essere risolto con un aggiustamento manuale all'algoritmo di selezione, forzandolo a non selezionare termini di grado più elevato in quella situazione; attuando tale aggiustamento le metriche d'errore venivano riportate in linea.

Lo stesso problema osservato per il modello polinomiale si ha anche con la struttura ad albero: ad ogni valore al di fuori dello spazio utilizzato per la calibrazione viene associato il valore costante del nodo più vicino disponibile. Ne segue che il modello ad albero non è in grado di modellare andamenti non costanti al di fuori dello spazio considerato, né risulta possibile effettuare aggiustamenti manuali come nel modello polinomiale per forzare uno specifico andamento. Il modello ad albero quindi non risolve la limitazione osservata per la funzione polinomiale, ma anzi la rende ancora più significativa.

Nella figura seguente, compariamo l'andamento marginale del modello ad albero con quello della funzione polinomiale relativamente al fattore di rischio in esame.



Andamento marginale rispetto alla mortalità catastrofica: confronto tra modello polinomiale (a sinistra) e modello ad albero (a destra) per la BEL relativa alle polizze di protezione della Compagnia 1.

3.2.2 Metodi basati su foreste casuali

Passiamo all'implementazione delle foreste casuali sul portafoglio di attività tramite l'algoritmo del pacchetto R *randomForest*. Anche qui, proponiamo l'implementazione relativa all'impostazione base dell'algoritmo e quella con il parametro m scelto in modo da minimizzare l'errore out-of-bag tramite la funzione *tuneRF()*, sempre del pacchetto *randomForest*. Il numero di alberi nella foresta è stato fissato a 1'000. È infatti opinione largamente diffusa in letteratura che il numero di alberi della foresta in questo contesto non sia un parametro da ottimizzare, ma semplicemente da fissare "sufficientemente grande".

Compagnia 1								
Categoria	Impostazioni standard				Tuning dei parametri			
	T	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	T	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Obbligazioni	4s	13.95%	1.58%	1.58%	33s	10.96%	1.25%	1.25%
Azioni	4s	99.44%	27.08%	24.94%	31s	84.30%	23.60%	21.36%
Compagnia 2								
Categoria	Impostazioni standard				Tuning dei parametri			
	T	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	T	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Obbligazioni	4s	8.70%	1.80%	1.80%	26s	8.57%	1.71%	1.72%
Azioni	3s	55.76%	15.55%	14.74%	23s	43.80%	12.81%	12.12%
Compagnia 3								
Categoria	Impostazioni standard				Tuning dei parametri			
	T	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	T	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Obbligazioni	7s	14.54%	1.48%	1.47%	34s	10.18%	1.24%	1.24%
Azioni	4s	101.31%	24.96%	21.38%	36s	82.16%	21.02%	18.41%
Derivati	4s	661.42%	13.34%	6.99%	27s	213.16%	4.45%	1.84%

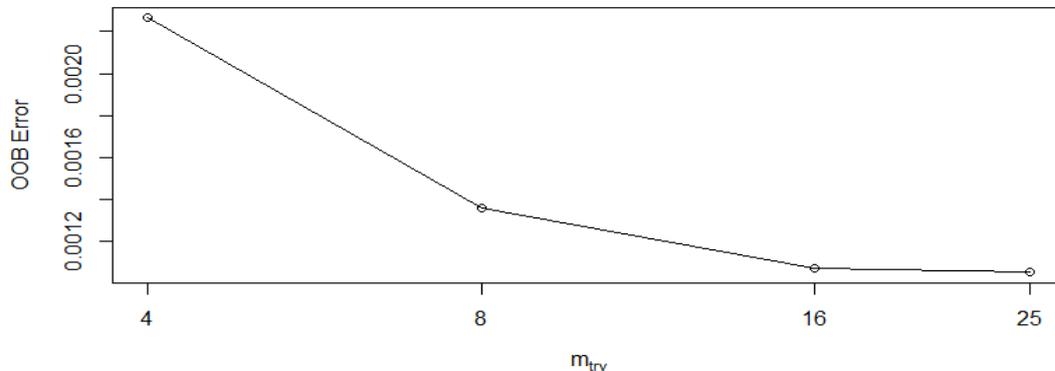
Tabella 3.3: *Statistiche di bontà di previsione per una proxy degli attivi basata su una foresta casuale.*

Come attendibile, i risultati della foresta esposti nella Tabella 3.3 in termini di errore medio (\mathcal{M}_2 e \mathcal{M}_3) sono sempre migliori dei corrispondenti risultati ottenuti con un singolo albero. Notiamo però che la foresta con le impostazioni standard

ottiene spesso risultati peggiori che un singolo albero ottimizzato.

Ciò presumibilmente è dovuto alla particolarità dei dati a disposizione, in cui la maggior parte delle variabili esplicative hanno un effetto marginalmente ridotto, mentre alcune hanno un effetto molto forte (ciò si può apprezzare nei grafici della sezione 2.3.2): con le impostazioni standard, solo $m = p/3$ variabili vengono utilizzate per calibrare ogni albero, quindi molti alberi risulteranno del tutto inefficaci nel spiegare l'andamento del portafoglio, dato che mancheranno loro le esplicative più importanti. Tale intuizione è confermata anche dall'algoritmo di tuning del parametro, che seleziona un valore di m sempre molto elevato, spesso pari proprio a p . Utilizzando il parametro ottimizzato, l'errore di previsione nella Tabella 3.3 risulta solitamente di nuovo minore di quello corrispondente nella Tabella 3.1. I risultati tuttavia sono accettabili solo sui portafogli obbligazionari, mentre gli errori rimangono estremamente elevati per le altre tipologie di attivi.

Nella figura seguente, riportiamo i risultati del tuning del parametro m per uno dei portafogli considerati (l'andamento è pressoché analogo per gli altri portafogli).



Andamento dell'errore out-of-bag al variare del parametro m , per uno dei portafogli considerati.

Ciò ci porta a considerare come valore ottimizzato di m proprio il numero massimo di variabili nel nostro modello; in altre parole, la foresta casuale in questo caso degenera in un semplice bagging di alberi.

Si è verificato inoltre che un aumento del numero di alberi utilizzato dal modello non comporta, di solito, una significativa variazione nelle statistiche considerate: 1'000 alberi a prima vista sembrano sufficienti a garantire la convergenza (abbiamo provato una calibrazione effettuata con 20'000 alberi su una LoB, con variazioni irrilevanti nelle metriche). Qualora si ritenesse di voler applicare un modello di questo tipo, analisi più complete in merito sarebbero opportune.

Una possibile estensione - computazionalmente più pesante - del metodo di tuning qui utilizzato consiste nell'ottimizzazione, oltre che di m , anche dei parametri relativi alla costruzione dei singoli alberi della foresta. Data tuttavia la debolezza dei risultati ottenuti con le foreste casuali e il significativo aumento del tempo di esecuzione che una procedura del genere comporterebbe, anche qui non abbiamo provveduto ad ulteriori analisi atte ad ottenere ad un'ottimizzazione più fine del modello: le metriche, infatti, sono sempre significativamente peggiori di quanto ottenuto tramite il modello polinomiale.

Procediamo ora con l'applicazione piena del modello sui dati relativi ai portafogli di passività, mostrando, nella Tabella 3.4, i risultati in previsione. Per brevità, i valori saranno riportati esclusivamente per la configurazione con il parametro m ottimizzato. Anche qui il tempo di esecuzione riportato include quello necessario per il tuning, che, tuttavia, per velocizzare l'algoritmo, qui è effettuato considerando soltanto 100 anziché 1'000 alberi; effettuare il tuning su 1'000 alberi comporta infatti un tempo di esecuzione di circa un'ora e tendenzialmente porta ad un valore ottimo analogo (o quantomeno molto simile) di m . Il modello finale su cui sono calcolate le statistiche considera invece, come nell'applicazione precedente, 1'000 alberi.

Compagnia 1								
Categoria	T. esecuzione	m/p	Set di validazione 1			Set di validazione 2		
			\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Pensioni	16m01s	0.62	3.35%	0.88%	0.87%	6.79%	2.33%	2.37%
Protezione	14m04s	0.60	4.04%	2.06%	2.13%	195.38%	5.76%	4.58%
Risparmio	19m28s	0.57	1.18%	0.37%	0.37%	8.26%	2.19%	2.15%
Unit-Linked	23m32s	0.62	6.66%	1.06%	0.82%	28.94%	1.63%	1.33%

Compagnia 2								
Categoria	T. esecuzione	m/p	Set di validazione 1			Set di validazione 2		
			\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Pensioni	9m44s	0.30	7.89%	1.23%	1.23%	/	/	/
Protezione	19m27s	0.60	3.39%	1.32%	1.29%	/	/	/
Unit-Linked	12m28s	0.29	6.37%	1.61%	1.57%	/	/	/

Compagnia 3								
Categoria	T. esecuzione	m/p	Set di validazione 1			Set di validazione 2		
			\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Protezione	34m24s	1.00	69.38%	8.53%	9.39%	76.08%	11.74%	13.77%
Risparmio	24m39s	0.62	1.38%	0.34%	0.34%	8.73%	1.63%	1.64%

Tabella 3.4: *Statistiche di bontà di previsione per una proxy dei passivi basata su una foresta casuale.*

Dalla Tabella 3.4 notiamo innanzitutto come il valore ottimale di m è, nelle nostre applicazioni, molto spesso vicino a $2/3p$ anziché al valore standard di $p/3$. I risultati ottenuti dalla foresta sull'applicazione piena del modello proxy sono significativamente migliori rispetto a quelli nel modello semplificato per i portafogli di attività. Tuttavia, il modello polinomiale risulta, nella maggior parte delle nostre applicazioni, ancora nettamente superiore, con un errore assoluto medio che molto spesso è meno della metà di quello riportato nella Tabella 3.4.

Osserviamo poi che, come il modello ad albero, anche la foresta non è in grado di modellare accuratamente il valore del portafoglio in scenari in cui i fattori di rischio assumono valori significativamente distanti da quelli utilizzati per calibrare il modello: l'errore nelle LoB di protezione risulta sempre estremamente elevato. Nel caso della terza compagnia, la performance nel portafoglio anomalo è significativamente migliore di quella ottenuta con l'algoritmo di selezione automatica dei termini polinomiali (ma anche molto peggiore di quella ottenibile configurando manualmente la funzione polinomiale del modello LSMC). Siamo portati a ritenere, tuttavia, che tale miglioramento sia maggiormente dovuto al caso piuttosto che ad un'effettiva maggiore capacità di estrapolazione.

3.2.3 Metodi basati sul gradient boosting

Sviluppiamo infine un modello per attivi e passivi utilizzando le tecniche di gradient boosting tramite la funzione $gbm()$ del pacchetto R *gbm*. Per il tuning del modello, considereremo quattro parametri, il cui significato è delineato nella sezione 3.1.3:

- Il numero n di alberi nel modello;
- Il grado massimo i di profondità di ogni albero;
- Il numero minimo di osservazioni o in ogni nodo terminale di ogni albero;
- Il parametro di apprendimento λ .

Il tuning sarà effettuato considerando una griglia di ricerca consistente nei seguenti valori possibili, per un totale di 160 modelli:

$$n = (250, 500, 1000, 5000)$$

$$i = (3, 4, 5, 6, 10)$$

$$o = (10, 40, 60, 100)$$

$$\lambda = (0.1, 0.05)$$

Ovviamente un processo di tuning più fine potrà portare ad un miglioramento della capacità predittiva del modello, nella nostra analisi ci siamo limitati ai pochi valori riportati sopra al fine di ottenere risultati in tempi ragionevoli. Inoltre, algoritmi di tuning più veloci della ricerca a griglia possono risultare interessanti per ottimizzare il processo di calibrazione in un'implementazione pratica del modello. Nella Tabella 3.5 sono riportati i risultati del modello ottimizzato per gli attivi. Osserviamo subito che i risultati ottenuti sono significativamente migliori di quelli relativi ad un singolo albero di regressione o ad una foresta casuale, riportati nelle Tabelle 3.1 e 3.3. Tuttavia, nonostante la complessità del modello e l'ottimizzazione effettuata, anche la gradient boosting machine non riesce ad ottenere risultati adeguati, risultando sempre nettamente peggiore del modello polinomiale. Come abbiamo già detto relativamente al modello ad albero, si ritiene che ciò sia dovuto all'estrema bontà del modello polinomiale, che in questo contesto cattura benissimo il valore calcolato in maniera precisa nel ridotto numero di scenari

disponibile, avvicinandosi ad un modello di interpolazione. I modelli di machine learning, elaborati per “imparare” da un gran numero di dati molto più irregolari, non ottengono risultati altrettanto buoni in questo contesto semplice.

Compagnia 1								
Categoria	T. Esecuzione	Parametri scelti				Metriche di validazione		
		n	i	o	λ	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Obbligazioni	2m10s	5000	3	10	0.1	4.79%	0.58%	0.57%
Azioni	1m34s	5000	5	100	0.05	18.14%	3.29%	3.28%
Compagnia 2								
Categoria	T. Esecuzione	Parametri scelti				Metriche di validazione		
		n	i	o	λ	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Obbligazioni	1m36s	5000	5	60	0.05	4.47%	0.71%	0.71%
Azioni	1m41s	1000	10	40	0.1	20.12%	2.45%	2.63%
Compagnia 3								
Categoria	T. Esecuzione	Parametri scelti				Metriche di validazione		
		n	i	o	λ	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Obbligazioni	2m48s	5000	3	10	0.1	4.55%	0.53%	0.53%
Azioni	1m44s	1000	10	60	0.05	42.77%	5.15%	4.34%
Derivati	2m03s	5000	3	10	0.05	117.87%	2.19%	1.07%

Tabella 3.5: *Statistiche di bontà di previsione per una proxy degli attivi basata su una gradient boosting machine ottimizzata.*

Osserviamo come il numero di alberi n nel modello è consistentemente scelto elevato. Il parametro n , come abbiamo osservato nella sezione 3.1.3, è strettamente legato all’adattamento al campione: valori alti portano ad un potenziale pericolo di overfitting, valori bassi a un peggiore adattamento al campione. Considerato che i dati sugli attivi sono composti da valori calcolati in maniera molto precisa, il risultato dell’algoritmo di tuning non dovrebbe stupire. Per quanto riguarda gli altri parametri, invece, non notiamo una particolare regolarità nella scelta effettuata. Ribadiamo tuttavia che vi sono metodi più evoluti per la ricerca del valore ottimo dei parametri: un processo di tuning più efficiente e/o più fine sarebbe auspicabile per un’implementazione pratica del modello. Tuttavia, considerata la

debolezza della performance delle tecniche studiate in questo capitolo nel modello semplificato per le attività, non analizzeremo ulteriormente questa applicazione del modello proxy, ritenendo che i risultati raggiunti con il modello polinomiale non possano essere migliorati mediante i modelli di machine learning ad albero. Nella Tabella 3.6 invece riportiamo i risultati della calibrazione per le passività delle tre compagnie. Il processo di tuning dei parametri n , i , o , λ ricalca quello effettuato nel caso precedente. Inoltre, avendo in questo caso a disposizione molti più dati, abbiamo utilizzato il gradient boosting stocastico proposto da Friedman (2002), utilizzando l'80% del dataset ad ogni iterazione.

Compagnia 1											
Categoria	T	Parametri scelti				Set di validazione 1			Set di validazione 2		
		n	i	o	λ	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Pensioni	57m15s	250	6	40	0.05	1.87%	0.51%	0.52%	4.46%	1.79%	1.79%
Protezione	41m38s	5000	5	40	0.05	1.25%	0.24%	0.22%	197.37%	5.20%	3.96%
Risparmio	55m49s	500	6	10	0.1	1.00%	0.30%	0.30%	8.18%	2.19%	2.16%
Unit-Linked	54m36s	250	4	100	0.05	4.82%	0.74%	0.65%	18.19%	0.83%	0.67%

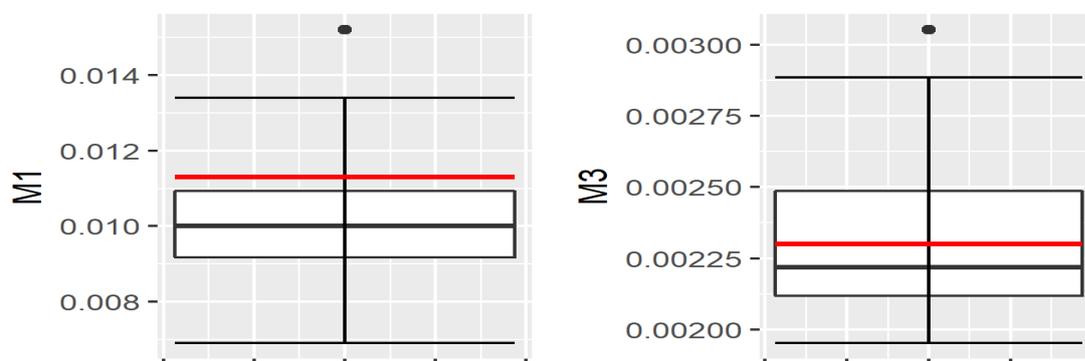
Compagnia 2											
Categoria	T	Parametri scelti				Set di validazione 1			Set di validazione 2		
		n	i	o	λ	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Pensioni	41m23s	250	3	5	0.1	2.12%	0.56%	0.56%	/	/	/
Protezione	43m22s	500	10	60	0.05	1.30%	0.38%	0.37%	/	/	/
Unit-Linked	59m20s	500	4	10	0.1	2.75%	0.83%	0.83%	/	/	/

Compagnia 3											
Categoria	T	Parametri scelti				Set di validazione 1			Set di validazione 2		
		n	i	o	λ	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
Protezione	1h2m58s	5000	3	100	0.1	69.16%	1.94%	3.67%	74.14%	15.82%	16.50%
Risparmio	1h6m35s	1000	10	60	0.1	0.87%	0.21%	0.21%	9.64%	1.55%	1.58%

Tabella 3.6: *Statistiche di bontà di previsione per una proxy dei passivi basata su una stochastic gradient boosting machine ottimizzata.*

I risultati nella Tabella 3.6 sono in linea con quelli del modello LSMC presentati nella sezione 2.3.2. In particolare, osserviamo un leggero peggioramento nelle me-

triche per le Compagnie 1 e 2 e un leggero miglioramento per la Compagnia 3. Nel nostro esempio di applicazione, abbiamo applicato una griglia relativamente poco granulare nella ricerca degli iperparametri, in modo da avere un'idea generale di quali valori possano essere adeguati per il problema in esame; un affinamento della griglia porterebbe sicuramente a risultati migliori, a scapito del tempo di esecuzione. Viceversa, ridurre il numero di combinazioni possibili potrebbe accorciare il tempo di esecuzione - si è visto, in particolare, che stimare un singolo modello richiede solitamente meno di un minuto, e che la maggior parte del tempo di tuning mostrato nella tabella sopra è dovuto ai giri effettuati con $n = 5000$. Vi è, quindi, ancora un discreto margine di miglioramento nei risultati qui presentati. Si noti tuttavia che, dato che stiamo utilizzando soltanto pochi scenari per calcolare le metriche out-of-sample, l'utilizzo del gradient boosting stocastico comporta che ogni giro dell'algoritmo condurrà ad una potenziale leggera instabilità nei risultati mostrati nella tabella sopra. Per comprenderne l'impatto, abbiamo provato ad utilizzare 100 seed diversi per la LoB Protezione della Compagnia 3. Nella figura seguente sono raffigurati i valori risultanti delle metriche \mathcal{M}_1 e \mathcal{M}_3 sul primo set di validazione.

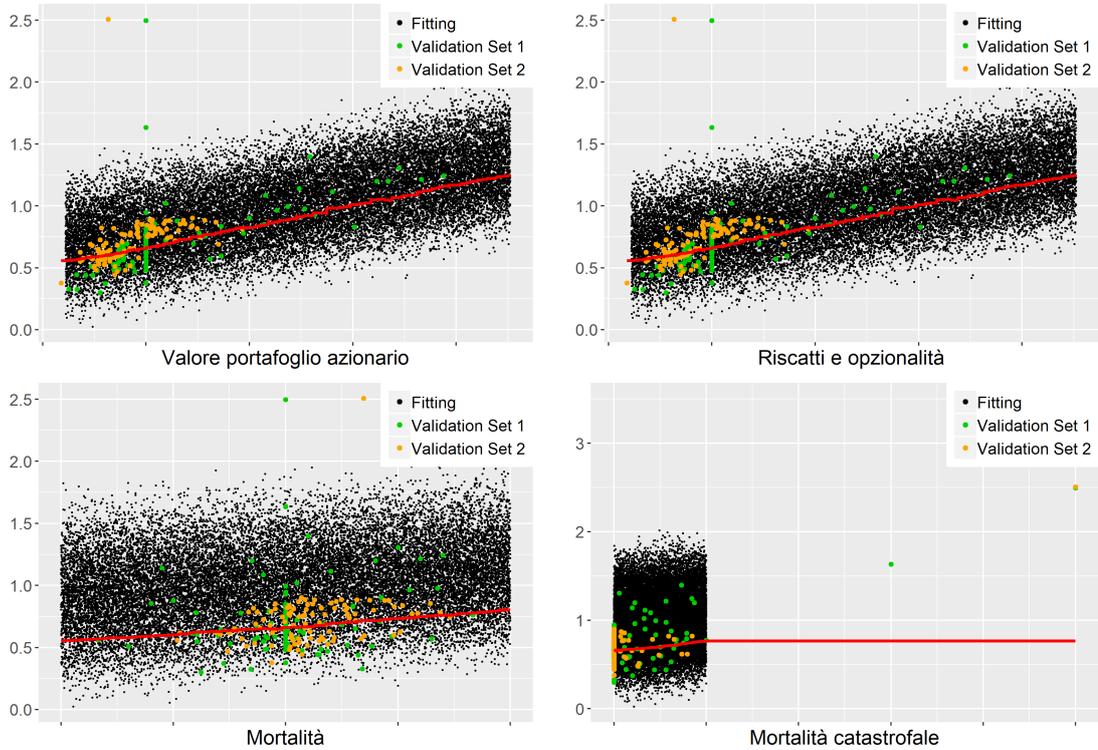


Metriche \mathcal{M}_1 e \mathcal{M}_3 sul primo set per la LoB Risparmio della Compagnia 3, considerando 100 giri distinti della gradient boosting machine stocastica; la linea rossa indica il valore ottenuto considerando tutto il set di training ad ogni iterazione.

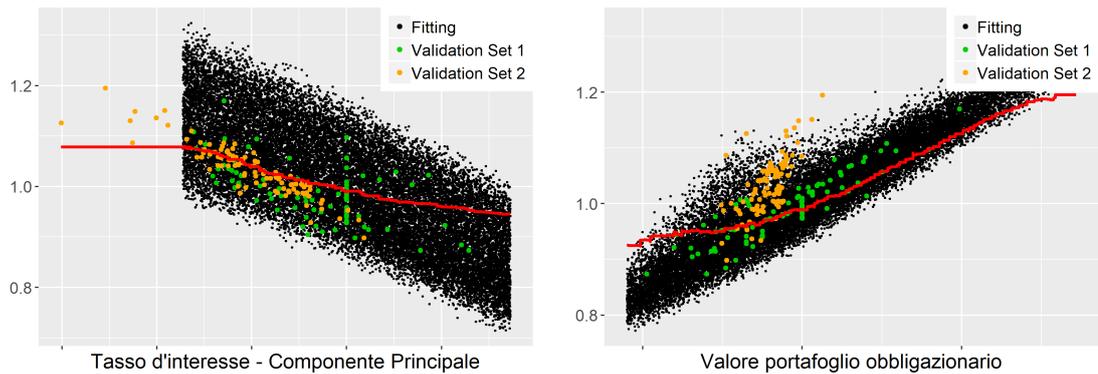
Le metriche out-of-sample, pur leggermente instabili, spaziano in un range di valori verosimilmente ritenuti accettabili. Qualora questa componente di aleatorietà sia ritenuta eccessiva, si potrebbe utilizzare il modello con l'intero dataset, che fornisce risultati stabili e un errore in linea con il modello stocastico. Nel seguito ripropo-

niamo l'analisi degli effetti marginali del modello. Come nel caso polinomiale, ci limitiamo ad alcune LoB e alle variabili più significative.

Compagnia 3 - Protezione



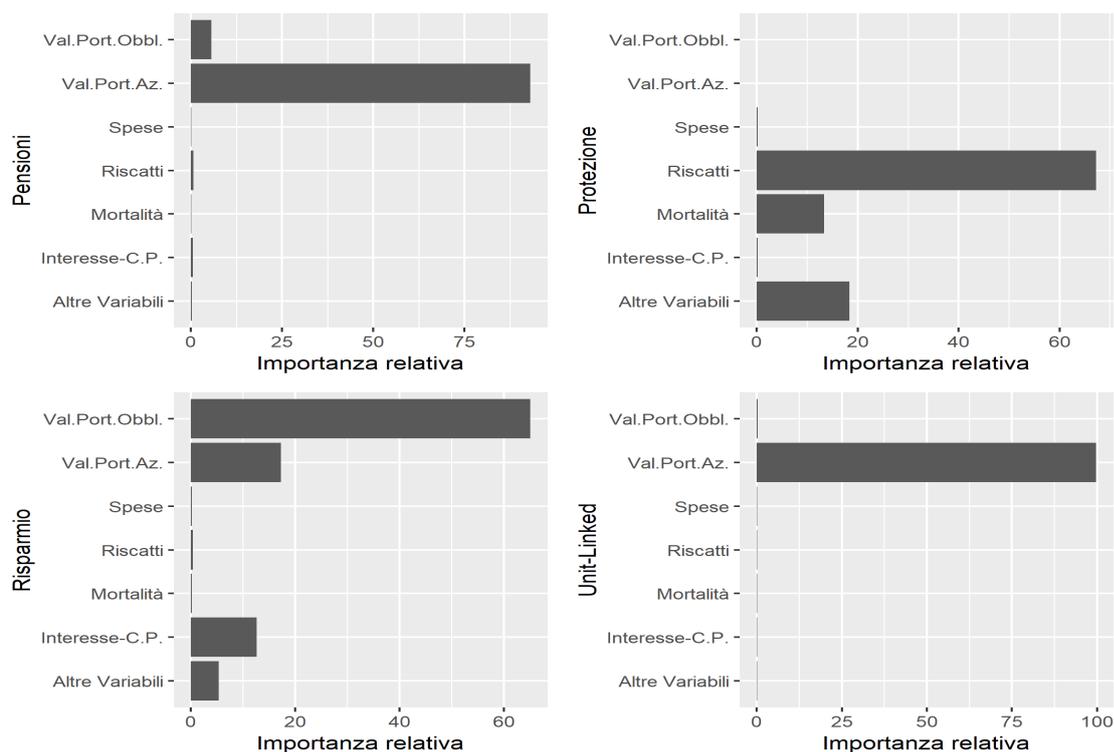
Compagnia 3 - Risparmio



Andamento marginale della proxy costruita per il valore della BEL, rispetto ad alcuni dei fattori di rischio considerati; punti neri rappresentano gli scenari utilizzati per calibrare il modello, i punti verdi e arancioni quelli su cui sono calcolate le metriche di validazione.

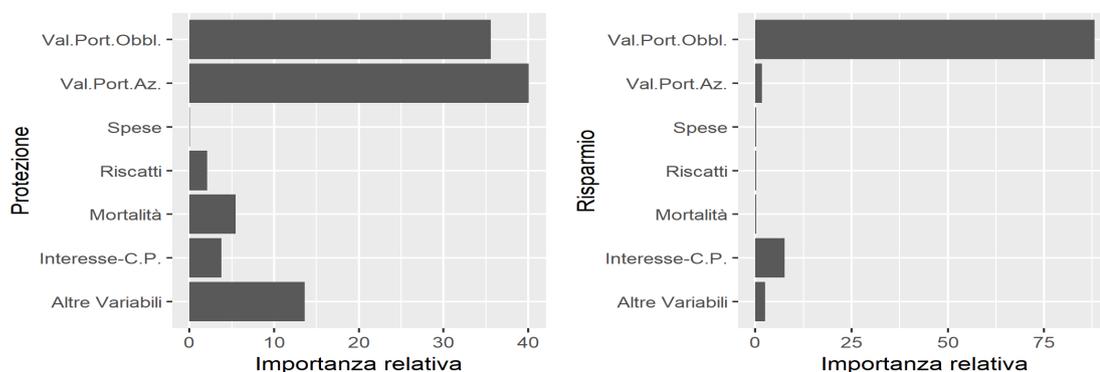
Gli andamenti nello spazio di fitting risultano molto simili a quelli della funzione polinomiale, seppure più irregolari in quanto frutto di una funzione costante a tratti. Al di fuori di tale spazio, invece, il valore in questo modello è necessariamente preso costante come negli altri modelli ad albero, che pertanto non sono adatti a risolvere problemi di estrapolazione. In ogni caso, in questo modello, fortemente basato sulle interazioni, l'analisi dei soli effetti marginali risulta meno significativa. Infine, possiamo anche considerare la misura di importanza relativa delle diverse variabili esplicative, come definita nella sezione 3.1.3. Esse, raffigurate nella figura seguente, confermano l'intuizione derivante dall'analisi grafica delle marginali: massima rilevanza è rivestita dall'andamento degli attivi, mentre i rischi biometrici risultano di importanza solitamente molto ridotta. Rispetto all'analisi delle marginali, tuttavia, la metrica d'importanza tiene conto, in maniera implicita, anche degli effetti di interazione, ed è quindi da considerarsi più completa.

Compagnia 1



Importanza relativa di alcune variabili inserite nel modello per le LoB della Compagnia 1.

Compagnia 3



Importanza relativa di alcune variabili inserite nel modello per le LoB della Compagnia 3.

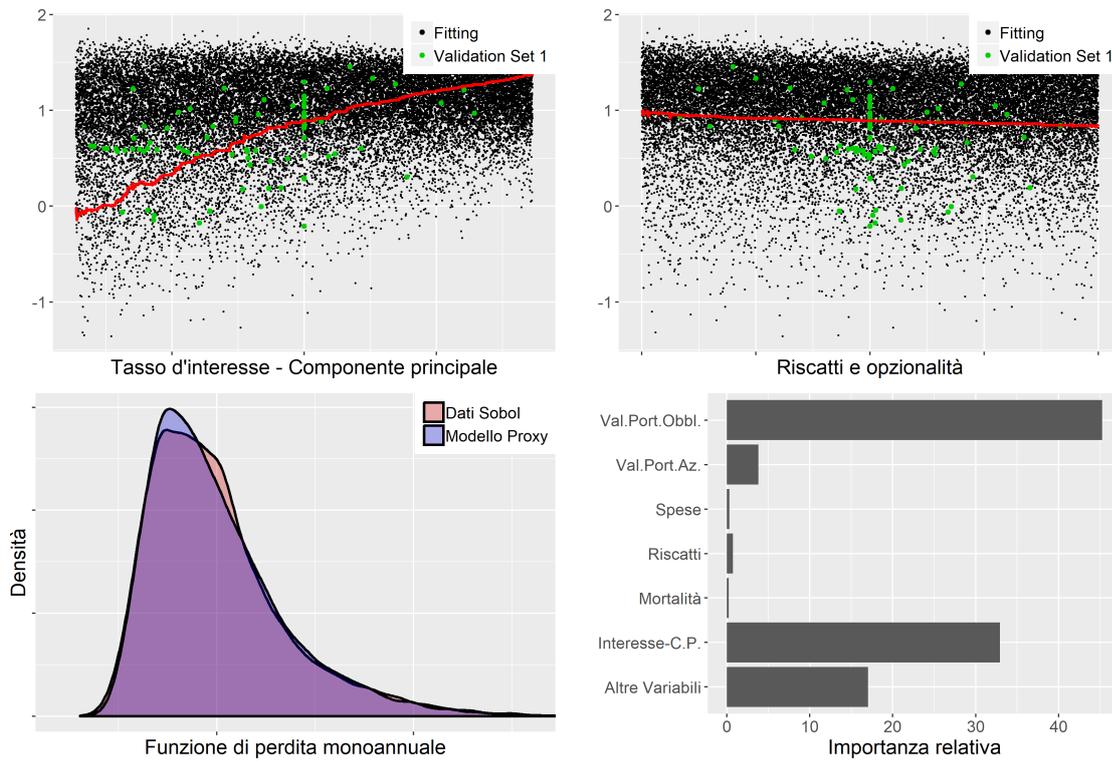
Ricalcando quanto effettuato nella sezione 2.3.3, procediamo ora con un'applicazione diretta al PVFP della Compagnia 3, in modo da ottenere una distribuzione simulata della perdita e un valore per l'SCR. Le forti ipotesi sottostanti a questa applicazione, nonché alcuni dettagli sugli scenari real-world utilizzati per stimare la perdita sono sempre rinvenibili nella sezione 2.3.3.

Nella Tabella 3.7 sono rappresentati i risultati del modello. Le impostazioni utilizzate per il tuning coincidono con quelle utilizzate per le BEL.

Modello per il PVFP						
	T	n	i	o	λ	\mathcal{M}_3
Compagnia 3	44m25s	5000	5	10	0.05	2.60%

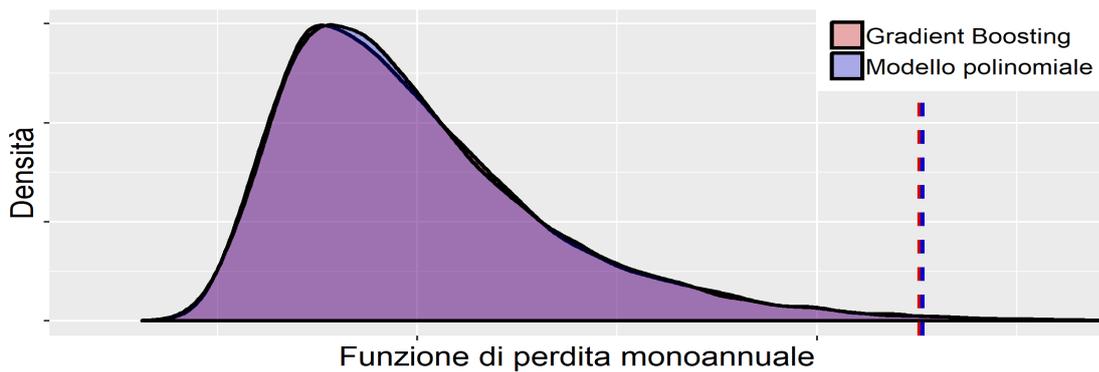
Tabella 3.7: *Parametri e bontà di previsione per il PVFP della Compagnia 3.*

Si nota che in questo contesto, in cui vi è maggiore instabilità dei dati e un andamento meno netto rispetto ai fattori di rischio, la flessibilità del gradient boosting consente di ottenere un modello dalle capacità predittive apparentemente migliori rispetto alla regressione polinomiale. Nella figura seguente riportiamo gli andamenti di alcune curve marginali del modello, l'importanza relativa dei fattori di rischio e il confronto tra perdita stimata e empirica sullo spazio di fitting.



Andamenti marginali della proxy, distribuzione della perdita e importanza relativa di alcuni fattori di rischio per la compagnia 3.

In ultimo, nella figura seguente compariamo la perdita sullo spazio di fitting ottenuta con il modello proxy basato sul gradient boosting con quella ricavata nella sezione 2.3.3 con il modello polinomiale; sia la forma della distribuzione che il quantile estremo sono quasi identici.



Comparazione della distribuzione empirica della perdita sullo spazio di fitting per i due modelli; le due linee tratteggiate rappresentano i rispettivi VaR al 99.5%.

Considerazioni finali

Nel corso di questa tesi ci siamo posti nel contesto del calcolo dei requisiti di capitale per un'impresa di assicurazione vita e abbiamo analizzato le diverse metodologie proposte per ottenere la distribuzione della funzione di perdita e il Value-at-Risk ad essa associato. Ci siamo soffermati sui metodi proxy, principalmente presentati nella sezione 1.3, che mirano ad approssimare il risultato ottenibile tramite un approccio pienamente stocastico al calcolo dell'SCR - ampiamente delineato nella sezione 1.2 - la cui applicazione diretta non è attuabile nella pratica per motivi computazionali.

Nel Capitolo 2 abbiamo esaminato il metodo Least-Squares Monte Carlo, il più recente e innovativo tra i modelli proxy, e abbiamo fornito gli strumenti matematico-statistici necessari ad impostare formalmente il modello e ricavare le sue principali proprietà, nonché alcuni suggerimenti per una sua implementazione pratica. Abbiamo poi posto l'attenzione sulla struttura di regressione polinomiale alla base della tecnica LSMC, analizzando alcune sue possibili estensioni; tuttavia, le difficoltà nella loro applicazione hanno finora fatto propendere, nelle implementazioni pratiche del modello, a mantenere la struttura polinomiale. I risultati ottenuti in questo contesto nella letteratura degli ultimi anni sono presentati nella sezione 2.2. Infine, nella sezione 2.3 abbiamo implementato il modello LSMC su tre compagnie di assicurazione vita, studiando in particolare la bontà di previsione del modello risultante in diverse configurazioni dell'algoritmo di determinazione della proxy, verificando empiricamente l'ottimalità dell'approccio basato sulle variabili antitetiche e analizzando la convergenza delle metriche considerate per misurare la bontà di previsione del modello. L'algoritmo utilizzato per la costruzione della funzione proxy, sviluppato indipendentemente in linguaggio R e presentato nell'Appendice

B di questa tesi, consente di attuare in via automatica l'intera procedura di selezione dei regressori e di stima del modello mediante un procedimento forward basato sui criteri informativi. Sul dataset da noi considerato, contenente 25'000 scenari, l'algoritmo sviluppato solitamente richiede un tempo totale di esecuzione che spazia dai pochi minuti ad un'ora per ogni funzione proxy. Si potrebbe cercare di ridurre ulteriormente il tempo di esecuzione tramite meccanismi di calcolo parallelo, che potrebbero condurre ad una significativa diminuzione dei tempi di calcolo, in particolare se l'esecuzione avviene su un server dedicato.

Nel Capitolo 3, infine, abbiamo proposto un'implementazione delle tecniche di machine learning tree-based nel contesto dei modelli proxy, come alternativa al metodo LSMC. L'obiettivo era la ricerca di un modello che possa svincolarsi dalle restrizioni insite nel modello di regressione polinomiale, consenta in maniera oggettiva di effettuare la selezione delle variabili esplicative e degli effetti di interazione rilevanti e, idealmente, porti a risultati migliori in termini di capacità di previsione, restando, allo stesso tempo, maggiormente interpretabile rispetto a tecniche "black-box" quali le reti neurali. Tuttavia, utilizzando i modelli di machine learning più tradizionali - quali i semplici alberi di regressione o le foreste casuali - i risultati ottenuti in previsione si sono rivelati peggiori di quelli del modello polinomiale; solo con le più avanzate tecniche di gradient boosting abbiamo ottenuto capacità previsive comparabili al più tradizionale modello LSMC, che, quantomeno sui dati da noi considerati, risulta essere molto adeguato per il problema in esame. Anche il modello di gradient boosting non ha mostrato un miglioramento globale decisivo rispetto al modello LSMC, pur ottenendo risultati più buoni su alcuni portafogli. Investigare ulteriormente sui modelli proposti in questa tesi, analizzandone ad esempio la performance su dataset diversi, con differenze nel budget computazionale, nella sua allocazione tra scenari interni ed esterni, nei modelli attuariali utilizzati per la proiezione dei flussi di cassa, potrebbe rivelarsi fruttuoso: si consideri, infatti, che le analisi sull'allocazione ottimale del budget sono sempre state effettuate solo guardando al modello polinomiale, e non è detto che la configurazione ottima sia la stessa anche per il modello alternativo. Inoltre, metodi più fini di tuning degli iperparametri del modello di gradient boosting potrebbero ridurre ancora l'errore commesso. Un discreto vantaggio dei modelli di machine learning,

infine, è dato dal fatto che vi sono già loro implementazioni efficienti e affidabili nei principali software di calcolo statistico, mentre l'algoritmo di selezione dei termini polinomiali va programmato manualmente, seguendo, ad esempio, la strada esposta nell'Appendice B di questa tesi.

Sebbene nella nostra applicazione non abbia avuto pieno successo il tentativo di migliorare il modello LSMC tramite l'utilizzo degli algoritmi di machine learning, le intuizioni alla loro base - come ad esempio i metodi di valutazione dell'errore, il trattamento degli iperparametri, i procedimenti di tuning - potrebbero essere utilizzati per migliorare l'algoritmo di selezione dei termini in un modello polinomiale. Ad esempio, si potrebbero considerare il grado massimo dei termini e delle interazioni come iperparametri da sottoporre ad un processo di tuning per ogni portafoglio, in modo da permettere ad un algoritmo automatico di selezionare valori in grado di riflettere meglio le specificità dei portafogli. Naturalmente, visto il maggiore onere computazionale della selezione polinomiale rispetto al gradient boosting, algoritmi particolarmente pesanti quali la ricerca a griglia non sarebbero computazionalmente applicabili, e si renderebbero necessari metodi di tuning più efficienti.

In ultimo, visto il discreto successo ottenuto dalla tecnica automatica di costruzione della struttura di regressione polinomiale nel nostro contesto, e visto il significativo potenziale di applicazione ad altri ambiti, ci proponiamo, come spunto per le nostre future ricerche, di estendere e generalizzare l'algoritmo sviluppato nell'Appendice B, aggiungendo funzionalità quali la possibilità di effettuare un procedimento di selezione stepwise o di considerare criteri alternativi di arresto, e velocizzando ulteriormente i tempi di esecuzione, in modo da fornire un valido strumento generale - integrato in un apposito pacchetto - per lo sviluppo automatico di modelli di regressione polinomiale sulla piattaforma R.

Appendice A

Sulla valutazione neutrale al rischio

La tecnica di valutazione neutrale al rischio è un metodo per la valutazione di derivati sviluppato da J. Cox e S. Ross nel 1976. L'idea è quella di definire una misura di probabilità - detta neutrale al rischio - sotto la quale il prezzo di non-arbitraggio di un'attività è pari al valore atteso del payoff, attualizzato tramite il tasso risk-free. Sotto la probabilità neutrale al rischio, tutte le attività finanziarie dell'economia hanno tasso di rendimento pari al tasso risk-free (quindi non vi è alcuna remunerazione per il rischio). La misura di probabilità neutrale al rischio è la misura di probabilità adottata da un individuo neutrale al rischio, ossia che sarà sempre indifferente tra una quantità aleatoria e il suo valore atteso.

La presente trattazione segue principalmente Bingham e Kiesel (2004) e mira, nell'ambito della matematica finanziaria a tempo continuo, a fornire in maniera sintetica un percorso formale volto a ricavare la formula della valutazione neutrale al rischio, che abbiamo utilizzato ampiamente nella trattazione dell'approccio stocastico alla valutazione del capitale di solvibilità.

Consideriamo un mercato finanziario aperto nel continuo in un intervallo $[0, T]$ e rappresentiamo l'incertezza tramite lo spazio di probabilità filtrato $(\Omega, \mathcal{A}, P, \mathcal{F})$ ove la filtrazione $\mathcal{F} = (\mathcal{F}_t)_{0 \leq t < T}$ con $\mathcal{F}_t \in \mathcal{A} \forall t$ è completa ($\forall t$ \mathcal{F}_t contiene gli eventi di probabilità nulla di \mathcal{A}) e continua a destra ($\forall t$ $\mathcal{F}_t = \bigcap_{u > t} \mathcal{F}_u$).

Interpretativamente, Ω rappresenta lo spazio degli stati possibili del mercato, \mathcal{A} è una σ -algebra di eventi di Ω , P una misura realistica di probabilità, \mathcal{F}_t la σ -algebra

degli eventi osservabili in t (ossia l'informazione disponibile in t).

Il mercato sia competitivo (agenti price-takers) e privo di frizionalità (assenza di tassazione, di costi di transazione, di vincoli alla vendita allo scoperto; perfetta divisibilità dei titoli).

Supponiamo che sul mercato siano trattate $d + 1$ attività, i cui prezzi sono definiti dai processi stocastici $S_0(t), \dots, S_d(t)$. Identificheremo per comodità le attività con i processi stocastici dei prezzi loro associati. Supponiamo che il processo dei prezzi $\mathbf{S} = (S_0, \dots, S_d)$ sia una semi-martingala continua, adattata (cioè tale che $\forall t$ $\mathbf{S}(t)$ sia \mathcal{F}_t -misurabile) e strettamente positiva sullo spazio $(\Omega, \mathcal{A}, P, \mathcal{F})$ e che $S_0(t)$ sia un'attività che non paga dividendi quasi certamente strettamente positiva (quindi utilizzabile come numerario). Possiamo allora considerare anche il processo dei prezzi scontati rispetto al numerario $S_0(t)$, $\tilde{\mathbf{S}}(t) = \mathbf{S}(t)/S_0(t)$.

Chiamiamo portafoglio o strategia d'investimento ogni processo a valori in \mathbb{R}^{d+1} $\varphi(t) = (\varphi_0(t), \dots, \varphi_d(t))$, prevedibile e localmente limitato, che soddisfi alle condizioni $\int_0^T E(\varphi_0(t))dt < +\infty$, $\sum_{i=0}^d \int_0^T E(\varphi_i^2(t))dt < +\infty$ necessarie per l'esistenza dell'integrale stocastico $\int_0^t \varphi(u) d\mathbf{S}(u) < +\infty$. Interpretativamente, $\varphi_i(t)$ rappresenta il numero di quote dell'attività i detenute all'epoca t e determinate dopo aver osservato i prezzi $\mathbf{S}(t-)$. Per l'assenza di vincoli sulle vendite allo scoperto e per la perfetta divisibilità dei titoli si ha $\varphi_i(t) \in \mathbb{R}$.

Definiamo allora il processo del valore della strategia:

$$V_\varphi(t) = \varphi(t) \mathbf{S}(t) = \sum_{i=0}^d \varphi_i(t) S_i(t) \quad t \in [0, T] \quad (\text{A.1})$$

e il processo del guadagno della strategia:

$$G_\varphi(t) = \int_0^t \varphi(u) d\mathbf{S}(u) = \sum_{i=0}^d \int_0^t \varphi_i(u) dS_i(u) \quad t \in [0, T] \quad (\text{A.2})$$

Analogamente possiamo introdurre il processo relativo al valore scontato $\tilde{V}_\varphi(t) = V_\varphi(t)/S_0(t)$ e il processo del guadagno scontato $\tilde{G}_\varphi(t) = G_\varphi(t)/S_0(t)$.

Definiamo infine autofinanziante una strategia φ se vale la relazione $V_\varphi(t) = V_\varphi(0) + G_\varphi(t) \forall t \in [0, T]$.

Possiamo allora enunciare il seguente teorema, per la cui dimostrazione si rimanda a Bingham & Kiesel (2004):

Teorema di invarianza rispetto al numerario

Una strategia d'investimento autofinanziante rimane autofinanziante anche in seguito ad un cambio di numerario.

Corollario

Una strategia d'investimento φ è autofinanziante se e solo se $\tilde{V}_\varphi(t) = \tilde{V}_\varphi(0) + \tilde{G}_\varphi(t) \forall t \in [0, T]$.

Tale risultato mostra che una strategia autofinanziante è determinata univocamente a partire dal suo valore iniziale e dalle componenti $\varphi_1(t), \dots, \varphi_d(t)$, ossia a partire da ogni insieme di processi prevedibili $\varphi_1(t), \dots, \varphi_d(t)$ che garantiscono l'esistenza degli integrali stocastici $\int_0^t \varphi_i d\tilde{S}_i$ $i = 1, \dots, d$, fissato arbitrariamente $\tilde{V}_\varphi(t) = v$ si può costruire una strategia autofinanziante ponendo:

$$\varphi_0(t) = v + \sum_{i=1}^d \int_0^t \varphi_i(u) d\tilde{S}_i(u) - \sum_{i=1}^d \varphi_i(t)\tilde{S}_i(t) \quad \forall t \in [0, T]$$

Definiamo ora formalmente l'opportunità di arbitraggio come una strategia autofinanziante φ per cui vale $V_\varphi(0) = 0$, $P(V_\varphi(T) \geq 0) = 1$ e $P(V_\varphi(T) > 0) > 0$.

Introduciamo quindi una misura neutrale al rischio (o martingala equivalente) Q , ossia una misura di probabilità definita su (Ω, \mathcal{A}) tale per cui Q è equivalente a P e il processo dei prezzi $\tilde{S}_k(t)$ è una martingala¹ rispetto a Q .

¹Ricordiamo che un processo stocastico $M = \{M_t\}_{0 \leq t < +\infty}$ è una martingala rispetto a Q se:

- M è adattato a \mathcal{F} , ossia $M(t)$ è \mathcal{F}_t -misurabile $\forall t$.
- $\mathbb{E}(|M(t)|) < +\infty \quad \forall t \in [0, +\infty[$
- $\mathbb{E}(M(t) | \mathcal{F}_s) = M(s) \quad Q - q.c. \quad \forall (s, t) : 0 \leq s \leq t$

Più in generale, possiamo considerare il processo M come martingala locale rispetto a Q se esiste una successione $\tau_n : \Omega \rightarrow [0, +\infty[$ di tempi d'arresto (ossia variabili aleatorie per cui $\{\tau \leq t\} \in \mathcal{F}_t \forall t \geq 0$) tale che:

- τ_n è P -q.c. crescente, ossia $P(\tau_n \leq \tau_{n-1}) = 1$, e la successione è P -q.c. divergente.
- il processo arrestato $M_{t \wedge \tau_n}$ è una martingala per ogni n .

Il seguente teorema fornisce una condizione necessaria e sufficiente affinché una misura equivalente a P sia neutrale al rischio: sotto una misura neutrale al rischio, infatti, il processo del prezzo di una qualunque attività trattata sul mercato deve evolvere con drift pari al tasso d'interesse risk-free.

Teorema di caratterizzazione delle misure neutrali al rischio

Assunto $S_0(t) = e^{r(t)}$, una misura Q equivalente a P è una misura neutrale al rischio se e solo se rispetto ad essa, ogni attività i trattata sul mercato ha processo del prezzo della forma $dS_i(t) = r(t)S_i(t)dt + dM_i(t)$ ove M_i è una martingala locale rispetto a Q .

Data una strategia autofinanziante φ , diremo che essa è mansueta se vale che $\tilde{V}_\varphi(t) \geq 0 \forall t \in [0, +\infty[$. Inoltre diremo che essa è Q -ammissibile se il processo del guadagno scontato $\tilde{Q}_\varphi(t)$ è una Q -martingala. Indicheremo con Φ l'insieme delle strategie autofinanzianti mansuete e con $\Phi(Q)$ l'insieme delle strategie autofinanzianti Q -ammissibili. Infine, chiameremo ammissibile ogni strategia autofinanziante mansueta e Q -ammissibile.

Il seguente teorema mette in relazione l'esistenza di misure neutrali al rischio con l'esistenza di opportunità di arbitraggio sul mercato.

Teorema

Esista una misura neutrale al rischio. Allora il mercato finanziario non contiene opportunità di arbitraggio in Φ e in $\Phi(Q)$.

Abbiamo visto quindi che l'esistenza della misura neutrale al rischio implica l'assenza di opportunità di arbitraggio tra le strategie mansuete e Q -ammissibili. Si vorrebbe ora dimostrare il viceversa, in modo da ottenere nel contesto della matematica finanziaria a tempo continuo un analogo del teorema fondamentale dell'asset pricing in tempo discreto. Tuttavia, ciò non è possibile senza una ridefinizione più stringente del concetto di opportunità di arbitraggio e ulteriori condizioni sui portafogli da considerare. Per ulteriori dettagli e una derivazione del teorema fondamentale dell'asset pricing in tempo continuo si veda Bingham & Kiesel (2004).

Definiamo contingent claim ogni variabile aleatoria non-negativa X su (Ω, \mathcal{A}, P) . Interpretativamente, si può vedere un contingent claim come un contratto che

paga $X(\omega)$ in T se si verifica lo stato del mondo ω . Un contingent claim è detto replicabile se esiste almeno una strategia ammissibile φ , detta strategia replicante, tale che $V_\varphi(T) = X$. Un mercato finanziario è detto completo se ogni contingent claim è replicabile. Osserviamo che se un contingent claim è replicabile, allora il suo prezzo di non-arbitraggio in t sarà $\Pi_X(t) = V_\varphi(t)$.

Assumeremo d'ora in poi che esista una misura neutrale al rischio Q e limitiamo la nostra analisi ai contingent claim X integrabili sotto Q (ossia tali per cui vale $\mathbb{E}_Q(X/S_0(T)) < +\infty$).

Teorema (Formula di valutazione neutrale al rischio)

Esista una misura neutrale al rischio Q e sia X un contingent claim replicabile. Allora il processo del prezzo di non-arbitraggio di X è dato da:

$$\Pi_X(t) = S_0(t) \mathbb{E}_Q \left[\frac{X}{S_0(T)} \middle| \mathcal{F}_t \right] \tag{A.3}$$

Dimostrazione. Essendo X replicabile, esiste una strategia ammissibile φ tale che $V_\varphi(T) = X$ e $\Pi_X(t) = V_\varphi(t)$. Dato che $\varphi \in \Phi(Q)$, il processo del valore scontato $\tilde{V}_\varphi(t)$ è una Q -martingala, e quindi vale:

$$\Pi_X(t) = V_\varphi(t) = S_0(t) \tilde{V}_\varphi(t) = S_0(t) \mathbb{E}_Q \left[\tilde{V}_\varphi(T) \middle| \mathcal{F}_t \right] = S_0(t) \mathbb{E}_Q \left[\frac{V_\varphi(T)}{S_0(T)} \middle| \mathcal{F}_t \right]$$

Da cui, essendo $V_\varphi(T) = X$, si ottiene proprio la formula (A.3). □

Corollario

Dato un mercato finanziario completo con strategie di negoziazione ammissibili (mansuete e Q -ammissibili per qualche misura neutrale al rischio Q), per ogni contingent claim replicabile X il prezzo di non-arbitraggio $\Pi_X(t)$ esiste ed è unico.

Dimostrazione. Fissato X , consideriamo le strategie replicanti $\varphi_1 \in \Phi(Q_1)$ e $\varphi_2 \in \Phi(Q_2)$. Essendo $V_{\varphi_1}(t) = V_{\varphi_2}(t)$ si ha $\Pi_X^{Q_1}(t) = V_{\varphi_1}(t) = V_{\varphi_2}(t) = \Pi_X^{Q_2}(t)$, ove l'uguaglianza centrale deriva dall'assenza di opportunità di arbitraggio tra le strategie ammissibili, le altre dalla definizione di replicabilità. □

Per conseguenza di quest'ultimo risultato, in particolare, il prezzo di non-arbitraggio non dipende dalla scelta della misura neutrale al rischio Q .

Appendice B

Funzioni R per l'implementazione efficiente del modello LSMC

In relazione al modello LSMC applicato nella sezione 2.3, forniamo alcune funzioni utilizzabili per implementare la procedura di calibrazione della funzione proxy tramite il software statistico open-source R. Seppure ove possibile si consiglia l'utilizzo delle funzioni di calcolo già presenti nel programma o in pacchetti ufficialmente supportati, si è sentita l'esigenza di proporre funzioni nuove, in modo da:

- Velocizzare il tempo necessario per l'esecuzione, costruendo funzioni specifiche che effettuino soltanto le operazioni strettamente necessarie per il modello, mentre le funzioni ufficiali, pur portando allo stesso risultato, richiedono un tempo di esecuzione significativamente maggiore.
- Costruire in maniera automatica la funzione proxy tramite la regressione polinomiale e gli algoritmi di selezione visti nella sezione 2.1.4. Infatti per tali procedimenti, generalmente poco utilizzati nella letteratura statistica, non sono presenti funzioni adeguate all'interno dei pacchetti pubblicamente disponibili.

Insieme alle funzioni, che commenteremo debitamente e confronteremo, qualora possibile, con quelle utilizzate nella versione base di R, forniremo la documentazione in package-style, in modo che esse possano venire implementate, senza ulteriori accorgimenti, all'interno di un pacchetto.

a) Una funzione per il calcolo efficiente dei criteri informativi

La prima funzione che proponiamo calcola le principali statistiche utili per le nostre applicazioni, ossia R^2 , R^2 corretto, AIC e BIC (per i dettagli tecnici sulla loro costruzione si veda la sezione 2.1.4), a partire da un vettore rappresentante le osservazioni della variabile dipendente e da una matrice contenente i valori osservati dei regressori del modello considerato.

Le statistiche considerate possono essere agevolmente calcolate in R tramite funzioni integrate nel programma a partire da un oggetto di tipo lm . Tuttavia, la costruzione di un oggetto lm è computazionalmente impegnativa, e, considerato che nel nostro modello per ottenere una singola funzione proxy dovremo effettuare migliaia di confronti tra modelli, risulta significativo il risparmio ottenuto utilizzando tecniche computazionalmente più leggere.

La strada seguita, in particolare, si basa sulla funzione integrata $lm.fit()$ che automaticamente fornisce la stima dei minimi quadrati del vettore dei coefficienti tramite la tecnica della decomposizione QR presentata nella sezione 2.1.1, oltre che i residui del modello, pur essendo significativamente meno impegnativa computazionalmente delle funzione $lm()$. Sulla base di queste informazioni, la funzione proposta calcola manualmente i criteri informativi. Osserviamo che le formulazioni utilizzate per il calcolo di AIC e BIC sono una tra le possibili varianti, scelta per garantire la consistenza con le formule integrate in R .

```
#' A function that computes the aic, bic, R squared and
#' adjusted R squared of a linear regression model
#'
#' @param y vector; contains the observations of the response variable
#' @param x matrix; contains the observations of the regressors
#'
#' @return AIC, BIC, R^2 and Adjusted R^2 of the linear model y ~ x
#' @export
#'
compute.information <- function(y, x) {
  # number of observations
  n <- length(y)
  # number of regressors
  k <- ncol(x)
  # lm.fit computes only the coefficients and the residuals
  # and not the whole object created by "lm"
  fit <- lm.fit(x, y)
  beta <- fit$coefficients
```

```

res <- fit$residuals
s <- y - mean(y)
# residual deviance
ssr <- as.numeric(t(res) %*% res)
# total deviance
sst <- as.numeric(t(s) %*% s)
rsquared <- 1 - ssr / sst
aic <- 2 * k + n * (log(ssr / n) + log(2 * pi) + 1)
bic <- log(n) * k + n * (log(ssr / n) + log(2 * pi) + 1)
adjrsquared <- rsquared - (1 - rsquared) * (k - 1) / (n - k)
results <- list(
  "aic" = aic,
  "rsq" = rsquared,
  "arsq" = adjrsquared,
  "bic" = bic)
return(results)
}

```

Osserviamo che potremmo calcolare unicamente l'indicatore relativo al criterio effettivamente utilizzato per la selezione delle variabili, anziché tutti e quattro i valori. La differenza computazionale tuttavia è minima e la possibilità di verificare ex-post l'andamento di altri indicatori nel processo di selezione può essere rilevante per determinare eventuali differenze qualora si valutassero criteri alternativi.

Per visualizzare il risparmio ottenuto in termini di tempo di esecuzione, abbiamo effettuato un confronto per diversi numeri di regressori su un recente personal computer di fascia media:

Tempo medio di esecuzione	25 regressori	75 regressori	150 regressori
compute.information()	26 ms	150 ms	598 ms
algoritmo integrato in R	43 ms	191 ms	721 ms

Confronto tra tempi medi di esecuzione per il calcolo dei criteri di informazione di un modello di regressione al variare del numero di regressori, fissato il numero di osservazioni a 25'000.

Considerato che tale operazione viene effettuata migliaia di volte durante la calibrazione, il risparmio complessivo è, in media, di 5-15 minuti per ogni funzione proxy. Utilizzando una granularità a livello di portafoglio, prendendo ad esempio una grande compagnia o addirittura un gruppo assicurativo, in tal modo si possono risparmiare diverse ore per un giro completo del modello.

b) Una funzione per la determinazione dei termini polinomiali ammissibili secondo un algoritmo adaptive

Gli algoritmi di selezione dei regressori presentati nella sezione 2.1.4 nel contesto della regressione polinomiale non sono implementati né nella versione base di R né nei pacchetti pubblicamente disponibili. Ne segue che per costruire un modello Least-Squares Monte Carlo nella sua forma tradizionale, gli algoritmi di selezione dovranno essere programmati manualmente.

Nel seguito proponiamo una funzione che ad ogni passo di un procedimento forward, a partire dal principio statistico di marginalità e da limitazioni imposte in input alla forma funzionale della proxy, determina quali nuovi termini polinomiali risultano ammissibili secondo un algoritmo di tipo adaptive, come presentato nella sezione 2.1.4.

Al di là delle imposizioni manuali quali la potenza massima dei termini principali e di interazione, l'algoritmo proposto si basa sull'osservazione che ogni possibile nuovo termine da testare ad un dato passo della procedura di selezione, se non è un termine già ammissibile negli step precedenti (ma non ancora scelto), allora necessariamente il vettore indicante le potenze dei diversi risk factor del nuovo termine è ottenibile come somma dei vettori delle potenze del termine selezionato durante l'ultima iterazione e di quello di un qualsiasi altro termine precedentemente selezionato. In questo modo abbiamo ottenuto un sovrainsieme dei nuovi termini ammissibili. Per selezionare i nuovi termini ammissibili a partire da quest'ultimo, si andrà a verificare, termine per termine, se il principio statistico di marginalità risulta soddisfatto. Solo se tutte le condizioni implicate dal principio di marginalità sono soddisfatte, il termine sarà aggiunto alla lista dei termini ammissibili per l'iterazione successiva dell'algoritmo di selezione.

```
#' A function that updates the list of possible candidates
#' after each iteration of the stepwise selection algorithm
#'
#' @param candidate matrix; the matrix to be updated
#' @param chosen numeric; the number of the row in candidate
#' which was selected during the last iteration
#' @param chosen_matrix matrix; the matrix containing
#' the terms chosen in previous iterations
#' @param maxPower_SingleTerms numeric; highest power
#' reachable by a single regressor
```

```

#' @param maxPower_CrossTerms numeric; highest power
#' reachable by a cross term
#' @param maxPower_CrossTermsTerms numeric; highest power
#' reachable by a single regressor in a cross term
#'
#' @return the candidate matrix for the next iteration
#' @export
#'

update.candidate <- function(candidate,
                              chosen,
                              chosen_matrix,
                              maxPower_SingleTerms,
                              maxPower_CrossTerms,
                              maxPower_CrossTermsTerms) {
  # Extracts the chosen vector from the matrix
  chosen_vector <- candidate[chosen,]
  # For each new possible candidate, check if all conditions are met
  # If yes it is added to the matrix containing the old candidates
  for (i in 1:nrow(chosen_matrix)) {
    potential <- chosen_vector + chosen_matrix[i,]
    zero <- matrix(0, nrow = 1, ncol = length(potential))
    # Checks that the term hasn't already been chosen
    if (!(any(apply(chosen_matrix, 1, function(x, want)
                   isTRUE(all.equal(x, want)), potential)))) {
      # Checks that the term is admissible
      # according to the chosen parameters
      if (max(potential) <= maxPower_SingleTerms) {
        # count values different from 0 in the potential candidate
        # to see if it is a single term or a cross term
        non_zeros <- sum(potential != 0)
        # if non_zeros <= 1 it is a single term
        # therefore we do not check cross term conditions
        if (non_zeros <= 1 | ( non_zeros > 1 &
                               max(potential) <= maxPower_CrossTermsTerms &
                               sum(potential) <= maxPower_CrossTerms))
          {
            # Builds a matrix containing all terms
            # that have to be already chosen
            # for the new term to be added
            # (if the term is X^3, it would be X^1 and X^2)
            potential_matrix <- matrix(potential,
                                       nrow = 1,
                                       ncol = length(potential))
            val <- list()
            # Creates all possible powers for each risk factor
            # there is one vector for each risk factor
            for (j in 1:length(potential)) {
              val[[j]] <- seq(from = 0, to = potential[j])
            }
          }
        }
    }
  }
}

```

```

}
# Builds a matrix with zero in the first row
# the potential new candidate in the second row
# followed by all possible combinations of powers
potential_matrix <- rbind(zero,
potential_matrix, as.matrix(expand.grid(val)))
# deletes duplicates and the first two rows
# therefore ensuring that the potential_matrix
# does neither include the intercept
#nor the potential new candidate
potential_matrix <- unique(potential_matrix)
potential_matrix <- potential_matrix[-c(1, 2),]
if (is.vector(potential_matrix)) {
  potential_matrix = t(matrix(
  potential_matrix,
  nrow = length(potential_matrix),
  ncol = 1))
}
# Checks if the terms considered have already been chosen
# aux is true if they have, false if they have not
# aux <- c()
# for (r in 1:nrow(potential_matrix)) {
  if (any(apply(chosen_matrix, 1, function(x, want)
  isTRUE(all.equal(x, want, check.attributes = FALSE)),
  potential_matrix[r,]))) {
    aux[r] <- TRUE
  } else{
    aux[r] <- FALSE
  }
}
# If all conditions are met the term is added to candidate
if (all(aux)) {
  candidate <- rbind(candidate, potential)
}
}
}
}
}
# Removes the chosen term from the candidates
candidate <- candidate[-chosen,]
# Returns the upgraded candidate matrix
return(candidate)
}

```

c) Una funzione per la costruzione del vettore di osservazioni per un termine polinomiale

La terza funzione proposta è meramente un tecnicismo con l'obiettivo di costruire il vettore di osservazioni relativo ad un termine polinomiale, a partire dalla matrice design dei fattori di rischio e dal vettore contenente le potenze del termine da costruire.

```
#' A function that extracts the observations of a polynomial
#' term from the risk factors design matrix
#'
#' @param selectedTerm character; the powers of the selected term
#' @param X matrix; the design matrix of the risk factors
#' @param maxPower numeric; the maximum power of the polynomial
#'
#' @return the observations of the polynomial term selected
#' @export
#'
extract.term <- function(selectedTerm, X, maxPower) {
  terms <- c()
  for (i in 1:maxPower) {
    factors <- which(selectedTerm == i)
    terms <- cbind(terms, X[, factors] ^ i)
  }
  computedTerm <- apply(terms, 1, prod)
  return(computedTerm)
}
```

Osserviamo che per un leggero risparmio computazionale potremmo utilizzare la funzione *rowProds()* del pacchetto *matrixStats* anziché la funzione base *apply()*. Tuttavia in questa sede abbiamo deciso di limitarci all'utilizzo di funzioni presenti nella versione base di R.

d) Una funzione per l'implementazione di un algoritmo forward di selezione dei termini polinomiali

In ultimo, forniamo la funzione che, sfruttando tutte le altre funzioni finora presentate, esegue l'intera procedura di selezione dei regressori, fornendo in output il numero di passi eseguito, il valore del criterio d'informazione, di R^2 e R^2 corretto ad ogni passo, la matrice design del modello finale, nonché il polinomio scritto come formula e in forma matriciale - ossia una matrice con numero di colonne pari al numero di regressori e numero di righe pari al numero di iterazioni

svolto dall'algoritmo, in cui ogni riga rappresenta le potenze selezionate durante quell'iterazione.

Una volta eseguito questo algoritmo a partire dalle osservazioni della variabile risposta e dei fattori di rischio, sarà sufficiente utilizzare la formula trovata in un modello lineare per ottenere la funzione proxy definitiva, da utilizzare poi per valutare in via semplificata il portafoglio in ogni scenario.

```
#' A function that performs the forward selection algorithm
#' in a polynomial regression setting
#'
#' @param YFit numeric; the observations of the response variable
#' @param XFit matrix; the observations of the possible regressors
#' @param informationCriterion character; the criterion used
#' to perform the selection
#' @param maxPower_SingleTerms numeric; highest power
#' reachable by a single regressor
#' @param maxPower_CrossTerms numeric; highest power
#' reachable by a cross term
#' @param maxPower_CrossTermsTerms numeric; highest power
#' reachable by a single regressor in a cross term
#' @param maxNumber_Terms numeric; maximum number of terms
#' in the polynomial. Default value of null means no limit
#'
#' @return a list with statistics and terms for each iteration,
#' the full chosen matrix and the number of iterations performed
#' @export
#'
select.regressors <- function(YFit,
                             XFit,
                             informationCriterion,
                             maxPower_SingleTerms,
                             maxPower_CrossTerms,
                             maxPower_CrossTermsTerms,
                             maxNumber_Terms = NULL) {
  # Builds up the polynomial with the intercept term
  # and computes the intercept's statistics
  candidate <- diag(c(rep(1, dim(XFit)[2])))
  colnames(candidate) <- colnames(XFit)
  ZeroModel <- lm(YFit ~ 1)
  if (informationCriterion == "AIC") {
    Information <- AIC(ZeroModel)+1
    Information <- rbind(Information, AIC(ZeroModel))
  } else{
    Information <- BIC(ZeroModel)+1
    Information <- rbind(Information, BIC(ZeroModel))
  }
  R_squared <- summary(ZeroModel)$r.squared
```

```

Adj_r_squared <- summary(ZeroModel)$adj.r.squared
chosen_matrix <- c()
Polynomial_terms <- c()
ChosenXF <- matrix(1, nrow = dim(XFit)[1], ncol = 1)
g <- 0
stop <- 0

# adaptive forward selection algorithm
while (Information[g + 2] < Information[g + 1] &
      (is.null(maxNumber_Terms) ||
       g + 1 <= maxNumber_Terms) & stop != 1) {
  aic <- c()
  bic <- c()
  temp_r <- c()
  temp_r_adj <- c()
  ChosenX <- list()
  # computes all candidate models
  for (l in 1:dim(candidate)[1]) {
    # extracts the potential new term in matricial view
    Selected_candidate <- candidate[l,]
    # extracts the value of the potential new term in each scenario
    maxPower <- max(maxPower_SingleTerms, maxPower_CrossTerms)
    DesignMatrix <- extract.term(Selected_candidate,
                                as.matrix(XFit), maxPower)
    # adds the vector above as a new column in the design matrix
    DesignMatrix <- cbind(ChosenXF, DesignMatrix)
    # saves the potential new predictors matrix in a list
    ChosenX[[l]] <- DesignMatrix
    # computes the statistics of this model
    results <- compute.information(YFit, DesignMatrix)
    aic <- rbind(aic, results$aic)
    temp_r <- rbind(temp_r, results$rsq)
    temp_r_adj <- rbind(temp_r_adj, results$arsq)
    bic <- rbind(bic, results$bic)
  }
  # chooses the best model according to the selected criterion
  if (informationCriterion == "AIC") {
    chosen <- which.min(aic)
    Information <- rbind(Information, aic[chosen])
  } else {
    chosen <- which.min(bic)
    Information <- rbind(Information, bic[chosen])
  }
  g <- g + 1
  # computes the new statistics and updates the candidate matrix
  # this is done only if the conditions of the while cycle
  # are still valid therefore ensuring that the element
  # after the last valid iteration is not taken
  if (Information[g + 2] < Information[g + 1] |

```

```

    length(Information) == 3) {
# Remembers the chosen predictors matrix
# previous candidates plus the candidate chosen now
ChosenXF <- ChosenX[[chosen]]
R_squared <- rbind(R_squared, temp_r[chosen])
Adj_r_squared <- rbind(Adj_r_squared, temp_r_adj[chosen])
# Indicates which predictor is chosen in the current iteration
polynomial_terms <- rbind(polynomial_terms, paste(paste(
  colnames(candidate)[which(candidate[chosen, ] != 0)],
  "^", candidate[chosen, which(candidate[chosen, ] != 0)]),
  collapse = "_"))
# Adds the candidate chosen in the current iteration
# to the matrix of previous chosen predictors
chosen_matrix <- rbind(chosen_matrix, candidate[chosen, ])
# update.candidate gives the candidates for the next iteration
candidate <- update.candidate(
  candidate,
  chosen,
  chosen_matrix,
  maxPower_SingleTerms,
  maxPower_CrossTerms,
  maxPower_CrossTermsTerms
)
# Conversion is necessary in the special case
# in which there is only one candidate
if (is.vector(candidate)) {
  candidate <- t(as.matrix(candidate))
}
# Prematurely exits the cycle if there are no more candidates
if (length(candidate) == 0) {
  stop <- 1
}
}
}
regressionResults <- list(
  "Information" = Information,
  "R_squared" = R_squared,
  "Adj_r_squared" = Adj_r_squared,
  "polynomial_terms" = polynomial_terms,
  "chosen_matrix" = chosen_matrix,
  "g" = g,
  "chosen_regressors_matrix" = ChosenXF
)
return(regressionResults)}

```

Bibliografia

- [1] Baek, S., Karaman, F., Ahn, H. (2005) *Variable Selection for Heteroscedastic Data Through Variance Estimation*, *Communication in Statistics-Simulation and Computation* 34(3), pp. 567-583
- [2] Bauer, D., Bergmann, D., Reuss, A. (2009) *Solvency II and Nested Simulations – a Least Squares Monte Carlo Approach*, Working Paper, Ulm University
- [3] Bauer, D., Reuss, A., Singer, D. (2012) *On the Calculation of the Solvency Capital Requirement based on Nested Simulations*, *ASTIN Bulletin* 42, pp. 453-499
- [4] Beutner, E., Pelsser, A., Schweizer, J. (2016) *Theory and Validation of Replicating Portfolios in Insurance Risk Management*, *SSRN Electronic Journal* 2557368
- [5] Bingham, N.H., Kiesel, R. (2004) *Risk-Neutral Valuation: Pricing and Hedging of Financial Derivatives*, 2nd Edition, Springer
- [6] Björck, Å. (1996) *Numerical Methods for Least Squares Problems*, Society for Industrial and Applied Mathematics
- [7] Born, R. (2018) *Künstliche Neuronale Netze im Risikomanagement*, Master Thesis, Universität zu Köln
- [8] Breiman, L. (2001) *Random Forests*, *Machine Learning*, Vol. 45, Issue 1, pp. 5-32
- [9] Cardi, G.G., Rusnak, R. (2007), *When the SST Standard Model Underestimates Market Risk*
- [10] Cathcart, M.J. (2012), *Monte Carlo Simulation Approaches to the Valuation and Risk Management of Unit-Linked Insurance Products with Guarantees*,

Ph.D. Thesis, School of Mathematical and Computer Sciences, Heriot-Watt University

- [11] CEIOPS (2010), *CEIOPS' Advice for Level 2 Implementing Measures on Solvency II: SCR standard formula - Article 111 j, k - Undertaking-specific parameters*
- [12] Chilla, F.M.W. (2014) *Economic Capital Forecasting of Interest Rate Risk in a Multi-Year Framework using Non-parametric Least Squares Monte Carlo for a With-Profit Insurance Product*, Master Thesis, Erasmus University of Rotterdam
- [13] Christiansen, M., Niemeyer, A. (2012) *The fundamental definition of the Solvency Capital Requirement in Solvency II*, Working Paper, Ulm University
- [14] DAV (2015), *Proxy-Modelle für die Risikokapitalberechnung*, DAV Ausschuss Investment, Arbeitsgruppe Aggregationstechniken, Technical Report
- [15] Directive 2009/138/EC of the European Parliament and of the Council on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II)
- [16] EIOPA (2014), *The underlying assumptions in the standard formula for the Solvency Capital Requirement calculation*
- [17] European Commission (2010), *QIS 5 Technical Specifications*, Annex to Call for Advice from CEIOPS on QIS5
- [18] European Insurance CFO Forum (2009), *Market Consistent Embedded Value Principles*
- [19] Floryszczak, A., Le Courtois, O., Majri, M. (2016), *Inside the Solvency 2 Black Box: Net Asset Values and Solvency Capital Requirements with a Least-Squares Monte-Carlo Approach*, Insurance: Mathematics and Economics, Vol. 71, November 2016, pp. 15-26
- [20] Friedman, J.H. (2001), *Greedy function approximation: A gradient boosting machine*, Annals of Statistics, Vol. 29, no. 5, pp. 1189–1232
- [21] Friedman, J.H. (2002), *Stochastic gradient boosting*, Computational Statistics & Data Analysis, Vol. 38, no. 4, pp. 367–378
- [22] Glasserman, P. (2003), *Monte Carlo Methods in Finance Engineering*, Springer

- [23] Gordy, M.B., Juneja, S. (2010), *Nested Simulation in Portfolio Risk Measurement*, Management Science, Vol. 56, No. 10, pp. 1833-1848
- [24] Ha, H., Bauer, D. (2018) *A Least-Squares Monte Carlo Approach to the Calculation of Capital Requirements*, Working Paper, Georgia State University
- [25] Hastie, T., Tibshirani, R., Friedman, G. (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd Edition, Springer
- [26] Hejazi, S.A., Jackson, K.R. (2017), *Efficient valuation of SCR via a neural network approach*, Journal of Computational and Applied Mathematics, Volume 313, March 2017, pp. 427-439
- [27] Higham, D.J (2004), *An Introduction to Financial Option valuation: Mathematics, Stochastics and Computation*, Cambridge University Press
- [28] Hörig, M. (2013), *Using least-squares Monte Carlo in a multi-year context*, InsuranceERM Expert Papers
- [29] Hörig, M., Murray, K., Phelan, E., Leitschkis, M. (2013), *An application of Monte Carlo proxy techniques to variable annuity business: a case study*, Milliman Research Report
- [30] Jonen, C. (2011), *Efficient Pricing of High-Dimensional American-Style Derivatives: A Robust Regression Monte Carlo Method*, Ph.D. Thesis, Universität zu Köln
- [31] Krah, A.S., Nikolić, Z., Korn, R. (2018), *A Least-Squares Monte Carlo Framework in Proxy Modeling of Life Insurance Companies*, Risks, MDPI, Vol. 6(2):62, June 2018
- [32] Koursaris, A. (2010), *Calculating the Solvency Capital Requirement, parts 1-3*, InsuranceERM Expert Papers
- [33] Koursaris, A. (2011), *A primer in replicating portfolios*, InsuranceERM Expert Papers
- [34] Lan, H., Nelson, B.L., Staum, J. (2010), *A Confidence Interval Procedure for Expected Shortfall Risk Measurement via Two-Level Simulation*, Operations Research, Vol. 58, No. 5, pp. 1481-1490
- [35] Larcher, M.N.S. (2015), *Modelli statistici per la valutazione del capitale con un'applicazione a Solvency II*, Tesi di Laurea Magistrale in Scienze Statistiche, Università degli Studi di Padova

- [36] Longstaff, F.A., Schwartz E.S. (2001) *Valuing American Options by Simulation: A Simple Least-Squares Approach*, The Review of Financial Studies, Spring 2001 Vol. I4. No. I, pp. 113-147
- [37] Liu, M., Nelson, B.L., Staum, J. (2010) *An efficient simulation procedure for point estimation of expected shortfall*, Proceedings of the 2010 Winter Simulation Conference
- [38] Manukian, A.V. (2012) *Approccio stocastico alla modellizzazione del requisito di capitale in ottica Solvency II*, Tesi di Laurea Magistrale, Università degli Studi di Trieste
- [39] McNeil, A.J., Frey, R., Embrechts, P. (2005) *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton Series in Finance
- [40] Morrison, S., Turnbull, C., Vysniauskas, N. (2013) *Multi-year projection of market consistent liability valuations*, Moody's Analytics Research
- [41] Natolski, J., Werner, R. (2014) *Mathematical analysis of different approaches for replicating portfolios*, European Actuarial Journal, Vol. 4, Issue 2, pp. 411–435
- [42] Natolski, J. (2017) *Mathematische Fundierung und Analyse replizierender Portfolios in der deutschen Lebensversicherung*, Ph.D. Thesis, University of Augsburg
- [43] Natolski, J., Werner, R. (2017) *Mathematical Analysis of Replication by Cash Flow Matching*, Risks 2017, 5(1), 13, MDPI
- [44] Niederreiter, H. (1992) *Random Number Generation and Quasi Monte Carlo Methods*, Society for Industrial and Applied Mathematics Philadelphia
- [45] Nikolić, Z., Jonen, C., Zhu, C. (2017), *Robust Regression Technique in LSMC Proxy Modeling*, Der Aktuar, 2017 (1), 8–16
- [46] Nilsson, M., Sandberg, E. (2018), *Application and Evaluation of Artificial Neural Networks in Solvency Capital Requirement Estimations for Insurance Products*, Master Thesis, KTH Royal Institute of Technology
- [47] Olivieri, A., Pitacco, E. (2015), *Introduction to Insurance Mathematics. Technical and Financial Features of Risk Transfers*, 2nd Edition, Springer
- [48] Pelsser, A. (2003) *Pricing and hedging guaranteed annuity options via static option replication*, Insurance: Mathematics and Economics, Vol. 33, Issue 2, pp. 283-296

- [49] Pelsser, A., Schweizer, J. (2016) *The difference between LSMC and replicating portfolio in insurance liability modeling*, European Actuarial Journal, Vol. 6, Issue 2, pp. 441-494
- [50] Pfeifer, D., Strassburger D. (2008) *Solvency II: stability problems with the SCR aggregation formula*, Scandinavian Actuarial Journal, 2008:1, pp. 61-77
- [51] Pitacco, E., Denuit, M., Haberman, S., Olivieri, A. (2009) *Modelling Longevity Dynamics for Pensions and Annuity Business*, Oxford University Press
- [52] Sandström, A. (2007) *Solvency II: Calibration for skewness*, Scandinavian Actuarial Journal, 2007:2, pp. 126-134
- [53] Sobol' I.M. (1967) *On the distribution of points in a cube and the approximate evaluation of integrals*, USSR Computational Mathematics and Mathematical Physics, Vol. 7, Issue 4, pp. 86-112
- [54] Society of Actuaries (2016) *Economic Scenario Generators: A Practical Guide*
- [55] Spanos, A. (2006) *The Curve-Fitting Problem, Akaike-type Model Selection, and the Error Statistical Approach*, Virginia Tech working paper
- [56] The IFoA Aggregation and Simulation Working Party (2016), *Simulation based capital models – testing, justifying and communicating choices*, presented to the Institute and Faculty of Actuaries 27th June 2016
- [57] The Proxy Model Working Party (2014), *Heavy Models, Light Models and Proxy Models*, presented to the Institute and Faculty of Actuaries 24th February 2014, pp. 51-76
- [58] Teuguaia, O.N., Ren, J., Planchet, F. (2014), *Internal model in life insurance : application of least squares monte carlo in risk assessment*, ISFA University Papers
- [59] Vounzi, P. (2017), *Calcul de la Probabilité Distribution Forecast du Best Estimate via l'approche LSMC à partir d'un modèle de flexing*, Mémoire d'actuariat, Institut des Actuaire & Société Generali Vie
- [60] Wood, S.N. (2017), *Generalized Additive Models: An Introduction with R*, Second Edition, Chapman & Hall/CRC Texts in Statistical Science