

Operational Risk Management

LA MISURAZIONE DEI RISCHI

Area Risk Management
Servizio Rischi Operativi e Reputazionali

- ✓ ***tipo documento:*** manuale
- ✓ ***autore:*** Servizio Rischi Operativi e Reputazionali
- ✓ ***revisore:*** Servizio Rischi Operativi e Reputazionali
- ✓ ***agente autorizzante e proprietario:*** Servizio Rischi Operativi e Reputazionali
- ✓ ***date di Sviluppo e approvazione:*** giugno 2017
- ✓ ***versione numero*** 26

Storia delle modifiche al documento

Versione 24 rispetto alla versione 23

- Aggiunta Sezione 1.6 su richiesta della Funzione Audit
- Eliminato il termine “margine di intermediazione” da due punti nel Capitolo 2
- Aggiornato lo schema in Figura 1

Cambiamenti minimi apportati in alcuni punti di Sezione 1.3 riguardanti le tabelle “Eventi AQ_Delta P.A.R”, “Macroeventi AQ_Delta P.A.R” ed i recuperi non assicurativi.

Versione 25 rispetto alla versione 24

- Aggiunta nota in Sezione 4.6 per chiarire i criteri di allocazione del capitale regolamentare (Loss Data Collection e Deduzione per perdite attese).
- Spostata dalla sezione 4 alla Sezione 3.3.1 la descrizione della perdita totale lorda usata per i dati interni.

Versione 26 rispetto alla versione 25

- Sezione 3.3.1: modifica della profondità storica del data set dei dati interni ed esclusione delle perdite relative ad eventi di confine con il rischio di credito (“credit boundary”);
- Sezione 3.3.2: introduzione dello scaling dei dati esterni;
- Sezione 3.3.3: modificati i filtri utilizzati per la costruzione del data set di calcolo sia relativamente alla profondità storica per i dati interni, sia per l’inclusione di tutti i dati esterni del DIPO ad eccezione delle revocatorie fallimentari;
- Sezione 4.1.1: Modifica alla costruzione del data set per le analisi di severity;
- Aggiunta nuova Sezione 4.1.4 relativa alla definizione della metodologia di scaling;
- Sezione 4.2.5: Inserito il dettaglio delle periodicità e dei controlli che si effettuano relativamente alle assunzioni di correlazione adottate tra le domande di scenario. Inoltre sono stati descritti i criteri adottati per evitare aggregazioni o frammentazioni artificiose tra gli scenari, definendo gli Ambiti ed i Temi.

INDICE

1	GLOSSARIO	7
2	INTRODUZIONE.....	9
3	LE COMPONENTI DEL PROCESSO DI MISURAZIONE	10
3.1	PREMESSA	10
3.2	LA SOLUZIONE ADOTTATA.....	12
3.3	GLI INPUT DELLA COMPONENTE QUANTITATIVA	13
3.3.1	<i>Dati Interni</i>	<i>13</i>
3.3.2	<i>Dati Esterni</i>	<i>15</i>
3.3.3	<i>Costruzione data set per il calcolo del requisito patrimoniale.....</i>	<i>17</i>
3.4	GLI INPUT DELLA COMPONENTE QUALITATIVA.....	19
3.5	PROCESSO DI PASSAGGIO IN PRODUZIONE DEL MODELLO DI CALCOLO E AGGIORNAMENTO DELLA DOCUMENTAZIONE METODOLOGICA.	20
3.6	CALENDARIO DEL PROCESSO DI CALCOLO DEL REQUISITO PATRIMONIALE E PRODUZIONE DEI FLUSSI DI SEGNALAZIONE DI VIGILANZA.	21
4	LOSS DISTRIBUTION APPROACH	23
4.1	ANALISI DATI STORICI.....	23
4.1.1	<i>Data set di Analisi</i>	<i>23</i>
4.1.1.1	Costruzione del Dataset per le analisi di frequency	23
4.1.1.2	Costruzione del Dataset per le analisi di severity	24
4.1.2	<i>Exploratory data analysis</i>	<i>25</i>
4.1.3	<i>Distribuzione di frequency</i>	<i>26</i>
4.1.4	<i>Distribuzione di severity.....</i>	<i>27</i>
4.1.4.1	Descrizione della metodologia di scaling dei dati esterni di coda	29
4.1.4.2	Determinazione della soglia per l'applicazione della EVT.	31
4.1.4.3	Stima parametri della Generalized Pareto Distribution e intervalli di confidenza	38
4.1.4.4	Test di Goodness of Fit.....	43
4.1.4.5	Criteri e linee guida per la scelta della soglia EVT	44
4.1.4.6	Stima della distribuzione del corpo.	46
4.1.5	<i>Distribuzione delle perdite aggregate per Event Type.....</i>	<i>47</i>
4.1.5.1	Convoluzione Frequency - Severity.....	48
4.1.5.2	Informazione contenuta nella distribuzione di perdita.....	51
4.1.6	<i>Intervalli di Confidenza del VaR.....</i>	<i>52</i>
4.1.6.1	Valutazione dell'errore di simulazione	52
4.1.6.2	Analisi di sensitivity del VaR	53
4.1.7	<i>Assunzioni.....</i>	<i>53</i>
4.2	SCENARIO ANALYSIS	54
4.2.1	<i>Introduzione.....</i>	<i>54</i>
4.2.2	<i>Stime soggettive.....</i>	<i>54</i>
4.2.3	<i>Analisi quantitativa: approccio attuariale.....</i>	<i>55</i>
4.2.4	<i>Calcolo dei parametri distribuzionali a partire dalle stime soggettive....</i>	<i>57</i>
4.2.5	<i>Calcolo del VaR qualitativo per classe di rischio interna</i>	<i>61</i>
4.2.6	<i>Calcolo del VaR qualitativo per Società.....</i>	<i>64</i>
4.2.7	<i>Archiviazione dei risultati</i>	<i>64</i>
4.3	INTEGRAZIONE QUALI-QUANTITATIVA.....	65
4.3.1	<i>Indicatore di continuità/discontinuità</i>	<i>66</i>
4.3.1.1	La costruzione degli indicatori	68
4.3.1.2	Gli indicatori per comparto	69
4.3.2	<i>La costruzione dei pesi per la componente di scenario</i>	<i>70</i>
4.3.2.1	Indicatori di continuità/discontinuità per event type.....	70

4.3.2.2	Assegnazione dei pesi per event type.....	73
4.3.3	Calcolo della distribuzione mistura per l'integrazione tra la componente storica e la componente di scenario	75
4.4	INTEGRAZIONE TRA EVENT TYPE: APPROCCIO T-COPULA	77
4.4.1	Input alla meta-t-copula	78
4.4.2	Stima empiriche dei parametri di interesse.....	80
4.4.3	Prima esecuzione del "metodo di discesa"	81
4.4.4	Seconda esecuzione del "metodo di discesa".....	82
4.4.5	Calcolo del VaR complessivo diversificato	82
4.4.6	Allocazione del beneficio di diversificazione: Expected Conditional Loss	83
4.5	CALCOLO DELLA DETRAZIONE PER LE PERDITE ATTESE	84
4.5.1	Il perimetro di interesse.....	85
4.5.2	Il calcolo della detrazione	86
4.6	ALLOCAZIONE DEL CAPITALE REGOLAMENTARE	87
4.7	PROVE DI STRESS, ANALISI DI SENSITIVITY E BACKTESTING DEL MODELLO AMA.	89
4.7.1	Stress di modello	89
4.7.2	Scenario Stress.....	90
4.7.3	Reverse stress	90
4.7.4	Stress congiunto con gli altri rischi	90
4.7.5	Backtesting del modello AMA	91
4.8	BENCHMARKING DELLE PERDITE INTERNE RISPETTO ALLE PERDITE DI SISTEMA	92
5	APPROFONDIMENTI TEORICI.....	93
5.1	ANALISI DATI STORICI	93
5.1.1	Pre-analisi	93
5.1.1.1	Autocorrelazione Seriale.....	93
5.1.1.2	Autocorrelation Plot	93
5.1.1.3	Plot of Records Development	94
5.1.2	Strumenti d'analisi grafica	95
5.1.2.1	Box Plot	95
5.1.2.2	Mean Excess Plot.....	96
5.1.2.3	Hill Plot - stimatore di Hill.....	96
5.1.2.4	Rapporto Massimo - Somma	96
5.1.3	Metodi di Stima.....	98
5.1.3.1	Minimi Quadrati (Least Squares -LS).....	98
5.1.3.2	Maximum Likelihood Estimation (MLE)	98
5.1.3.3	Maximum Penalized Likelihood Estimation (MPLE)	101
5.1.3.4	Probability Weighted Moments (PWM)	105
5.1.3.5	Minimum Density Power Divergence Estimator (MDPDE)	107
5.1.3.6	Maximum Goodness of Fit Estimator (MGF)	109
5.1.4	Test Statistici.....	111
5.1.4.1	Test di autocorrelazione	111
5.1.4.2	Test sulla bontà di adattamento	116
5.1.5	Extreme Value Theory	120
5.1.5.1	Logica Block Maxima.....	120
5.1.5.2	Logica Peaks Over Threshold (POT)	121
5.2	APPROFONDIMENTI SULLE DISTRIBUZIONI UTILIZZATE	123
5.2.1	Distribuzioni di frequenza	123
5.2.1.1	Poisson	123
5.2.1.2	Binomiale Negativa.....	123
5.2.2	Distribuzioni di severity.....	124
5.2.2.1	Weibull	124
5.2.2.2	Lognormale	126
5.2.2.3	Gamma.....	128
5.3	CONVOLUZIONE DI DISTRIBUZIONI.....	129

5.3.1	<i>Convoluzione via simulazione Monte Carlo</i>	129
5.4	SULLE FUNZIONI COPULA E LA DIPENDENZA DI CODA.....	131
5.4.1	<i>Definizione di funzione copula</i>	131
5.4.2	<i>Dipendenza di coda (tail dependance)</i>	132
5.4.3	<i>Proprietà fondamentali delle copule Gaussiane e t-Student</i>	133
5.5	METODI DI STIMA DEI PARAMETRI DELLA GPD E PROPRIETÀ ASINTOTICHE	135
5.5.1	<i>Metodo della massima verosimiglianza (MLE)</i>	135
5.5.2	<i>Metodo dei momenti pesati in probabilità (PWM)</i>	136
5.5.3	<i>Proprietà asintotiche degli stimatori MLE, PWM ed MPLE</i>	136
5.5.4	<i>Proprietà asintotiche MDPDE</i>	137
5.6	INDICE DI CONTINUITÀ/DISCONTINUITÀ: UN ESEMPIO	141
6	BIBLIOGRAFIA	143

INDICE DELLE FIGURE

Figura 1 Schema del modello interno di misurazione dei Rischi Operativi.	10
Figura 2 Attività di misurazione	11
Figura 3 Approccio AMA MPS.....	12
Figura 4 Time series plot della frequenza di accadimento.....	27
Figura 5 peso w_1 al variare della numerosità n , con $nc = 200$ e $n_0 = 20$	30
Figura 6 peso w_2 al variare della statistica K di Kolmogorov-Smirnov.....	31
Figura 7 Mean Excess Plot	33
Figura 8 Plot rapporto Massimo su somma.....	34
Figura 9 Hill plot del parametro di shape	35
Figura 10 Grafico di confronto fra i vari stimatori – scelta della soglia	37
Figura 11 QQplot.....	38
Figura 12 distribuzioni cumulate delle probabilità	39
Figura 13 Grafico dei periodi di ritorno	40
Figura 14 Distribuzione dei residui: verifica poisson process	41
Figura 15 Boxplot distribuzione simulata del parametro di shape della GPD, metodo bootstrap.	42
Figura 16 KS e AD quadratico applicato alla coda destra.....	44
Figura 17 Analisi di sensitivity del VaR al variare della soglia monetaria	53
Figura 18 Scenario: Convoluzione di frequenza e impatto	57
Figura 19 Spazio Frequenza/Impatti	57
Figura 20 informazioni individuate nei vari comparti	67
Figura 21 Indicatori di continuità/discontinuità (dicembre 2014).....	69
Figura 22 Pesi assegnati ai comparti per ciascun event type (dicembre 2014)	70
Figura 23 Score per event type (dicembre 2014).....	73
Figura 24 mapping	73
Figura 25 Attribuzione peso alle classi di rischio (dicembre 2014).....	74
Figura 26 Attribuzione peso w post add on (dicembre 2014)	74
Figura 27 Attribuzione del peso w all'ET2R.....	75
Figura 28 Perdite trimestrali e percentili distribuzione di perdita	92
Figura 29 Autocorrelation plot.....	94
Figura 30 Plot dello sviluppo dei record	95
Figura 31 Boxplot delle perdite.....	96
Figura 32 Schema di convoluzione.....	130

1 Glossario

Nel presente documento metodologico sono utilizzati i seguenti termini:

- **AMA (Advanced Measurement Approach):** approccio avanzato per la determinazione del requisito patrimoniale sui rischi operativi. L'utilizzo delle stime interne per il calcolo dell'assorbimento patrimoniale è condizionato al preventivo riconoscimento formale dell'Autorità di Vigilanza del rispetto dei requisiti quali/quantitativi previsti dalla normativa.
- **Analisi di Scenario:** modulo rivolto al Top-Management finalizzato alla quantificazione del capitale a rischio sulla base di stime soggettive indipendenti e all'individuazione degli interventi di gestione dei rischi operativi (mitigazione, ritenzione e trasferimento).
- **Assessment:** modulo rivolto al Middle Management finalizzato alla valutazione indipendente della qualità dei presidi (in altri termini, qualità della gestione dei fattori di rischio) posti in essere per la gestione ed il controllo dei singoli eventi di rischio e delle relative priorità d'intervento.
- **Assorbimento patrimoniale:** quantità di capitale allocato a fronte di una determinata esposizione al rischio operativo calcolata mediante il Modello Interno (cfr. Modello Interno più avanti).
- **Business Line (BL):** modello, introdotto dall'accordo di Basilea II, in base al quale si classificano le aree di attività di un'azienda/gruppo bancario.
- **Caso peggiore (Worst case):** impatto dell'evento di rischio nel caso peggiore possibile, definito ad un certo intervallo di confidenza.
- **DIPO:** (Database Italiano delle Perdite Operative) iniziativa di data pooling promossa dall'Associazione Bancaria Italiana.
- **Evento di rischio:** fatto/atto, aziendale o extra-aziendale, al manifestarsi del quale può conseguire un danno per la società.
- **Event Type (ET):** modello, introdotto dall'accordo di Basilea II, in base al quale si classificano gli eventi di rischio operativo di un'azienda/gruppo bancario.
- **Frequenza attesa (frequency):** numero di volte in cui una determinata tipologia di evento si manifesta nel periodo d'analisi.
- **Impatto medio:** impatto medio della tipologia di eventi oggetto di analisi.
- **Impatto tipico:** impatto che si verifica più frequentemente nella classe di eventi oggetto di analisi.

- **LDA (Loss Distribution Approach):** Approccio per la stima del capitale a rischio, basato sulla modellizzazione separata del numero di eventi di perdita (frequency) e degli importi delle singole perdite (severity).
- **LDC (Loss Data Collection):** Raccolta e catalogazione dei dati di perdita.
- **OR:** Operational Risk.
- **ORM:** Operational Risk Management.
- **Perdita Attesa:** valore atteso della distribuzione delle perdite potenziali.
- **Perdita Inattesa:** differenza fra il valore della distribuzione di perdita aggregata per un definito quantile (es. al 99,9%) e il suo valore atteso.
- **PTL:** Perdita Totale Lorda dell'evento di rischio operativo
- **Questionario Assessment:** è lo strumento attraverso cui si svolge l'Assessment, definito in relazione ad una stessa unità organizzativa e costituito da una lista di processi di competenza, sui quali si richiedono sia una valutazione di "adeguatezza del presidio" rispetto ai rischi operativi tipici del processo, sia opportune misure di mitigazione in caso di presidio non adeguato.
- **Questionario Analisi di Scenario:** è lo strumento attraverso cui si svolge lo Scenario, è costituito sulla base delle analisi dei risultati della LDC e dell'Assessment.
- **Severity:** è la distribuzione di impatto delle perdite operative
- **Stime soggettive:** valutazione della perdita attesa ed inattesa effettuata sulla base di valutazioni qualitative dei responsabili di business, espressa in un contesto coerente da un punto di vista statistico secondo predefinite logiche di analisi di scenario.
- **VaR (Value at Risk):** misura di rischio calcolata come somma tra perdite attese (EL) ed inattese (UL), dato un orizzonte temporale di un anno ed un intervallo di confidenza del 99,9% (ovvero la probabilità che le perdite subite assumano al massimo quel valore). Si definiscono Quantitative, Qualitative ed Integrated VaR.

2 INTRODUZIONE

Il presente documento illustra il macroprocesso “Misurazione” con l’obiettivo di descrivere la metodologia statistica utilizzata al fine di indagare il profilo di rischio complessivo del Gruppo Montepaschi e determinare la quantificazione del capitale a rischio.

3 LE COMPONENTI DEL PROCESSO DI MISURAZIONE

3.1 Premessa

La direttiva del Gruppo Montepaschi in “materia di gestione e governo dei rischi operativi” approvata dal CdA, schematizza il framework del Gruppo relativo all’adozione di un approccio avanzato di misurazione (c.d. AMA - Advanced Measurement Approach), come riportato in Fig.1:

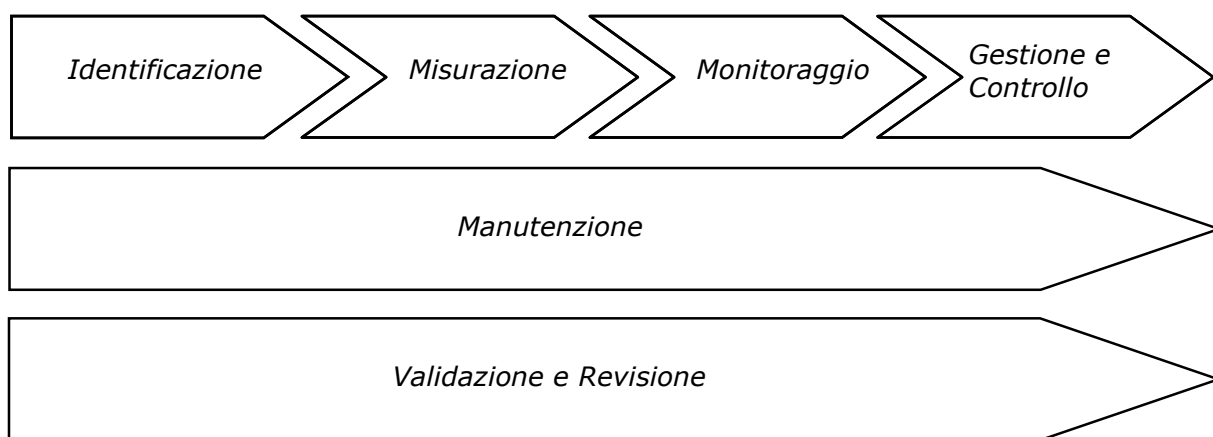


Figura 1 Schema del modello interno di misurazione dei Rischi Operativi.

Per completezza espositiva, si richiama la descrizione sintetica presente in normativa di ciascuna componente del framework:

- Identificazione: volta alla ricerca e al censimento delle perdite operative, di natura oggettiva o soggettiva, attraverso l’analisi di opportune fonti;
- Misurazione: volta alla determinazione del capitale economico e regolamentare a livello di Gruppo e per le singole società;
- Monitoraggio: finalizzato alla costante verifica degli assorbimenti patrimoniali e delle indicazioni strategiche sulla propensione al rischio sia in termini gestionali che patrimoniali;
- Gestione/controllo: finalizzato all’individuazione delle azioni di mitigazione/trasferimento/ritenzione del rischio;
- Manutenzione: volto ad assicurare nel continuo un progressivo aggiornamento del modello adottato nella gestione dei rischi operativi, della normativa, dei processi, degli applicativi.

- Validazione e revisione, per garantire le opportune attività di controllo di compliance regolamentare, interna ed esterna, del modello. Fino al 2013 l'attività di validazione è stata svolta in regime di self-validation, a seguire è stata istituita un'apposita funzione di Validazione dedicata a tale compito.¹

Il documento si focalizza sulla misurazione, le cui attività si articolano nelle seguenti analisi:

- **Analisi Quantitativa dei dati storici (Interni e DIPO)**, il cui obiettivo è quello di analizzare i dati interni di perdita rilevati tramite Loss Data Collection integrati con i dati esterni provenienti dal processo di data-pooling (DIPO) al fine di determinare il capitale a rischio per ogni unità del Gruppo e per il Gruppo Montepaschi nel suo complesso (*Quantitative VaR*). Il modello utilizzato è quello attuariale, che prevede analisi separate per le distribuzioni di frequenza ed di impatto e la stima della distribuzione aggregata di perdita per convoluzione.
- **Analisi Quantitativa su stime soggettive**, il cui obiettivo è quello di analizzare le stime soggettive di Scenario Assessment per la quantificazione del VaR per ogni dimensione di analisi rilevante (ET, UO etc.) e per il Gruppo nel suo complesso (*Qualitative VaR*). La metodologia adottata permette di misurare il rischio con un output confrontabile ed integrabile con quello basato sui dati di perdita.
- **Integrazione**, il cui obiettivo è quello di quantificare il Requisito Patrimoniale minimo di Gruppo per i rischi operativi. Il processo comprende una metodologia di integrazione fra gli output del processo di analisi quantitativa sui dati interni ed esterni e di analisi quantitativa sulle stime soggettive in modo da ottenere un Integrated VaR.

Le attività di misurazione possono essere schematizzate come segue:

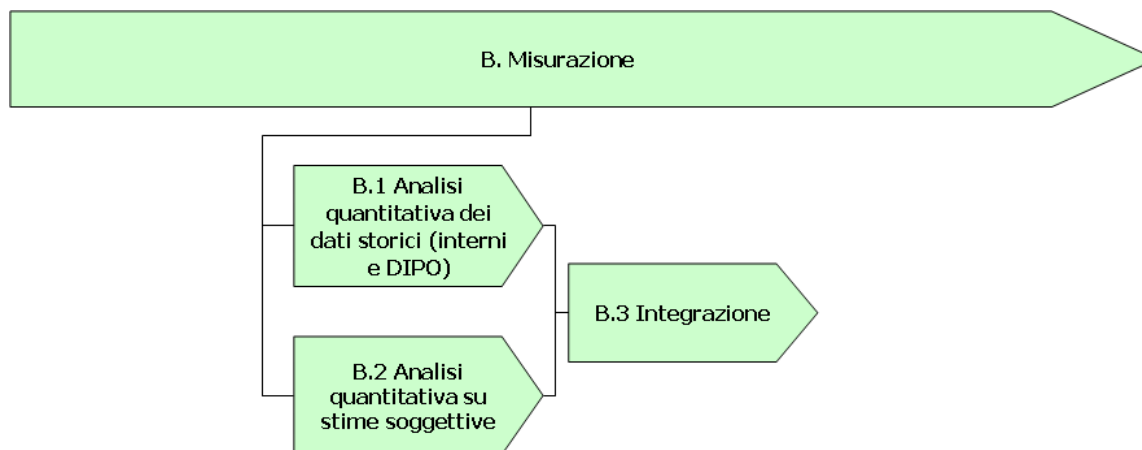


Figura 2 Attività di misurazione

¹ Da “Regolamento 1” l'attività è attualmente in carico al *Servizio Validazione e Monitoraggio*, definito presso l'Area Validazione, Monitoraggio e Risk Reporting, Direzione Rischi.

3.2 La soluzione adottata

L'approccio adottato dal Gruppo Montepaschi può essere così schematizzato:

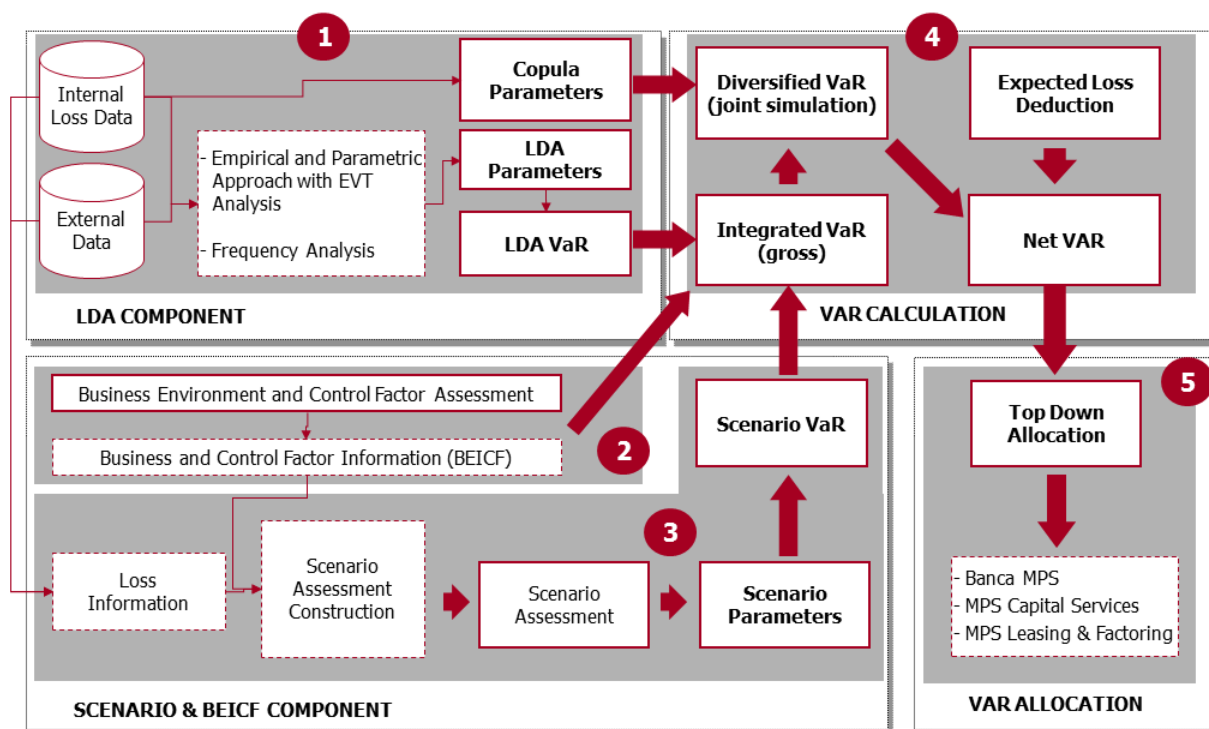


Figura 3 Approccio AMA MPS

Per ciò che concerne la componente quantitativa l'approccio di tipo attuariale modella separatamente la frequenza con cui si verificano gli eventi di perdita e l'impatto economico con il quale ciascuno di essi si può manifestare. La soluzione adottata da MPS prevede di utilizzare i dati interni di perdita per modellizzare la frequenza di accadimento degli eventi ed il set campionario dei dati esterni (forniti semestralmente dal consorzio interbancario DIPO²) integrati con i dati MPS per la distribuzione degli impatti. Per quest'ultima l'adozione della teoria dei valori estremi (EVT) permette la modellizzazione separata del corpo e della coda tramite l'utilizzo di due distribuzioni diverse.

La componente qualitativa del modello di misurazione al fine di costruire i questionari per le analisi di scenario:

- Utilizza i risultati del processo di Assessment sulla qualità del contesto operativo valutata attraverso i fattori di rischio e di controllo. Tale valutazione è supportata dall'utilizzo di **Key Risk Indicators** (KRIs), costruiti per il monitoraggio dei principali cambiamenti del contesto operativo.

² Database Italiano delle Perdite Operative, <http://www.dipo-operationalrisk.it/>

- Analizza le perdite storiche più rilevanti raccolte nel processo di LDC al fine di individuare i fenomeni importanti per la costruzione dello scenario.
- Integra le informazioni con opinioni di esperti a livello locale e di gruppo (auditing, organizzazione, legale, compliance, esperti del business, ecc.)
- Integra con le informazioni provenienti dall'estero, da media/internet, da perdite esterne individuate come estreme e da informazioni contenute nelle relazioni dell'Arbitro Bancario e Finanziario.

Ogni questionario di scenario è organizzato per sezioni ognuna delle quali definisce un'area di criticità ed i relativi eventi di rischio sui quali si raccoglie l'auto-valutazione del Top Management al fine di determinare il VaR qualitativo.

La componente d'integrazione prevede di determinare una misura unica di VaR a partire dai VaR storico e qualitativo, il cosiddetto VaR integrato, che rappresenti una stima attendibile e robusta del profilo di rischio complessivo del Gruppo Montepaschi.

3.3 Gli input della componente quantitativa

3.3.1 Dati Interni

I dati interni sono utilizzati per modellizzare la frequenza di accadimento.

Per quanto riguarda l'importo si è scelto di utilizzare la perdita totale lorda, che include anche accantonamenti e perdite stimate. In questo modo si considerano accaduti anche quegli eventi la cui perdita è solo stimata, anche se non ha ancora avuto una sua realizzazione economica, cioè non è stata contabilizzata.

Per gli eventi interni, il Gruppo Montepaschi adotta una soglia di raccolta pari a 0€, fatta eccezione di dati provenienti da alcune fonti specifiche quali le partite irrecoverabili (per le partite irrecoverabili acquisite automaticamente la soglia di raccolta è pari a 50€, mentre per quelle processate manualmente la soglia è 500€, vedasi il Manuale delle Fonti). Tuttavia nel modello di calcolo si applica una soglia di 5000€ ai dati di perdita interni facenti parte delle classi di rischio nelle quali vengano utilizzati anche dati di sistema (DIPO), al fine di mantenere la coerenza con il database DIPO.

D'altronde i dati al di sotto di tale soglia cumulativamente rappresentano una piccola percentuale del totale delle perdite e possono quindi non essere presi in considerazione nella determinazione del capitale a rischio senza perdita di significatività della stima.

La frequenza è stimata sugli Eventi di rischio e non sui singoli Effetti di perdita verificatisi, dato che la modalità di raccolta prevista dalla metodologia di LDC censisce raccogliendo sotto un unico evento di rischio tutte le perdite operative generate da un unico accadimento.

La profondità storica del set di dati utilizzato per il calcolo del requisito, a partire dalla semestrale di Giugno 2017, è stata estesa a 10 anni a seguito della Follow up Letter, ricevuta a

Febbraio 2017, relativa all'Internal Model Investigation 42 (IMI42) avvenuta a settembre 2015.

Nel database sono presenti una serie di date relative allo stesso evento:

- **Data di Accadimento:** la data in cui si verifica l'evento pregiudizievole.
- **Data Fine Accadimento:** viene valorizzata quando l'evento non è puntuale ma si verifica in un determinato arco di tempo (es una frode che è perpetrata per anni prima di essere rilevata). Anche in questo caso può non essere possibile identificare univocamente il giorno in cui l'evento ha cessato di manifestarsi.
- **Data di Rilevazione:** data in cui si è venuti a conoscenza dell'evento pregiudizievole e dell'effetto/effetti riconducibili al singolo evento.
- **Data di Prima Contabilizzazione:** data in cui gli effetti sono contabilizzati.

La data di accadimento ha il vantaggio di essere coerente con il flusso di dati DIPO, ma lo svantaggio di non essere sempre univocamente determinata, oppure di non essere valorizzata.

La data di contabilizzazione non presenta le criticità della data di accadimento, essendo sempre univocamente valorizzata. Tuttavia l'utilizzo di tale data comporterebbe la presenza di aggregazione dei dati di perdita attorno a periodi dell'anno tipici dal punto di vista contabile, corrispondenti alle date di bilancio, portando alla violazione del principio di indipendenza e identica distribuzione dei dati.

La data di rilevazione è univocamente assegnata e sconta meno l'affollamento di registrazioni su periodi dell'anno particolari (tipicamente fine ed inizio anno); inoltre permette di tenere conto di eventi accaduti nel passato che generano perdite a distanza di tempo (ad esempio le cause legali). In questi ultimi casi, infatti, l'utilizzo della data di accadimento porterebbe all'esclusione dal dataset di numerose perdite. Si osserva che la data di rilevazione per vincolo tecnico, è sempre minore o uguale alla relativa data di contabilizzazione.

Alla luce di queste valutazioni è stato ritenuto opportuno utilizzare nel modello la data di rilevazione.

Il Gruppo Montepaschi per escludere dall'analisi perdite accadute a troppa distanza dalla data di riferimento delle analisi e quindi scarsamente informative, si riserva inoltre di escludere ulteriormente dal dataset di calcolo del requisito patrimoniale le perdite con data di fine accadimento superiore ai 10 anni precedenti l'analisi.

Per tenere conto nel calcolo del requisito delle manifestazioni economiche (variazioni di accantonamenti o perdite spese) contabilizzate nel periodo di osservazione ma relative ad eventi di rischio operativi rilevati al di fuori di tale periodo, sono state predisposte due tabelle specifiche nell'applicativo OpriskEv : "Eventi AQ_Delta P.A.R" e "Macroeventi AQ_Delta P.A.R". Nella prima tabella vengono raccolte sistematicamente tutte le variazioni di importo (specificamente: perdite, accantonamenti, recuperi non assicurativi) contabilizzate nel periodo di

misurazione (e contabilizzate a partire dal primo gennaio 2010³), aggregate in base al codice dell'evento al quale esse si riferiscono,⁴ escludendo i codici relativi a Macroeventi; le variazioni relativi a codici di Macroeventi vengono invece raccolte nella seconda tabella. Delle variazioni riportate nelle due tabelle si selezionano quelle relative ad eventi/macroeventi con data di rilevazione precedente all'inizio del periodo di misurazione: i dati risultanti vengono utilizzati nel modello di misurazione esclusivamente per la stima della severity, dato che a livello di frequency essi sono già stati considerati in passato; si veda la Sezione 3.3.3 per dettagli.

Per allinearsi alla Circolare 575/2013 relativa ai requisiti prudenziali per gli enti creditizi e le imprese di investimento, a partire dalla semestrale di Giugno 2017, sono escluse dalla base dati di calcolo del requisito patrimoniale le perdite operative dovute a frodi di confine con il credito con posizione a contenzioso (credit boundary).

Tali perdite vengono individuate utilizzando le segnalazioni della funzione Audit e l'ammontare di perdita classificato a contenzioso nei database di rischio di credito viene utilizzato come importo di perdita per l'evento di rischio operativo. Si tratta in particolare di eventi classificati come Event Type 1 (frode interna) ed Event Type 2 (frode esterna).

Queste loss di confine con il credito, rappresentando l'ammontare di perdita delle pratiche a contenzioso, sono già trattate ai fini del Rischio di Credito e sono monitorate solo a fini gestionali nell'ambito dei rischi operativi, come previsto dall'Art. 30(1) degli EBA Final Draft RTS.

3.3.2 Dati Esterni

Il Gruppo Montepaschi ha aderito al consorzio interbancario DIPO (Database Italiano delle Perdite Operative), iniziativa intrapresa a livello nazionale. Non ha ritenuto opportuno aderire a consorzi internazionali in quanto le banche partecipanti presentano differenze sostanziali sia per dimensione, sia per tipologia di business.

Il Gruppo Montepaschi ha scelto di utilizzare il flusso di ritorno dei dati DIPO per arricchire la propria distribuzione interna in modo da perfezionare la stima della distribuzione degli importi di perdita (severity). Il Gruppo si riserva la possibilità di valutare l'inclusione o meno, nel proprio dataset di calcolo, delle perdite operative di Sistema per le quali sia evidente l'assoluta incoerenza con il proprio business. A tal fine il Gruppo Montepaschi adotta la seguente procedura semestrale, in corrispondenza dell'arrivo di un nuovo flusso di ritorno DIPO.

Per ciascuna classe di rischio interna al modello AMA che preveda l'utilizzo dei dati esterni si procede:

1. individuando "outlier potenziali" nei dati esterni.

Per ciascuna classe di rischio, il massimo importo tra i dati esterni viene marcato come "outlier potenziale" se:

³ Non si considerano le variazioni contabilizzate in data precedente al primo gennaio 2010, in quanto i criteri di censimento adottati prima di tale data rendono poco affidabile la base dati delle variazioni.

⁴ Specificamente, per ciascun codice evento si calcola un unico importo "Delta", aggregando tutti i movimenti contabilizzati relativi a tale evento, mediante la formula "Perdite+Accantonamenti-Recuperi non assicurativi".

- a. il massimo esterno è maggiore di 10 volte il secondo importo massimo tra i dati esterni, si tratta quindi di un outlier rispetto ai dati di sistema, oppure
 - b. il massimo esterno è maggiore di 10 volte l'importo massimo tra i dati interni, si tratta quindi di un outlier rispetto ai dati interni.
2. Nel caso che al passo 1 sia stato individuato un outlier potenziale, il dato viene escluso dal dataset di calcolo. Tuttavia il Gruppo si riserva la facoltà di includerlo comunque, in via prudenziale, nel caso di classi di rischio scarsamente popolate o sulla base di analisi sulle distribuzioni come, ad esempio, sui rapporti tra i percentili elevati delle distribuzioni dei dati esterni e dei dati interni.
3. In tutti gli altri casi il massimo esterno viene incluso nel dataset di calcolo.

Nel caso in cui il dato esterno sia considerato “outlier” e venga escluso dal dataset di calcolo, tale dato viene utilizzato comunque in sede di Analisi di Scenario.

I dati esterni vengono utilizzati solo per la stima di severity, per descrivere la frequenza aggregata annuale, infatti, si è scelto di utilizzare i soli dati interni, al fine di riflettere sia la dimensione sia il contesto operativo aziendale interno al Gruppo.

Diversamente per la severity si è ritenuto opportuno aumentare la numerosità delle serie storiche con i dati delle altre banche aderenti al DIPO al fine di ottenere una migliore stima della relativa distribuzione. In questo modo il Gruppo Montepaschi raggiunge lo scopo di:

1. aumentare l'affidabilità della stima prodotta dovuta essenzialmente al maggiore numero di perdite operative presenti nel set di dati DIPO a disposizione per il fitting;
2. Ottenere una stima del capitale regolamentare conservativa dovuta all'introduzione nel set di dati estremi non presenti nella raccolta dati interna;

Relativamente ai dati di sistema, a partire da giugno 2017, il Gruppo Montepaschi ha adottato un metodo di riscaldamento (“*scaling*”) della distribuzione dei dati DIPO, descritto nella Sezione 4.1.4.1. Questo meccanismo di *scaling* è stato introdotto al fine di scongiurare oscillazioni inattese del requisito AMA per effetto di fenomeni esterni di dimensioni significative, ritenuti non coerenti con l'attuale e prospettico il profilo di rischio del Gruppo Montepaschi

Per alcuni casi specifici il Gruppo Montepaschi si riserva la possibilità di utilizzare i propri dati interni anche per la modellizzazione della severity, in particolare nel caso in cui si perda l'allineamento con DIPO. Ovviamente è necessario che, in tali casi, la numerosità del dataset disponibile sia sufficiente per poter effettuare un'analisi statistica.

3.3.3 Costruzione data set per il calcolo del requisito patrimoniale

Il modello AMA utilizza dati interni MPS e dati esterni provenienti dal consorzio DIPO. I dati interni vengono selezionati per data di rilevazione in una finestra temporale ampia almeno 10 anni: per il calcolo del requisito di fine anno (es: 31/12/2017) si considerano gli ultimi **10 anni** (data rilevazione compresa tra gennaio 2008 e dicembre 2017), mentre per quelle successive si tiene ferma la data di rilevazione iniziale (nell'esempio gennaio 2008), aggiungendo man mano un trimestre, fino alla fine anno successiva in cui si riaggiorna la data di inizio considerando nuovamente una finestra di **dieci anni** (es: al 31/12/2018 si selezionano dati interni con data di rilevazione compresa tra gennaio 2009 e dicembre 2018).

Con questa modalità vengono selezionati i dati interni provenienti dalle seguenti tabelle:

- Eventi (che non siano parte di un MacroEvento);
- MacroEventi;
- Dubbi Esiti per i dati relativi ai dubbi esiti sui reclami;
- Sospesi Contabili per i dati relativi ai sospesi contabili.

Tali dati vengono utilizzati per il calcolo della distribuzione di frequenza, si veda la Sezione 4.1.1.1. Essi vengono inoltre utilizzati per la stima della distribuzione di severità (si veda la Sezione 4.1.1.2), assieme ai dati DIPO (che vengono selezionati per data di accadimento a partire dal primo gennaio 2003) ed alle variazioni contabili su perdite interne rilevate precedentemente alla data inizio analisi (gennaio 2008 nell'esempio sopra), ma che hanno prodotto manifestazioni economiche all'interno della finestra temporale considerata. Le variazioni contabili vengono selezionate dalle seguenti tabelle:

- Eventi AQ_Delta P.A.R. (Perdite, Accantonamenti, Recuperi non assicurativi);
- MacroEventi AQ_Delta P.A.R.

Sono escluse le variazioni contabili di eventi e macroeventi con data di prima contabilizzazione precedente alla data inizio analisi⁵.

Il modello AMA attualmente prevede 14 classi di rischio. Gli Event Types Frodi interne (ET1), Rapporti di impiego e sicurezza sul lavoro (ET3), Danni a beni materiali (ET5) e Disfunzioni dei sistemi (ET6) corrispondono a singole classi interne del modello AMA. Gli Event Types Frodi esterne (ET2), Clienti prodotti e prassi operative (ET4) ed Esecuzione consegna e gestione dei processi (ET7), sono suddivisi rispettivamente in sottoclassi:

- Altre frodi esterne (ET2AFE), Rapine (ET2R), Carte clonate (ET2CC), Sistemi tradizionali di pagamento (ET2STP);
- Anatocismo (ET4A), Cause (ET4C), Reclami (ET4R);

⁵ Il filtro è stato concordato con l'AdV coerentemente con quanto viene richiesto dal Comitato di Basilea per il Quantitative Impact Study (QIS). Si escludono tutti gli eventi che hanno data di prima contabilizzazione antecedente alla data di inizio analisi, anche se registrano aggiustamenti contabili nella finestra temporale di analisi.

- Cause ed altro (sigla: ET7C), Partite diverse (ET7P), Reclami (ET7R).

Per alcune classi vengono utilizzati solo dati interni: ET2STP, ET4R, ET7P ed ET7R. Per queste classi si usa un filtro sugli importi pari a zero. Per le altre classi vengono utilizzati anche dati esterni DIPO: per queste si utilizza un filtro sugli importi pari a 5,000 EUR (pari alla soglia di raccolta usata dal DIPO). La tabella riporta la suddivisione in classi di rischio interne del modello AMA, l'utilizzo o meno di filtri sugli importi e dati esterni e il dettaglio dei filtri utilizzati per includere i dati interni e DIPO⁶ nelle varie classi di rischio:

ET	Classe di rischio	Filtro importi	su	Filtri dati interni	Filtri dati esterni (DIPO)
1	Frodi interne	5000		ET1	ET1
2AFE	Frodi esterne: Altre frodi esterne	5000		Dal 2.2.1.2 (Frode): 2. 2. 1. 2. 3 Contraffazione di valori 2. 2. 1. 2. 5 Falsificazione e/o utilizzo fraudolento dei poteri di delega / Appropriazioni indebite 2. 2. 1. 2. 6 Altre frodi/atti illeciti di controparti esterne. 2. 2. 1. 2. 7 Erogazione di credito sulla base di documentazione falsa o contraffatta tutto il 2.2.2 Sicurezza dei Sistemi Tutti i sospesi contabili dell'ET2	da ET2 tutti i dati ET2 non inclusi in 2R e 2CC: 2.2.1 Furto e frode 2.2.1-(2.05 DIPO) Altre frodi da esterni 2.2.2 Sicurezza dei Sistemi
2R	Frodi esterne: Rapine	5000		2.2.1.1 Furti e Rapine	da ET2: 2.2.1-(2.03 DIPO) Furti/rapine, scassi ai Bancomat (da esterni)
2CC	Frodi esterne: Carte	5000		2. 2. 1. 2. 4 Utilizzo fraudolento di carte di credito/debito	da ET2: 2.2.1-(2.04 DIPO) Frodi su carte (da esterni)
2STP	Frodi esterne: Sistemi Tradizionali di pagamento	-		da 2.2.1.2 Frode: 2. 2. 1. 2. 1 Utilizzo di assegni contraffatti/trafugati 2. 2. 1. 2. 2 Utilizzo di banconote false	-
3	Rapporti di impiego	5000		ET3	ET3
4A	Clienti: Anatocismo	5000		2. 4. 1. 1. 4 Danni causati ai clienti per una non corretta interpretazione delle norme (effetti retroattivi)	da ET4: Eventi Sistemici codice 509, 510. 4.2 Attività/Pratiche operative o di mercato improprie, tranne Eventi Sistemici con codice 501, 502, 503, 504, 508, 505, 506, 511
4C	Clienti: Cause legali e altro	5000		contiene tutti i dati dell'ET4 non inclusi in ET4A ed ET4R	da ET4: Tutto quanto non incluso nell'ET4A, tranne Eventi Sistemici con codice 507
4R	Clienti: Reclami	-		Eventi dell'ET4 che non ricadono nell'anatocismo e la cui fonte	-

⁶ La classificazione dei dati esterni è fatta sulla base del Manuale DIPO versione 2.4. Il Gruppo Montepaschi si riserva la possibilità di modificare l'attribuzione della classe di rischio in caso di modifiche al Manuale DIPO o qualora emergano informazioni ulteriori per riclassificare i dati esterni sulle corrette classi di rischio per fenomeni omogenei.

ET	Classe di rischio	Filtro su importi	Filtri dati interni	Filtri dati esterni (DIPO)
			informativa è inclusa tra partite diverse, reclami, rimborsi, compliance	
5	Danni a beni materiali	5000	ET5	ET5
6	Disfunzioni dei sistemi	5000	ET6	ET6
7C	Esecuzione: Cause ed altro	5000	contiene tutti i dati dell'ET7 non inclusi in ET7P ed ET7R	ET7 ET4: Eventi Sistemici con codice 507
7P	Esecuzione: Partite diverse	- (*)	Eventi dell'ET7 la cui fonte informativa è Partite Diverse	-
7R	Esecuzione: Reclami	-	eventi dell'ET7 la cui fonte informativa è Reclami, Compliance, Rimborsi, Carte Clonate, Finanza	-

(*) La soglia di raccolta dati è pari a 0 EUR per la quasi totalità degli eventi, con 3 eccezioni che contengono pochissimi eventi, ovvero 50 EUR per le partite irrecuperabili automatiche, 500 EUR per le partite irrecuperabili manuali e 5000 EUR per le rettifiche a conto economico.

Per l'ET2R e l'ET2CC vengono utilizzati dati appartenenti a classi create appositamente dal DIPO, non corrispondenti ad Event Types di livello 2 di Basilea.

I dati DIPO relativi alle revocatorie fallimentari non vengono utilizzati per il calcolo del requisito patrimoniale, in quanto già considerati ai fini del calcolo del requisito a fronte dei rischi di credito.

3.4 Gli input della componente qualitativa

Gli input della componente qualitativa utilizzati per la stima del VaR basato su stime soggettive, derivano dai due seguenti processi:

- **Assessment:** finalizzato all'autovalutazione della qualità dei presidi (contesto operativo) posti in essere per la gestione ed il controllo dei singoli eventi di rischio. Tale valutazione è supportata dall'utilizzo di **Key Risk Indicators** (KRIs), costruiti per il monitoraggio dei principali cambiamenti del contesto operativo.

Analisi di Scenario: finalizzato ad ottenere le stime soggettive del Top Management su frequenza, impatto medio e caso peggiore di eventi di rischio potenziali. Con una solida metodologia statistica alla base le stime sono convertite nel Qualitative VaR.

Nell'Analisi di Scenario gli eventi di rischio sottoposti al giudizio del Top Management sono il risultato di un processo di analisi degli output dell'Assessment e della LDC, arricchite da valutazioni di esperti (Audit, Organizzazione, ecc.), da informazioni provenienti dall'estero, dai media, da eventi estremi rilevati sul sistema e da informazioni contenute nelle relazioni dell'Arbitro Bancario e Finanziario. Le stime sono ottenute a fronte di questionari che evidenziano gli eventi di rischio potenzialmente dannosi e sono legate ad un'area di criticità effettivamente riscontrata nelle precedenti analisi: tale area riguarda sia l'efficacia e l'efficienza operativa, sia l'ambiente di controllo.

In questo modo la metodologia definita permette di cogliere attraverso l'Assessment i fattori del contesto operativo e del sistema dei controlli interni; a sua volta l'Assessment confluisce per costruzione, attraverso lo Scenario, nelle stime soggettive mirate alla valutazione dei rischi che insistono su aree di carenza. Il risultato è che le valutazioni qualitative dell'Assessment opportunamente integrate con tutte le altre informazioni possedute sono convertite, attraverso la metodologia statistica alla base delle Analisi di Scenario, in misure quantitative.

In base al risultato delle analisi di Scenario vengono individuate, quindi, le criticità e i relativi possibili interventi da porre in essere per la loro mitigazione, interventi suggeriti dalle stesse strutture che hanno risposto ai questionari. Tali interventi vengono concordati con le singole strutture, che si fanno carico delle azioni di mitigazione da porre in essere.

Per un approfondimento sull'Assessment e sull'Analisi di Scenario si rimanda al documento sulla Identificazione.

3.5 Processo di passaggio in produzione del modello di calcolo e aggiornamento della documentazione metodologica.

Il modello AMA è implementato in codice R (<http://cran.r-project.org/>) all'interno di un ambiente di esecuzione ed archiviazione dei codici e dei risultati, denominato ASIA, predisposto e mantenuto dal Consorzio Operativo di Gruppo. L'implementazione del modello è conforme a quanto descritto nel presente Manuale. Il modello può essere modificato per una serie di motivi:

1. allineamento a nuove norme di vigilanza;
2. allineamento alle “best practices” nazionali e/o internazionali;
3. interventi correttivi richiesti dalle Autorità di Vigilanza, dalla funzione di Revisione Interna o di Convalida Interna;
4. interventi per esigenze tecniche (ad esempio miglioramento dell'implementazione, efficientamento del motore Monte Carlo, ecc.).

Il regolamento EU 529/2014 distingue tra estensioni e modifiche sostanziali (art.6) e non sostanziali (art.7) ai modelli AMA. L'introduzione di estensioni e modifiche sostanziali è soggetta a previa autorizzazione delle autorità competenti. La domanda di autorizzazione deve essere corredata di una serie di documenti tra i quali le relazioni sulla verifica indipendente o validazione. Le estensioni e modifiche non sostanziali sono soggette a notifica *ex ante* (due mesi prima dell'attuazione), se rientrano nelle categorie e/o nella misura specificate dal suddetto Regolamento. Le altre modifiche che con certezza non rientrano nel suddetto Regolamento sono soggette ad un diverso iter. La decisione di quali modifiche non sostanziali debba essere oggetto di notifica *ex-ante* è assunta dalle Funzioni aziendali competenti sentito il parere delle Funzioni aziendali di Controllo interessate.

In caso di modifiche sostanziali o soggette a notifica *ex ante* si prevedono le seguenti fasi:

- A) la documentazione metodologica viene innanzitutto sottoposta alla funzione di revisione interna e alla funzione di convalida;
- B) viene reso disponibile il codice alle 2 funzioni sopra citate;
- C) una volta che le modifiche siano state validate internamente viene richiesta l'autorizzazione all'Autorità di Vigilanza;
- D) successivamente all'autorizzazione da parte dell'Autorità di Vigilanza le modifiche vengono implementate nel codice del modello: con la nuova versione viene effettuata una esecuzione in prova su ASIA e, mediante un messaggio di notifica, il Servizio Validazione e Monitoraggio viene invitato a verificare il passaggio del nuovo codice;
- E) una volta ricevuta l'approvazione del passaggio in produzione dal SVM, la nuova versione viene eseguita in ASIA ed il passaggio in produzione viene dichiarato effettivo e completo;
- F) viene aggiornato il Manuale di Misurazione e ne viene chiesta l'approvazione al Servizio Validazione e Monitoraggio tramite mail.

Nel caso di modifiche non sostanziali e non soggette a notifica *ex ante*, come ad esempio miglioramenti implementativi del codice che non alterino la conformità alla documentazione metodologica, vengono eseguiti solo i seguenti passi:

- A) la modifica viene descritta e sottoposta all'esame della funzione di convalida che, solo in casi particolari, prevedrà l'emanazione di un'Istruttoria di Convalida per l'autorizzazione al passaggio in produzione;
- B) viene reso disponibile il codice alla funzione sopra citata che, dopo la presa visione, è disponibile per il passaggio in produzione.

3.6 Calendario del processo di calcolo del requisito patrimoniale e produzione dei flussi di segnalazione di vigilanza.

I dati necessari per il calcolo del requisito patrimoniale AMA e BIA, individuale e consolidato, vengono acquisiti:

- entro il decimo giorno lavorativo del mese per la trimestrale di inizio anno;
- entro il sesto giorno lavorativo del mese per le trimestrali in corso di anno.

Nel caso in cui alcune fonti della Loss Data Collection non chiudano le elaborazioni del mese precedente entro la data di acquisizione concordata, le eventuali variazioni verranno comunque prese in considerazione successivamente, segnalando che le elaborazioni sono state effettuate con i dati disponibili alla data sopra riportata.

Il calcolo del requisito e l'invio dei relativi flussi informativi alla Funzione Segnalazione di Vigilanza vengono effettuati entro i tre giorni lavorativi successivi all'acquisizione dei dati:

- entro il tredicesimo giorno lavorativo del mese per la trimestrale di inizio anno;
- entro il nono giorno lavorativo del mese per le trimestrali in corso di anno.

Il calendario viene sottoposto a revisione al termine di ogni anno, per verificarne la compatibilità con le esigenze di acquisizione dei dati di requisito da parte della Funzione Segnalazione di Vigilanza.

4 Loss Distribution Approach

4.1 Analisi dati storici

In questo capitolo si descrive la metodologia per il calcolo del VaR sui dati storici (interni e di sistema).

La metodologia utilizzata dal Gruppo Montepaschi prevede l'aggregazione dei dati di perdita sulle 7 classi di Event Type di Basilea 2.

Tutte le elaborazioni sono implementate dal Servizio Rischi Operativi attraverso il modulo di analisi quantitativa della suite Op Risk Evolution di List e sull'ambiente ASIA sviluppato internamente e che utilizza il software statistico "R".

L'analisi dei dati storici si compone dei seguenti step:

- Costruzione dei data set per l'analisi quantitativa;
- Explanatory Data Analysis;
- Individuazione della soglia ai fini del calcolo del VaR con metodologia EVT;
- Fitting della coda della distribuzione di impatto (GPD con scaling);
- Fitting del corpo della distribuzione di impatto (distribuzione empirica);
- Fitting della distribuzione delle frequenze (Poisson);
- Calcolo dei VaR attraverso convoluzione delle distribuzioni di severity (corpo e coda) e frequenza.

Viene eseguita con cadenza semestrale un'analisi sullo stato di granularità del modello per identificare eventuali ulteriori classi di rischio, che garantiscano una migliore omogeneità dei dati e quindi una maggiore efficienza del modello.

4.1.1 Data set di Analisi

4.1.1.1 Costruzione del Dataset per le analisi di frequency

L'attuale metodologia prevede di utilizzare i soli dati interni di perdita per la determinazione della distribuzione di frequency. Per ogni classe di rischio identificata si considera solo il set di dati costituito da tutti gli eventi di perdita interni della classe analizzata al quale si applicano i seguenti filtri:

- Nel caso di classi di rischio per le quali siano presenti sia dati interni sia di sistema, l'importo di perdita deve essere maggiore od uguale a 5000 €. Nel caso di classi di rischio contenenti solo dati interni la soglia è pari a 0€.

- **Data di rilevazione minima = 10 anni dalla data di riferimento dell'analisi.**

L'ipotesi di considerare come soglia limite per l'importo 5000 € deriva dalla necessità di congruenza con i dati DIPO, raccolti sopra i 5000 €. Come tipologia di perdita si è scelto di utilizzare la Perdita Totale lorda (PTL) per essere conservativi nella stima della frequenza. Infatti la PTL è data dalla somma di Perdita effettiva lorda (PEL), accantonamenti e perdite stimate. In questo modo, per determinare la frequenza media si considerano accaduti anche quegli eventi la cui perdita è solo stimata, anche se non ha ancora avuto una sua realizzazione economica, cioè non è stata contabilizzata.

La normativa prevede di utilizzare la perdita al lordo dei recuperi assicurativi e derivanti da altri meccanismi di trasferimento del rischio. Pertanto sarebbe possibile dedurre, per il calcolo del requisito, tutti i recuperi non assicurativi (intragruppo, altro tipo e da esterno). Tuttavia al momento per ragioni di prudenzialità la perdita utilizzata è la Perdita Totale Lorda, ad ogni modo il Gruppo Montepaschi si riserva la possibilità di applicare questa metodologia in futuro.

La profondità storica di analisi considerata è di 10 anni per i dati interni.

Prima di poter eseguire il fitting per la distribuzione di frequency è opportuno verificare che gli elementi dei dataset così costituiti rispondano effettivamente alle ipotesi alla base delle stime, e cioè che siano elementi di un campione i.i.d (identicamente ed indipendentemente distribuiti).

Il primo elemento per la verifica dell'assunzione di dati identicamente distribuiti è la consistenza dei dataset creati in termini di ET. A questo fine la qualità del processo di raccolta dei dati di perdita interni garantisce un'applicazione uniforme per tutte le società delle definizioni e dei criteri adottati a livello di gruppo: l'identificazione dei singoli eventi di perdita ed il successivo mapping rispetto al modello degli Event Type rispondono a metodologie ed a modalità operative uniformemente adottate da tutte le società del gruppo.

L'indipendenza dei singoli dati è invece un'ipotesi più difficile da garantire, soprattutto per i database interni, in quanto gli eventi di perdita di un'unica realtà bancaria tendono ad essere auto-correlate dipendendo dalle evoluzioni del contesto operativo interno. Questo fenomeno è mitigato dalle diverse società del Gruppo che partecipano alla raccolta di dati: la diversificazione della fonte certamente riduce la dipendenza fra i dati generata dai fenomeni interni alle singole società, anche se non attenua quella da problematiche di contesto operativo a livello di Gruppo.

4.1.1.2 Costruzione del Dataset per le analisi di severity

Per il fitting della distribuzione di severity le scelte metodologiche prevedono di utilizzare le informazioni delle altre banche provenienti dal dataset DIPO integrandole con i dati di perdita del Gruppo Montepaschi **mediante un processo di scaling delle rispettive distribuzioni di coda**. Si considerano i record del flusso di ritorno FEG del dataset DIPO che hanno un valore di PEL (perdite effettive lorde) maggiori di 5000 €. La profondità storica di analisi considerata è di 10 anni per i dati interni, mentre per i dati di Sistema si utilizza l'intera serie a disposizione, a partire dal 2003.

La diversità di trattamento tra la serie storica basata sui dati interni e quella basata sui dati esterni (DIPO), dal punto di vista della finestra di riferimento, è legata al fatto che i dati di Sistema,

generalmente, rappresentano gli estremi delle distribuzioni e uno shift della finestra temporale comporterebbe senz'altro instabilità nelle valutazioni. Si tratta infatti di perdite di importo elevato che popolano quella zona delle distribuzioni che è più difficile da valutare e che è caratterizzata da pochi dati. Il database DIPO è inoltre in continuo riempimento sia perché migliora la raccolta dati a livello di Sistema, sia per i ritardi tra l'accadimento e la rilevazione di un evento.

Per questi motivi precauzionalmente si è deciso di mantenere invariata la data di inizio della base dati esterna, almeno fino a quando non si avranno sufficienti elementi per rivedere tale decisione.

Per la severity l'assunzione di dati identicamente distribuiti proviene dall'idea che le banche alimentanti il database DIPO, per la maggior parte operanti prevalentemente sul territorio italiano, hanno caratteristiche non troppo dissimili e, di fatto, *non sono distinguibili rispetto agli importi di perdita operativa ad i quali sono potenzialmente esposte* (Moscadelli, 2004) ⁷.

La molteplicità delle banche alimentanti garantisce inoltre l'ipotesi di campioni indipendenti, dato che normalmente non c'è correlazione fra le perdite registrate da banche diverse⁸. L'utilizzo di un set di dati di un consorzio di banche raccolto in un arco temporale mitiga la possibilità di fenomeni di *non-stazionarietà*⁹ dei dati, aumentando l'indipendenza dei singoli campioni estratti in termini di importo di perdita.

4.1.2 Exploratory data analysis

Gli obiettivi di una fase di pre-analisi dei dati quale l'Exploratory Data Analysis (EDA) sono principalmente:

- Valutare la sussistenza delle ipotesi di base sulle quali si poggiano le successive analisi di inferenza.
- Eseguire una prima semplice analisi che permetta una selezione più efficace delle tecniche e degli strumenti da adottare per le successive analisi.

Per il primo obiettivo la metodologia definita prevede l'utilizzo degli strumenti:

- Il "plot di auto-correlazione".
- Il "plot of records development".
- Il test di Durbin Watson.
- Il test di Ljung Box.

⁷ Questa assunzione non vale invece per la "ripetitività" delle perdite, sia in assoluto sia rispetto a perdite di particolari importi. Questo giustifica l'adozione per la frequency del solo set di dati interni.

⁸ Un'eccezione può essere il caso di perdite causate da eventi esterni, quali calamità naturali. In questo caso sarebbe normale trovare cluster di grosse perdite concentrate in un periodo di tempo ristretto e quindi correlate in una certa misura.

⁹ In un processo *stazionario* in senso forte le probabilità congiunte delle variabili casuali sono insensibili ad una qualsiasi traslazione dell'origine dei tempi, ossia non dipendono da quest'ultima ma solo dalla loro posizione relativa (distanza temporale).

- Il test di Dickey Fuller aumentato.

I test vengono applicati ai data set sia di severity sia di frequency. Per la pre-analisi dei dati la metodologia definita si avvale di vari strumenti quali:

- Trend Chart delle perdite.
- Time series plot sia per le frequenze sia per le severity.
- Distribuzione cumulata ed istogramma empirico delle perdite.
- Box Plot per le severity
- Q-Q plot (contro un set di varie distribuzioni teoriche)

La visione del *Trend Chart* e dei *time series plot* permette di avere una prima evidenza dell'andamento delle perdite in funzione del tempo. Con il *time series plot* della severity è possibile identificare i valori estremi di perdita e percepire in prima approssimazione la loro frequenza di accadimento. Il *time series* della frequency invece evidenzia l'andamento delle frequenze giornaliere/settimanali/mensili ed i periodi dove si sono registrate un numero maggiore di perdite. I *time series plot*, inoltre, consentono di evidenziare eventuali clusters di perdite che potrebbero indebolire l'assunzione di dati i.i.d.. Il *Trend chart*, infine, riassume le informazioni dei due singoli *time series plot* evidenziando i periodi con le maggiori perdite aggregate giornaliere/settimanali/mensili.

Le distribuzioni cumulate empiriche, l'istogramma empirico ed il Box plot permettono di avere una prima idea sulle caratteristiche principali delle distribuzioni di severity e di frequency dei dati. Normalmente le distribuzioni empiriche costruite sui dati dei rischi operativi mostrano una marcata asimmetria destra e code molto pesanti.

Per riassumere le caratteristiche di ogni set di dati si utilizzano alcuni indici di statistica descrittiva, quali media, mediana, primo e terzo quartile (25° e 75° percentile della distribuzione) e gli indici di asimmetria (momento terzo della distribuzione fornisce un'idea della forma) e di curtosi (momento quarto, fornisce una misura della pesantezza delle code).

4.1.3 Distribuzione di frequency

I dati sulla frequency sono fittati con una Poisson con parametro pari alla media degli eventi registrati nell'arco di tempo considerato. La semplicità di tale fitting fornisce un buon risultato e, in generale, non si ritiene necessario il ricorso ad altre distribuzioni, quali la binomiale negativa o misture di Poisson.

La tipicità dei dati di perdita interni del Gruppo Montepaschi, la dinamicità del Database delle perdite ed i ritardi nella rilevazione di determinati eventi evidenziano un significativo gap temporale tra l'effettiva data di accadimento e quella di rilevazione, con una conseguente mancanza di dati per date di accadimento recenti. Per tale incompletezza è alto il rischio di una sottostima della frequenza media degli accadimenti prendendo a riferimento l'intero set di dati dell'arco temporale considerato.

Questo fenomeno è evidente dall'analisi di un tipico time series plot basato sulla data di accadimento, che evidenzia la diminuzione con il passare del tempo della frequenza media:

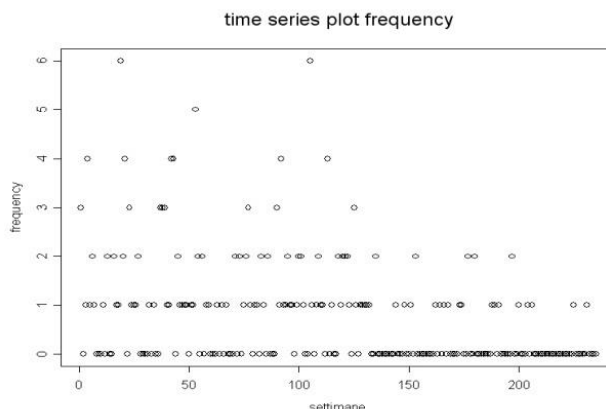


Figura 4 Time series plot della frequenza di accadimento

Questo problema non sussiste utilizzando la data di rilevazione della perdita.

4.1.4 Distribuzione di severity

La distribuzione della severity degli eventi di perdita stima la densità di probabilità degli importi di ciascuna perdita osservata. Può essere stimata secondo due modalità:

- **Analisi Empirica**: la distribuzione empirica costituita dagli importi di perdita registrati dalla LDC si considera quale la migliore distribuzione per la stima della severity.
- **Analisi Parametrica**: Il metodo consiste nel “fittare” la distribuzione con un’opportuna densità di probabilità (pdf). La pdf può essere sia una distribuzione parametrica singola che una somma di più distribuzioni, ciascuna pesata da un opportuno fattore. Date le caratteristiche peculiari evidenziate dei dati di perdita operativi, asimmetria e curtosi, il “fit” può avvenire in maniera separata per il *corpo* della distribuzione e per la *coda*, sulla quale si può applicare la teoria dei valori estremi (EVT).

La scelta fra i due modelli dipende dalla numerosità e dalla qualità dei dati su cui viene svolta l’analisi. La scelta di utilizzare l’intero set di dati DIPO a disposizione rende l’analisi empirica una buona possibilità per la stima della severity, ma va considerato quale fattore decisivo il fatto che un’analisi puramente empirica implica il porre un limite superiore (il massimo delle perdite registrate) all’entità dell’evento di perdita. Questa approssimazione non è tanto sostenibile per le perdite operative, per le quali il potenziale di perdita massimo non è limitato dall’esperienza storica. Di conseguenza la metodologia del Gruppo Montepaschi prevede un approccio misto per lo studio della severity:

- L’utilizzo di una distribuzione empirica per studiare il corpo della distribuzione;
- L’utilizzo delle distribuzioni parametriche previste dalle tecniche di EVT per lo studio della coda della distribuzione;

La metodologia basata sull'EVT consente di ottenere una rappresentazione parametrica della coda della distribuzione delle perdite, senza la necessità di avanzare ipotesi sulla forma funzionale dell'intera "distribuzione originaria" delle suddette perdite.

Si noti che la EVT non è assimilabile ad una mistura né ad un fit parametrico, in quanto non copre l'intero dominio della distribuzione empirica, ma solo la regione dei valori estremi. I metodi di stima sono comunque riconducibili a quelli normalmente utilizzati nei fit.

I due più importanti risultati (di natura asintotica) dell'EVT si esplicitano in due logiche differenti di selezione dei valori estremi assunti da una variabile casuale: la *Logica Block Maxima* e la *Logica Peaks Over Threshold* (POT)¹⁰.

In entrambi i casi l'importante conclusione a cui si perviene è la conoscenza della forma distributiva (asintotica) della coda della distribuzione campionaria, senza la necessità di fare alcuna supposizione sulla distribuzione delle perdite osservate. Ciò a cui si è interessati è la stima dei parametri caratterizzanti la suddetta forma distributiva asintotica, che nella logica Block Maxima è una Generalized Extreme Value distribution (GEV), mentre nella logica POT è una Generalized Pareto Distribution (GPD).

L'approccio POT prevede i seguenti passi:

1. Stima della soglia per la successiva applicazione della EVT;
2. Applicazione dell'EVT per la stima della distribuzione della coda;
3. Utilizzo della distribuzione empirica per la stima del corpo.
4. Definizione della distribuzione di severità come giunzione ("splicing") delle distribuzioni del corpo e della coda.

I diagnostici e metodi utilizzati nell'approccio POT per la determinazione della soglia EVT e la stima dei parametri della distribuzione GPD sono descritti nelle sezioni 4.1.4.2, 4.1.4.3, 4.1.4.4.

Per la modellizzazione della severità sopra la soglia EVT, la metodologia definita dal Gruppo Montepaschi prevede due casi:

- classi interne di rischio per le quali vengono utilizzati i soli dati interni per modellizzare la severità (vedasi sezione 3.3.3);
- classi interne di rischio per le quali vengono utilizzati dati interni ed esterni per modellizzare la severità.

Nel primo caso, viene usato l'approccio POT sopra descritto. Nel secondo caso viene usato un approccio POT che include un riscalamento ("scaling"): in particolare, dopo aver fissato la soglia EVT (secondo i criteri specificati in sezione 4.1.4.5), la distribuzione di coda viene ottenuta come media ponderata di due distribuzioni GPD, stimate separatamente sui dati di coda interni ed esterni (DIPO). I dettagli sono riportati nella sezione 4.1.4.1. In entrambi i casi, per la modellizzazione del corpo si utilizza una distribuzione empirica, vedasi sezione 4.1.4.6.

¹⁰ Si veda il capitolo 5 per gli approfondimenti teorici sulle due logiche di analisi EVT.

4.1.4.1 Descrizione della metodologia di scaling dei dati esterni di coda

La metodologia definita dal Gruppo Montepaschi prevede due casi:

1. classi interne di rischio per le quali la severità viene stimata sui soli dati interni (vedasi sezione 3.3.3);
2. classi interne di rischio per le quali la severità viene stimata sui dati interni ed esterni.

Nel primo caso, dopo aver identificato la soglia EVT che separa corpo e coda dei dati, una singola distribuzione GPD viene stimata a partire dai dati di coda. Nel secondo caso, viene applicata una metodologia di riscaldamento: si modellizzano separatamente le distribuzioni G_I dei dati interni e G_E dei dati esterni di coda, entrambe delle GPD con medesima soglia EVT, per poi utilizzare, come distribuzione di coda, una media ponderata delle due distribuzioni così ottenute. In particolare, dato un peso w compreso tra 0 ed 1, la distribuzione di coda G viene definita da:

$$G^{-1}(p) = wG_I^{-1}(p) + (1 - w)G_E^{-1}(p)$$

per tutti i livelli di probabilità p nell'intervallo $[0; 1]$. In altre parole, si effettua una pesatura delle funzioni percentile G_I^{-1} e G_E^{-1} sull'asse delle probabilità p , con un peso w da determinare. Effettuare la pesatura sulle funzioni percentile garantisce che i percentili della funzione G siano compresi tra quelli delle funzioni G_I e G_E .

In questo approccio, il problema della scelta della soglia EVT viene affrontato con i metodi descritti nelle sezioni precedenti: i diagnostici EVT utilizzati per la scelta della soglia (grafico di ξ al variare della soglia, test di bontà dell'adattamento al variare della soglia, etc.) vengono separatamente prodotti per le due distribuzioni G_I e G_E . Il collocamento della soglia viene deciso principalmente sulla base dei diagnostici per la distribuzione G_I dei dati interni: si veda la sezione 4.1.4.5 per i criteri adottati.

Il calcolo del peso w si basa sulla combinazione di due fattori:

1. numerosità dei dati interni sui quali stimare G_I ;
2. dissimilarità tra le distribuzioni dei dati interni G_I ed esterni G_E sopra la soglia EVT.

Consideriamo innanzitutto i due fattori separatamente. Riguardo al punto 1, denotiamo con n la numerosità degli eccessi interni sui quali stimare la G_I ; si assume che il peso w dei dati interni debba tendere verso 1 quando la numerosità n tende ad infinito. Questa proprietà viene modellizzata imponendo che il peso sia una funzione del tipo:

$$w_1 = 1 - \exp(-n/n_c) \quad \text{se } n \geq n_0 \quad (2)$$

In questa formula:

- $n_0 = 20$ è la numerosità minima degli eccessi interni, richiesta per poter effettuare la stima della distribuzione G_I ;
- $n_c = 200$ è una numerosità richiesta per ottenere una stima ragionevolmente affidabile; il valore $n_c = 200$ è stato determinato considerando le numerosità medie degli eccessi sopra soglia nella stima di requisito patrimoniale a marzo 2015.

La Figura 5 mostra il grafico del peso w_1 al variare di n .

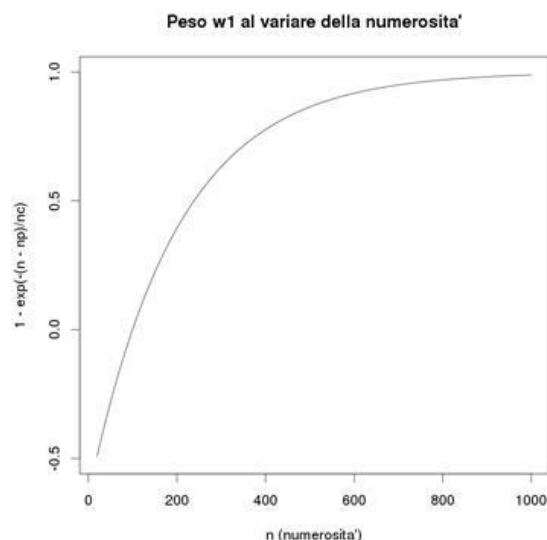


Figura 5 peso w_1 al variare della numerosità n , con $n_c = 200$ e $n_0 = 20$.

Riguardo al secondo fattore, ovvero la dissimilarità tra le distribuzioni, si utilizza un peso con la seguente forma funzionale:

$$w_2 = 1 - \exp(-f(K))$$

In questo caso K è la statistica del test di Kolmogorov-Smirnov per l'identità distribuzionale tra due campioni di dati: viene usata per misurare il grado di dissimilarità tra le distribuzioni G_I e G_E . La funzione f è una trasformazione di K che ha la seguente forma:

$$f(K) = -\log(0.5) + \frac{b}{\sqrt{1-K}} - b,$$

dove $b = 3$. L'idea alla base di questa trasformata è che le due distribuzioni G_I e G_E abbiano peso uguale (pari a 0.5) se la dissimilarità è bassa (K vicina a zero), mentre il peso dei dati interni aumenti nel caso di dissimilarità elevata (K vicina ad 1). Si veda Figura 6 per un esempio di grafico del peso w_2 al variare di K .

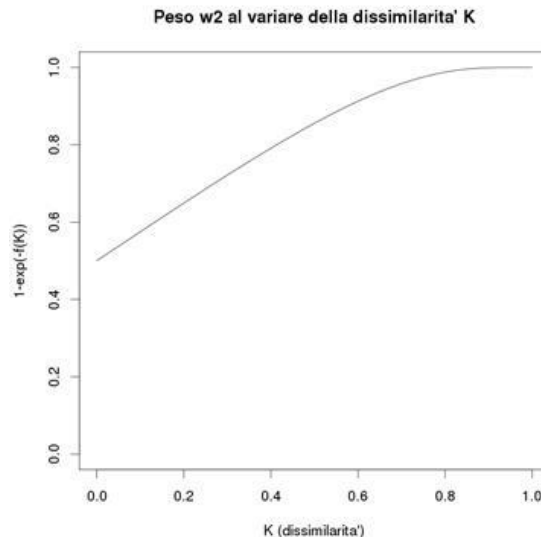


Figura 6 peso w_2 al variare della statistica K di Kolmogorov-Smirnov.

La scelta del parametro $b = 3$ è stata effettuata per conferire alla funzione $w_2(K)$ un andamento concavo al tendere di K ad 1. Si definisce infine il peso w da utilizzare nel modello come una combinazione tra i due effetti sopra descritti:

$$w = \max \left(0.5, 1 - \exp \left(- \left(\frac{n}{n_c} + f(K) \right) \right) \right) \quad \text{se } n \geq n_0 \quad (5)$$

Con questa formula:

- la distribuzione dei dati interni acquisisce un peso w pari ad almeno 0.5;
- il valore del peso tiene presente sia la numerosità dei dati interni, sia la dissimilarità rispetto ai dati esterni; in particolare, il peso dei dati interni aumenta sia in caso di pronunciata dissimilarità dei dati interni ed esterni, sia in presenza di elevata numerosità dei dati interni.

Il metodo di *scaling* non si applica nel caso in cui i dati interni presentino una ridotta numerosità complessiva o dei soli eccessi. In particolare, si stima una singola distribuzione di severità, fondendo i dati interni ed esterni (e non vengono calcolate due distribuzioni di coda separate per i dati interni ed esterni) nel caso in cui:

- la numerosità complessiva dei dati interni è inferiore a 200, oppure
- non si riescono ad individuare soglie EVT idonee con almeno $n \geq n_0 = 20$ eccessi interni.

4.1.4.2 Determinazione della soglia per l'applicazione della EVT.

L'obiettivo dell'analisi è stimare un livello di soglia per separare il corpo dalla coda della distribuzione. Per le scelte metodologiche fatte, la soglia individua l'importo fino al quale la stima della severity si avvale della distribuzione empirica ed oltre il quale si applicano per il fit le

tecniche di EVT (P. Embrechts, C. Kluppelberg, T. Mikosch, 1997 e J. Beirlant, Y. Goetgebeur, J. Segers e J. Teugels, 2004).

La stima si avvale delle informazioni qualitative di alcuni strumenti grafici di Exploratory Data Analysis integrandole con le indicazioni quantitative, che derivano dai metodi di stima dei parametri utilizzati e dai test di goodness of fit applicati.

I limiti impliciti nell'analisi qualitativa non consentono di individuare un valore univoco della soglia, ma, sulla base dei dati disponibili, permettono di definire un range di possibili valori all'interno del quale scegliere il posizionamento ritenuto più appropriato.

Il valore, all'interno del range individuato, deve necessariamente garantire un valore elevato della funzione di ripartizione (così come previsto dai teoremi dell'EVT) e isolare alla sua destra un numero sufficiente di dati su cui poter effettuare una stima robusta e stabile.

4.1.4.2.1 Pre-analisi grafica: comportamento a code pesanti ed applicabilità dell'EVT.

4.1.4.2.1.1 Mean Excess Plot

Data la distribuzione di severity, il *Mean Excess* (ME, media degli eccessi) per una certa soglia “u” è definito come il valore atteso degli eccessi oltre la soglia, condizionato al verificarsi di perdite superiori alla soglia stessa. La funzione per il calcolo del ME è la seguente:

$$e(u) = E(X - u | X > u) = \frac{1}{1 - F_X(u)} \cdot \int_u^{\infty} (x - u) \cdot f_X(x) dx$$

Equazione 1

Il Mean Excess Plot (MEP), cioè il grafico del valore del ME al variare della soglia, è in primo luogo utile per determinare la grossezza delle code e quindi per valutare la necessità di utilizzare le tecniche di EVT per il set di dati analizzato.

Risultati teorici affermano che la mean excess function è costante per un'esponenziale, mentre è lineare per una Generalized Pareto Distribution (GPD): se la distribuzione ha code con peso elevato, la pendenza è positiva, altrimenti è negativa.

Se per il set di dati a disposizione si riscontra una pendenza positiva e crescente all'aumentare della soglia si desume la presenza di code grosse: la pendenza positiva implica infatti che al diminuire del numero di osservazioni nella coda (e dunque all'aumentare della soglia), non diminuisce la distanza media dalla nuova soglia degli eccessi valutati sulle perdite residue. In questo caso si verificano quindi le condizioni di applicabilità dell'analisi EVT.

Al contrario, se il ME plot avesse una pendenza decrescente all'aumentare della soglia, si desumerebbe la presenza di una coda destra leggera che potrebbe essere fittata senza dover ricorrere all'analisi EVT.

La seguente figura rappresenta un tipico MEP per dati OpR:

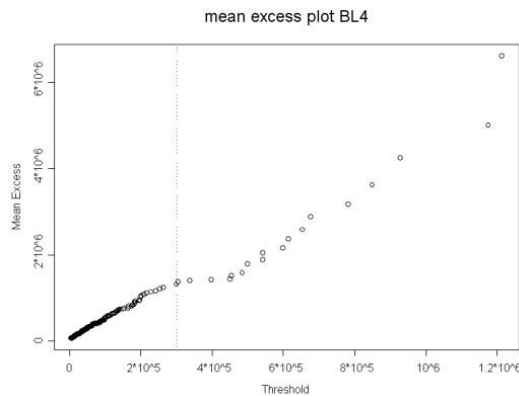


Figura 7 Mean Excess Plot

L'analisi del grafico, oltre a segnalare la presenza di una coda pesante, aiuta ad identificare possibili valori sui quali posizionare la soglia: la presenza di eventuali discontinuità nel grafico, infatti, può aiutare nella selezione dei possibili valori soglia per l'applicazione delle tecniche di EVT.

In ogni caso, le informazioni emerse dall'analisi del grafico vanno confrontate con quanto si rileva dalle successive analisi condotte con gli strumenti di seguito descritti.

4.1.4.2.2 Stima del range dei valori per il parametro di shape.

L'utilizzo dei grafici della funzione ME permettono di selezionare un set di possibili soglie, qualificando la coda in termini di numeri di eccessi presi dal campione che la formano. Il valore della soglia corrispondente in termini monetari è dato dall'importo campionario più grande che lascia alla sua destra il numero selezionato di eccessi.

La metodologia MPS prevede, una volta identificati attraverso le analisi grafiche i possibili valori "candidati soglia", l'utilizzo di strumenti quantitativi per la selezione di una sola fra le possibili soglie e la stima dei parametri "ξ" e "β" della GPD.

4.1.4.2.2.1 Rapporto Massimo su Somma

Il Rapporto Massimo su Somma (RMS) permette di analizzare in maniera qualitativa il valore dei momenti della distribuzione empirica, fornendo indirettamente indicazioni sui possibili valori del parametro "ξ" della distribuzione GPD con la quale si fitta la coda.

Consideriamo n variabili aleatorie iid X_1, \dots, X_n e definiamo $S_n(p) = |X_1|^p + \dots + |X_n|^p$ e $M_n(p) = \max(|X_1|^p, \dots, |X_n|^p)$.

Si dimostra che se il valore del rapporto massimo su somma, $R_n(p) = \frac{M_n(p)}{S_n(p)}$, valutato sul campione analizzato per una data potenza p -esima, converge quasi certamente a zero al crescere della numerosità del campione stesso, allora esiste ed è definito il momento assoluto p -esimo della distribuzione che si sta ricercando.

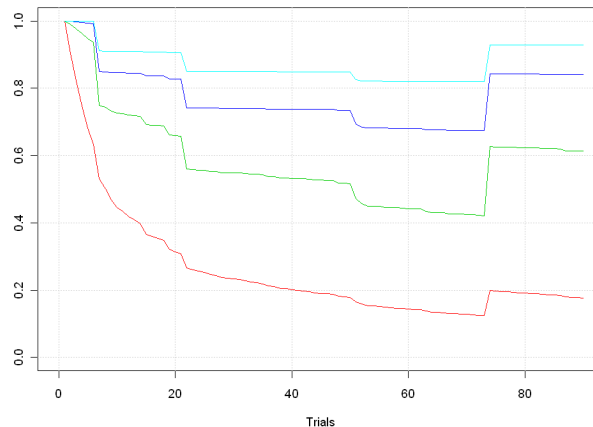


Figura 8 Plot rapporto Massimo su somma

Le quattro linee dal basso verso l'alto rappresentano l'andamento di $R_n(p)$ al crescere del campione per $p=1,2,3,4$. L'asse delle x rappresenta l'aumentare della numerosità del campione considerato, al crescere del numero di eventi di perdita raccolti. I grafici partono tutti da 1, dato che per campioni formati da 1 solo elemento il massimo coincide con la somma stessa. Il rapporto tende a decrescere all'aumentare del campione fino alla rilevazione di un nuovo massimo (un nuovo record per la serie) che spinge verso l'alto il valore di $R_n(p)$. Se dall'analisi del grafico si desume che il valore del rapporto per $p=1$ tende a 0 al crescere della numerosità del campione allora esiste il momento primo della GPD, e si è quindi certi che il parametro “ ξ ” è minore in valore assoluto ad 1. Se si desume che anche per $p=2$ tende a 0, allora esiste la varianza di X e $|\xi| \leq 1/2$. In generale se esiste il momento p-esimo, allora il valore di “ ξ ” è minore di $|1/p|$.

Dall'analisi del grafico riportato si evince che sicuramente il campione di eccessi considerato ha un momento primo finito, per cui la stima della ξ deve essere inferiore a 1. Il momento secondo, invece, non sembra tendere a 0 con certezza, indicando, di conseguenza un momento secondo infinito ed un intervallo per la stima di ξ compreso fra 0,5 e 1.

4.1.4.2.2.2 Hill Plot

L'Hill Plot (HP) riporta l'andamento dello stimatore di Hill del parametro “ ξ ” della GPD al variare del numero di osservazioni del campione fornendo un'indicazione del peso della coda della distribuzione. Infatti, se il grafico ha un andamento crescente, la coda è più pesante di quella di un'esponenziale, mentre se è decrescente la coda è più leggera.

All'aumentare delle osservazioni (o al diminuire della soglia), tale parametro tende a stabilizzarsi. Per quanto possibile, è consigliato scegliere il valore della soglia non prima che il parametro “ ξ ” sia stabile; in caso contrario, la scelta della soglia potrebbe influenzare enormemente il valore del parametro al variare del numero di osservazioni.

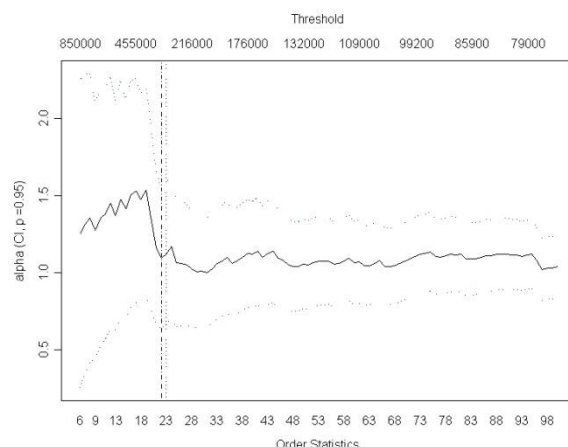


Figura 9 Hill plot del parametro di shape

Per il medesimo campione di dati il parametro “ ξ ” della GPD può essere stimato usando i metodi¹¹ :

- Massima Verosimiglianza (Maximum Likelihood Estimation, MLE);
- Probability Weighted Moments (PWM);
- Massima Verosimiglianza Penalizzata (Maximum Penalized Likelihood Estimation, MPLE);
- Maximum Goodness of Fit basato sulla statistica Anderson-Darling quadratica (MGF-AD2);
- Minimum Density Power Divergence (MDPD).

4.1.4.2.2.3 Confronto tra i vari stimatori

La scelta del metodo di stima da utilizzare per i parametri di shape, “ ξ ”, e di scale, “ β ”, della GPD è stato oggetto di particolare approfondimento, in quanto l’applicazione allo stesso set di dati dei vari metodi può risultare in stime di “ ξ ” e di “ β ” sensibilmente diverse. In particolare la stima del valore dello ξ è importante data l’elevata sensibilità della stima finale di VaR aggregato al suo valore stimato. Gli aspetti di cui tenere conto sono sostanzialmente i seguenti:

- Numerosità del campione di dati a disposizione: secondo quanto dimostrato da uno studio condotto da Coles e Dixon (referenza [18] nella bibliografia), e verificato con test realizzati dal gruppo di lavoro interno, la stima del parametro “ ξ ” con i metodi MPLE e PWM risulta più vicina a quella reale nel caso di campioni di modeste dimensioni. Per campioni sotto le 500 osservazioni la stima di tale parametro tramite MPLE e PWM risulta costantemente più accurata, per campioni di dimensioni maggiori, il divario con il metodo MLE cessa di esistere. Le medesime considerazioni valgono anche per il parametro “ β ”. Inoltre, Coles e Dixon affermano che il metodo MPLE offre comunque i

¹¹ Si veda il capitolo 5 per gli approfondimenti teorici sui metodi di stima dei parametri.

vantaggi di un approccio a massima verosimiglianza, come la possibilità di introdurre covariate.

- Valore dello ξ ricercato: lo stimatore PWM si dimostra sempre superiore a quello ML per valori dello ξ da stimare inferiori a 0,5. Al tendere del valore di ξ ad 1, invece la velocità di convergenza delle stime PWM diminuisce dato che la GPD non ha un momento secondo finito. Lo stimatore MPLE introduce una penalizzazione che modifica la verosimiglianza di una quantità asintoticamente trascurabile; pertanto, tutti i risultati asintotici standard che valgono per lo stimatore MLE valgono anche per lo stimatore MPLE.
- Stabilità della stima al variare del campione: la continua attività di raccolta dati (sia interni sia esterni) rende disponibili nuovi set di dati sui quali eseguire le stime. L'arrivo dei nuovi dati si può tramutare in valori stimati di ξ anche sensibilmente diversi rispetto ad i periodi passati, anche a parità di soglia monetaria selezionata per l'applicazione della EVT. Una forte instabilità della stima dello ξ anche in seguito a piccole variazioni del set di dati, può risultare in stime finali di VaR inaccettabilmente volatili da un periodo di analisi all'altro. Anche da questo punto di vista lo studio di Coles e Dixon, verificato da test interni effettuati su dati empirici, dimostra un comportamento dello MPLE e del PWM molto più stabile rispetto a quello del ML. Questa è una caratteristica, desiderabile per le stime di ξ , poiché permette di ottenere una certa stabilità anche per le stime di VaR aggregate: le stime MPLE e PWM si mostrano più stabili rispetto a quelle ML non presentando improvvisi salti al variare della numerosità del campione selezionato a parità di soglia.

In base alle considerazioni sopra esposte si ritiene che le stime tramite MPLE e PWM siano da preferire a quella tramite ML, soprattutto per la fondamentale caratteristica evidenziata di stabilità del valore ottenuto. Le stime ML, se applicate, possono determinare improvvisi salti nel valore stimato della ξ e conseguentemente nel valore finale del VaR (raddoppiandolo o triplicandolo). Questo problema di forte instabilità delle stime viene riscontrato spesso anche con gli stimatori MDPD ed MGF-AD2.

A partire dal 2013, in base ad analisi approfondite e per allinearsi alle best practice di sistema il Gruppo Montepaschi ha adottato gli stimatori MPLE.

Una possibile soluzione al problema della scelta della soglia può essere fornita dal grafico sull'andamento della stima del parametro di shape della GPD al variare della soglia monetaria o del numero di eccessi. Una indicazione sulla scelta della soglia può essere fornita dai punti in cui le varie stime convergono, cercati all'interno di ampi range di stabilità.

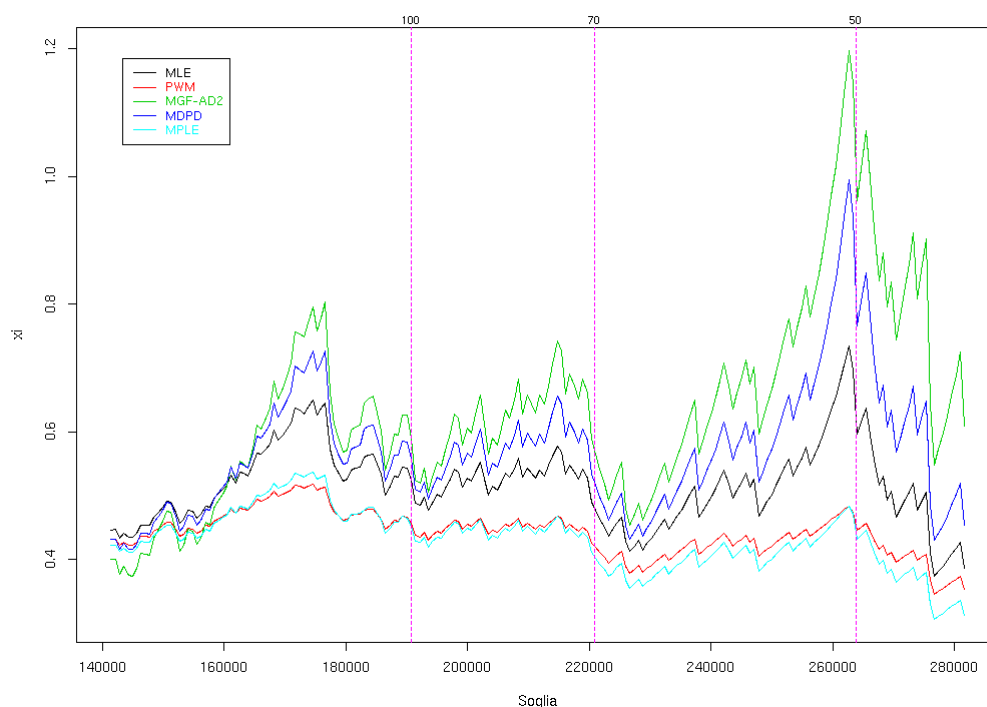


Figura 10 Grafico di confronto fra i vari stimatori – scelta della soglia

I casi estremi riguardano:

- Assenza di punti di contatto tra le curve;
- Presenza di più punti di contatto tutti ugualmente probabili.

Nel primo caso si sceglie il punto all'interno del range di stabilità in cui è minima la distanza tra le curve; nel secondo si analizza il trade-off tra numerosità campionaria e valore soglia abbastanza elevato da permettere il rispetto delle ipotesi per l'applicazione dell'EVT.

Infine, una volta definita una classe di potenziali valori di “ ξ ” è utile analizzare la loro attendibilità mediante l'utilizzo di strumenti grafici quali il q-q plot o analitici come i test di Kolmogorov-Smirnov e di Anderson-Darling che consentono di determinare quale delle soglie porta ad un fitting più accurato.

Per concludere quindi la scelta del metodo di stima deve contemplare i seguenti aspetti:

- Numerosità del campione;
- Stabilità della stima al variare della soglia;
- Stabilità della stima nel tempo (simulabile anche tramite bootstrap);
- Qualità del fitting.

4.1.4.3 Stima parametri della Generalized Pareto Distribution e intervalli di confidenza

La stima dei parametri della GPD, una volta fissata la soglia, avviene con il metodo MPLE su tutti i dati di severity oltre la soglia stessa. Questa stima viene comunque confrontata con quelle ottenute dagli altri stimatori considerati (PWM, MLE, MGF-AD2, MDPD).

Gli strumenti utilizzati¹² per stimare la compatibilità fra i dati e la distribuzione ipotizzata sono:

- Il confronto dei quantili (Q-Q plot);
- Il confronto delle probabilità cumulate;
- Il confronto dei periodi di ritorno;
- La verifica della dispersione dei residui;

4.1.4.3.1.1 Q-Q plot

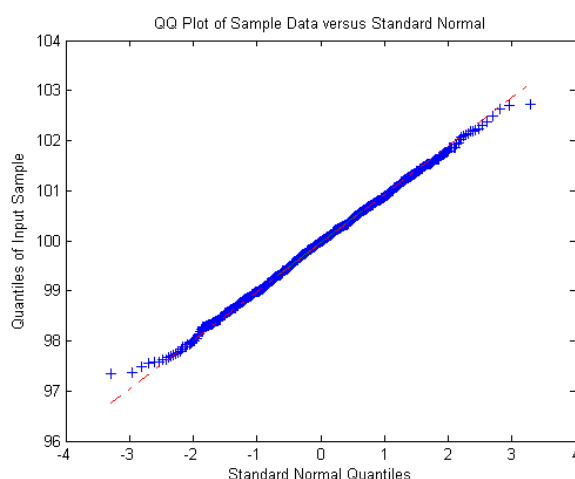


Figura 11 QQplot

Il grafico dei quantili permette di valutare la bontà dell'adattamento del modello ai dati; esso è ottenuto riportando su un piano cartesiano punti di ordinata pari ad un determinato quantile della distribuzione empirica dei dati e di ascissa pari al medesimo quantile della distribuzione parametrica ipotizzata.

In generale se il modello parametrico ipotizzato garantisce un buon fit dei dati, il Q-Q plot assume una forma lineare. In questo modo tale strumento grafico consente di valutare visivamente in modo rapido la capacità di diversi modelli parametrici di “fittare” i dati, evidenziando soprattutto la dispersione lungo le ascisse delle due distribuzioni comparate.

4.1.4.3.1.2 Probabilità cumulate

Il grafico delle probabilità cumulate permette di confrontare i dati empirici e con quelli della teorica stimata. Tale confronto può essere effettuato riportando sul grafico sia in scala log-lineare, sia in scala log-log.

¹² Si veda il Capitolo 5 per gli approfondimenti teorici sugli strumenti statistici utilizzati.

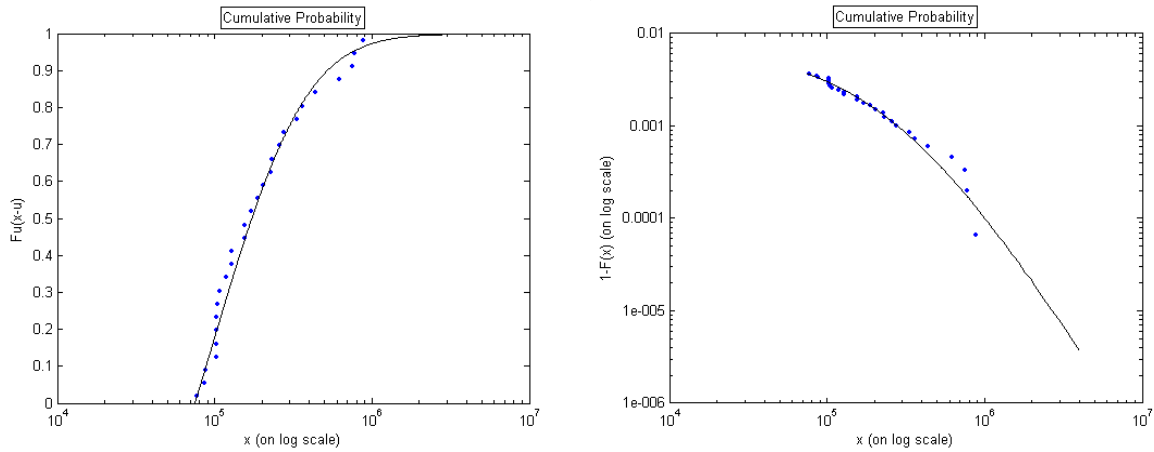


Figura 12 distribuzioni cumulate delle probabilità

4.1.4.3.1.3 Il confronto dei periodi di ritorno

Il periodo di ritorno è definito come il tempo di attesa medio tra due eventi estremi.

Sia X_i una sequenza di variabili aleatorie iid con distribuzione di probabilità F e sia u una data soglia. La sequenza $I_{\{X_i > u\}}$ descrive una successione di variabili aleatorie iid di Bernoulli con probabilità di successo $p = \bar{F}(u)$.

Il tempo di primo successo è definito come $L(u) = \min\{i \geq 1 : X_i > u\}$ ovvero si distribuisce come una variabile aleatoria geometrica con distribuzione:

$$P(L(u) = k) = (1 - p)^{k-1} p, \quad k = 1, 2, \dots$$

La variabile aleatoria

$$L_1(u) = L(u), \quad L_{n+1}(u) = \min\{i \geq L_n(u) : X_i > u\}, \quad n \geq 1$$

Descrive il periodo tra l'accadimento di due eventi successivi che superano la soglia.

Il periodo di ritorno dell'evento $\{X_i > u\}$ è definito come :

$$E(L(u)) = p^{-1} = (\bar{F}(u))^{-1}$$

che tende a ∞ per u che tende a ∞ .

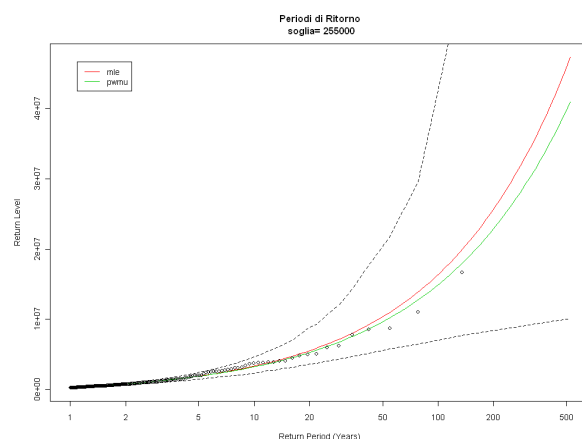


Figura 13 Grafico dei periodi di ritorno

4.1.4.3.1.4 Verifica della dispersione dei residui

Per quanto non rientri a far parte della categoria dei test, nel caso si compiano fit su funzioni continue è utile una visione dell'andamento dei residui.

Si stima la statistica di Davison, “residuo”, definita come:

$$W_i = \frac{1}{\xi} \ln \left[1 + \xi \frac{x_i - u}{\beta} \right]$$

Se gli eccessi sono iid e provengono da una GPD i corrispondenti residui dovranno essere iid e distribuiti esponenzialmente con media pari a 1 (Chavez-Demoulin_, 2004, Mc Neil, 2005) ¹³. Quindi se l'andamento medio dei residui non evidenzia curve particolari ma si mantiene piuttosto “piatto”, si ha un'indicazione ulteriore della bontà del fit. Al contrario, scostamenti “a campana” sono sintomo di un errore nella forma della distribuzione ipotizzata.

¹³ Cfr. pag 50 e pag 300 e seg.

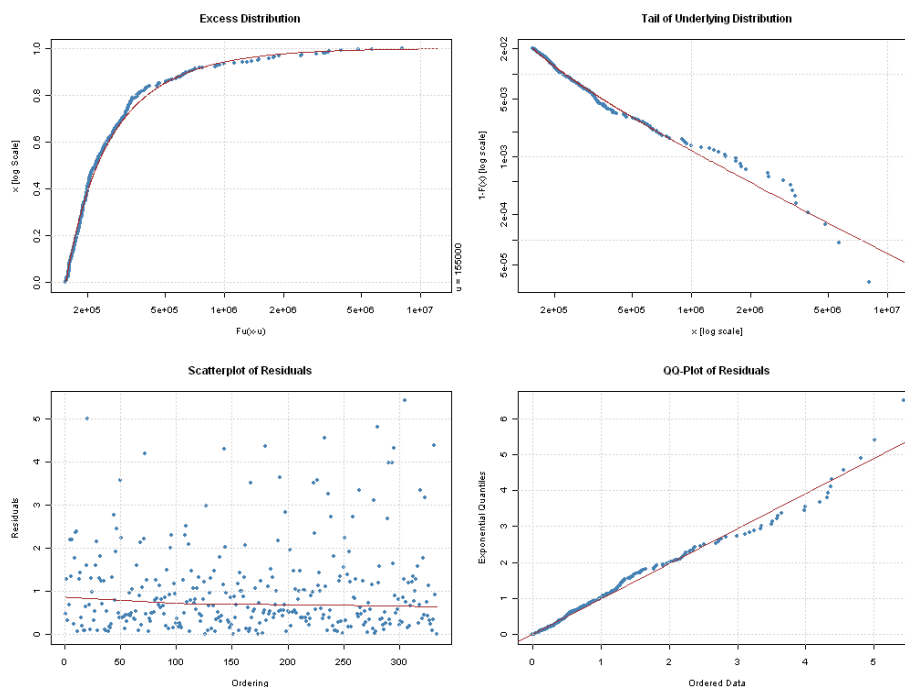


Figura 14 Distribuzione dei residui: verifica poisson process

4.1.4.3.1.5 Stabilità della stima

Per testare la stabilità della stima. Ovvero la non eccessiva dipendenza della stima dallo specifico campione di dati a disposizione si produce un box plot di dati simulati.

Utilizzando un metodo bootstrap, ovvero campionando con ripetizione n pseudo code formate dallo stesso numero di eccessi della coda reale, estratti in maniera casuale dalla distribuzione originale. Si ottengono vari campioni casuali e le relative stime di ξ .

Generalmente i metodi MPLE e PWM si dimostrano più stabili, in quanto producono oscillazioni meno ampie attorno al valor medio della stima. Inoltre, il metodo MPLE si dimostra particolarmente robusto rispetto alla presenza di perdite estreme.

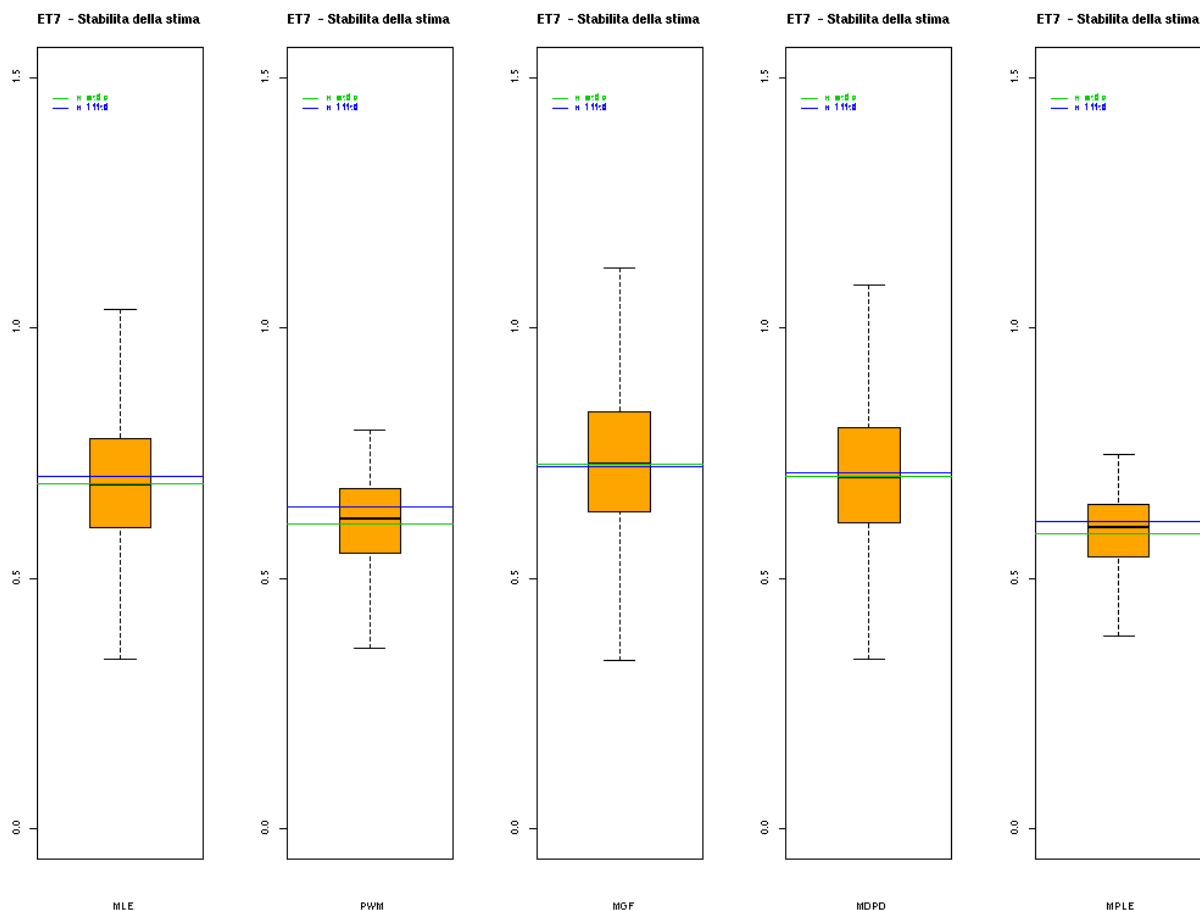


Figura 15 Boxplot distribuzione simulata del parametro di shape della GPD, metodo bootstrap.

4.1.4.3.1.6 Intervalli di confidenza delle stime dei parametri

Per ottenere un intervallo di confidenza delle stime dei parametri della GPD, si possono utilizzare due metodi:

- il metodo asintotico: deriva dall'applicazione della matrice di Fisher per la stima della varianza asintotica, si veda a tal proposito il Capitolo 5 contenente gli approfondimenti teorici.

Sfruttando il teorema asintotico che garantisce la normalità della distribuzione degli stimatori MLE, PWM ed MPLE si ottiene l'intervallo cercato a livello 95% e 99% applicando la relazione (cfr paragrafo 4.1.6):

$$P [\mu - Z_{\alpha/2} (\sigma/\sqrt{n}) \leq x \leq \mu + Z_{\alpha/2} (\sigma/\sqrt{n})] = (1 - \alpha)$$

Valori critici usuali di $Z_{\alpha/2}$ sono:

- $Z_{0.05} = 1.64$ (confidenza del 90%)
- $Z_{0.025} = 1.96$ (confidenza del 95%)

$$- Z_{0.005} = 2.57 \text{ (confidenza del 99\%)}$$

$Z_{\alpha/2} (\sigma/\sqrt{n})$ è la quantità che viene aggiunta e sottratta alla media campionaria per avere l'intervallo. Si chiama massimo errore di stima, ed è un indicatore della precisione della stima.

- il metodo numerico: nel caso in cui non sia possibile determinare la varianza asintotica a causa della singolarità della matrice di Fisher, oppure se non sono rispettate le condizioni di esistenza della stessa (csi MLE o MPLE inferiore a -0,5 oppure csi PWM superiore a 0,5), si può utilizzare il risultato ottenuto con il campionamento bootstrap descritto al paragrafo precedente.

Si calcola l'intervallo di confidenza considerando come estremi degli intervalli di confidenza i percentili delle distribuzioni di csi ottenute dal campionamento. Il 2,5% ed il 97,5% sono gli estremi dell'intervallo al 95% di confidenza, mentre lo 0,5% ed il 99,5% sono gli estremi nel caso di intervallo al 99% (l'area sottesa alla coda rappresenta l'1% dei casi estremi).

Si è osservato empiricamente che, nel caso siano disponibili entrambe le valutazioni, le differenze tra i due intervalli non sono eccessive, sebbene il bootstrap si fondi su un numero finito di simulazioni (1000).

4.1.4.4 Test di Goodness of Fit

- I Test parametrici;
- I Test non parametrici.

4.1.4.4.1 Test parametrici

I test parametrici sono quelli che si basano su una statistica descritta da una funzione densità di probabilità (pdf) dipendente da parametri, solitamente determinabili dai dati del problema. Un test parametrico attualmente utilizzato è quello del Chi-Quadro il quale emerge in maniera naturale per i problemi di fit di istogrammi. La statistica della somma degli errori quadratici del fit si distribuisce come un Chi-quadro. Dato quindi un certo valore di confidenza, la funzione di ripartizione (CDF) calcolata in tale valore fornisce la probabilità entro la quale il fit può essere rigettato.

Test parametrici più avanzati applicabili alle distribuzioni troncate a sinistra sono stati trattati da Chernobai, Rachev, Fabozzi (A. Chernobai, 2005) e permettono di testare la bontà di adattamento sulla coda destra della distribuzione, caratteristica fondamentale quando si tratta di tail fitting. In particolare sono stati considerati i test di Kolmogorov-Smirnov, e Anderson Darling corretti, si veda a tal proposito il Capitolo 5 contenente gli approfondimenti teorici.

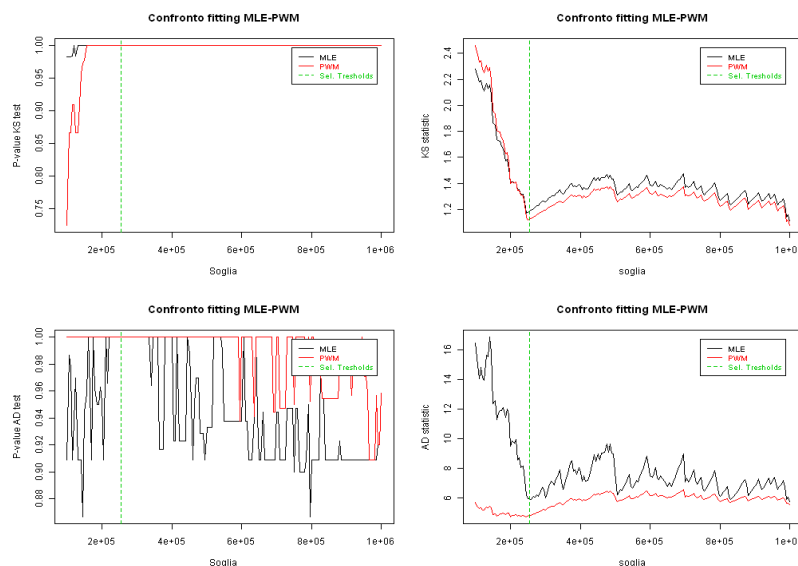


Figura 16 KS e AD quadratico applicato alla coda destra.

4.1.4.4.1.2 Test non parametrici

I test non parametrici dipendono dal valore asintotico di una statistica la cui pdf è indipendente dalle condizioni del problema, sono infatti anche detti metodi indipendenti dalla distribuzione. Tali test si utilizzano per testare l'ipotesi che le realizzazioni di un campione seguano una certa distribuzione di probabilità. Il test confronta la CDF teorica con quella empirica definita dal campione. I test non parametrici attualmente utilizzati sono:

- Kolmogorov – Smirnov
- Kramer - von Mises
- Anderson – Darling

4.1.4.5 **Criteri e linee guida per la scelta della soglia EVT**

Di seguito vengono descritte le linee guida utilizzate per la scelta della soglia monetaria EVT. I criteri adottati dall'analista sono riportati in ordine di priorità. Si distinguono due casi:

1. classi interne di rischio per le quali la severità viene stimata sui soli dati interni (vedasi sezione 3.3.3);
2. classi interne di rischio per le quali la severità viene stimata sui dati interni ed esterni.

Nel primo caso, le analisi ed i diagnostici descritti nelle sezioni precedenti vengono prodotte per una singola distribuzione di coda GPD. Nel secondo caso, le analisi e i diagnostici vengono prodotti per entrambe le distribuzioni G_I e G_E dei dati interni ed esterni. Il collocamento della soglia viene deciso principalmente sulla base dei diagnostici per la distribuzione G_I dei dati interni, esaminando i diagnostici per G_E solo per evitare, laddove possibile, di scegliere soglie non compatibili con la distribuzione dei dati esterni. Qualora non sia possibile effettuare lo

scaling (vedasi la sezione 4.1.4.1), poiché i dati interni ed esterni vengono fusi in un'unica base dati, le analisi vengono prodotte per una singola distribuzione GPD.

1. *Stabilità dei parametri della GPD, principalmente del parametro ξ .*

L'obiettivo è identificare degli intervalli di soglie in cui le stime dei parametri della GPD ed, in particolare, del parametro ξ siano il più stabile possibile. Lo strumento principale è il grafico delle stime del parametro ξ al variare della soglia, che presenta i risultati di 5 stimatori: MPLE, PWM, MLE, MDPD, MGF-AD2. Si considera principalmente la stabilità delle stime fornite dallo MPLE, che è attualmente lo stimatore prefissato per tutte le classi di rischio, seguita in ordine di importanza dagli altri stimatori (nell'ordine sopra elencato).

2. *Numero di eccessi.*

Il numero di eccessi dovrebbe idealmente essere compreso tra 20 e 300. Un numero di eccessi inferiore a 20 è difficilmente giustificabile, in quanto potrebbe corrispondere ad elevata incertezza campionaria e di stima. Un numero di eccessi superiore a 300 può anche essere utilizzato, sebbene si corra il rischio di “contaminare” le code con eventi “non estremi”.

3. *Test di bontà dell'adattamento.*

Si esaminano i seguenti test di bontà dell'adattamento: Kolmogorov-Smirnov (KS), Quadratic Class Upper Tail Anderson-Darling (AD2UP), Quadratic Class Anderson-Darling (AD2), Upper Tail Anderson-Darling (ADUP). Si preferiscono intervalli di soglie in cui venga superato il test KS e, laddove possibile, anche AD2UP.

4. *Distanza tra gli stimatori.*

Si preferiscono intervalli di soglie in cui la somma delle distanze della stima MPLE dalle stime ottenute mediante i restanti 4 stimatori sia minima. Tali soglie costituiscono infatti dei punti in cui si ha convergenza delle varie stime, il che rafforza l'ipotesi distribuzionale.

5. *Grafici diagnostici di adattamento per singole soglie.*

Arrivati a questo punto, di norma sono state identificate una o più soglie potenzialmente idonee. Vengono quindi analizzati grafici diagnostici per l'adattamento di ciascuna di tali soglie, quali:

- a. Grafici quantile-quantile della distribuzione GPD rispetto agli eccessi empirici;
- b. Grafici dei livelli di ritorno (“return level plots”);
- c. Grafici di tipo “tail plot”.

6. *Linearità del grafico Mean Excess e Stabilità dello Hill Plot.*

Come ulteriori verifiche di idoneità, in corrispondenza delle soglie prescelte vengono esaminate:

- a. la linearità del grafico dell'eccesso medio (“Mean Excess”);

- b. la stabilità del grafico di Hill (“Hill plot”).

Laddove siano operate scelte subottimali delle soglie rispetto ai criteri sopra esposti, basate sull’esperienza dell’analista e legate anche alla valutazione degli impatti sulla stabilità del requisito, queste debbono essere descritte in un documento sintetico prodotto ad ogni trimestrale inviato alla funzione di Convalida.

4.1.4.6 Stima della distribuzione del corpo.

Il fitting del corpo della distribuzione è eseguito mediante un’analisi empirica. In questo modo la distribuzione “sperimentale”, rappresentando l’unica fonte di informazione determinata dalla serie storica delle perdite rilevate, è considerata quale distribuzione vera della severity. Di conseguenza si ha la certezza di non distorcere l’informazione contenuta nei dati fino alla soglia EVT selezionata. La robustezza della metodologia di stima del corpo della distribuzione utilizzata è garantita dall’elevata numerosità delle perdite a disposizione (dal momento che il fitting di una distribuzione mediante un’analisi empirica risulta tanto più approssimativo quanto più è bassa la statistica a disposizione).

Per le classi di rischio nelle quali la severità viene stimata sui dati interni ed esterni (vedasi sezione 3.3.3), tali dati vengono fusi in un’unica base dati, sulla quale viene definita la distribuzione empirica.

La distribuzione empirica (S.A. Klugman, 2004) è ottenuta da un campione in cui ad ogni osservazione è associata una probabilità pari ad $1/n$. Più formalmente, la distribuzione empirica cumulata (*cdf*) è la seguente:

$$F_n(x) = \frac{\text{number of } x_j \leq x}{n}$$

La *cdf* è discreta e cresce di $1/n$ ad ogni punto. La funzione di probabilità (*pf*) è definita come segue:

$$f_n(x) = \frac{\text{number of } x_j = x}{n}$$

Comunque, è possibile approssimare la *cdf* con la notazione standard riportata nella definizione seguente:

siano $c_0 < c_1 < \dots < c_r$ i limiti per raggruppare il set di dati e n_j il numero di osservazioni nell’intervallo (c_{j-1}, c_j) , con $j = 1, \dots, r$ e $c_r = \infty$. La *cdf* può essere ottenuta nei limiti, tramite

$F_n(c_j) = \frac{1}{n} \sum_{i=1}^j n_i$. Il grafico della *cdf* empirica si ottiene collegando con una linea i punti $F_n(c_j)$ individuati.

Di seguito si riporta la definizione formale della *cdf* empirica:

$$\tilde{F}_n(x) = \begin{cases} 0, & x \leq c_0 \\ \frac{(c_j - x)F_n(c_{j-1}) + (x - c_{j-1})F_n(c_j)}{c_j - c_{j-1}}, & c_{j-1} \leq x \leq c_j \\ 1, & x \geq c_r \end{cases}$$

La derivata dell'ogiva (dove esiste) è un'empirica approssimazione della funzione di densità la cui definizione formale è la seguente:

$$\tilde{f}_n(x) = \begin{cases} 0, & x \leq c_0 \\ \frac{(c_j - x)F_n(c_{j-1}) + (x - c_{j-1})F_n(c_j)}{c_j - c_{j-1}} = \frac{n_j}{n(c_j - c_{j-1})}, & c_{j-1} \leq x \leq c_j \\ 0, & x \geq c_r \end{cases}$$

La stima della media della distribuzione empirica è la seguente:

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{j=1}^n x_j$$

La stima empirica dei momenti, puri e centrali, della popolazione sono dati dalle seguenti espressioni:

- k -esimo momento puro: $\mu'_k = E(X^k) = \frac{1}{n} \sum_{j=1}^n x_j^k$
- k -esimo momento centrale: $\mu_k = E[(X - \mu)^k] = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^k$
- k -esimo momento puro per dati raggruppati:

$$\hat{\mu}'_k = \sum_{j=1}^r \int_{c_{j-1}}^{c_j} x^k \frac{n_j}{n(c_j - c_{j-1})} dx = \sum_{j=1}^r \frac{n_j(c_j^{k+1} - c_{j-1}^{k+1})}{n(k+1)(c_j - c_{j-1})}$$

4.1.5 Distribuzione delle perdite aggregate per Event Type

Il calcolo della distribuzione delle perdite aggregate per ciascun Event Type è stato effettuato mediante convoluzione delle distribuzioni di frequency, fittata tramite una Poisson e di severity, dove il corpo è simulato con la distribuzione empirica e la coda con una GPD.

La convoluzione¹⁴ avviene tramite Monte Carlo: ogni simulazione di perdita aggregata prevede tante estrazioni sulla severity quanti sono gli accadimenti previsti dalla frequency in una precedente estrazione. Le severity così ottenute si sommano, andando a simulare così la perdita totale di un anno. L'iterazione del processo va a costruire la distribuzione delle perdite aggregate.

¹⁴ Si veda il Capitolo 5 per gli approfondimenti teorici sui differenti metodi di simulazione esistenti.

Per quanto riguarda le estrazioni dalla severity, si deve tener conto che nell'approccio EVT le estrazioni vengono distribuite fra estrazioni dal corpo ed estrazioni dalla coda in maniera proporzionale alla percentuale di probabilità corrispondente al livello di soglia fissato. Se cioè si ha un corpo che rappresenta il 99% della distribuzione, è logico che il 99% delle estrazioni della severity avvenga tramite la pdf che modella il corpo (Distribuzione Empirica) e solo il restante 1% sia estratto dalla funzione usata per ricostruire la coda (Generalized Pareto Distribution). L'algoritmo da utilizzare per generare le estrazioni di severity deve quindi essere in grado di scegliere la funzione da cui estrarre e la frequenza con cui deve estrarre da tale distribuzione.

4.1.5.1 Convoluzione Frequency - Severity

Di seguito si descrive la metodologia di simulazione utilizzata per la stima del VaR di ognuno degli event type.

4.1.5.1.1 Step 1: Simulazione delle Frequency

Una volta stimato dai dati empirici il valore del parametro λ^* per la distribuzione di Poisson si estraggono n valori dalla distribuzione stessa. Tecnicamente si simulano n valori da una distribuzione uniforme $U(0,1)$, $U^f_1, U^f_2, \dots, U^f_n$, i quali sono interpretati come n estrazioni casuali di percentili dalla Poisson di parametro λ^* . Per ottenere le frequenze simulate si inverte la funzione di distribuzione della Poisson per individuare i relativi quantili simulati.

Simulazione	$U(0,1)$	Frequenze
1	U^f_1	$P^{-1}_{(\lambda^*)}(U^f_1) = \lambda_1$
2	U^f_2	$P^{-1}_{(\lambda^*)}(U^f_2) = \lambda_2$
...		
n	U^f_n	$P^{-1}_{(\lambda^*)}(U^f_n) = \lambda_n$

4.1.5.1.2 Step 2: Simulazione dei percentili della Severity

In questa fase si simulano i percentili della severity da associare ad ognuna delle n estrazione eseguite per la frequenza. A tal fine, per ognuna delle n estrazioni fatte dalla distribuzione di frequenza, si simulano da una $U(0,1)$ un numero di percentili pari alla frequenza, $\lambda_1, \dots, \lambda_n$, simulata, ovvero:

Simulazione	Frequenze	Severity				
1	λ_1	U^{s1}_1	U^{s1}_2	...	$U^{s1}_{\lambda_1}$	
2	λ_2	U^{s2}_1	...	$U^{s2}_{\lambda_2}$		
...						
n	λ_n	U^{sn}_1	U^{sn}_2	U^{sn}_3	...	$U^{sn}_{\lambda_n}$

Ad ognuna delle n simulazioni, quindi, si è associato un vettore \mathbf{U}^{si} di lunghezza λ_i con $i=1,2,\dots,n$, i cui elementi sono i percentili della severity che devono essere invertiti utilizzando o la funzione empirica che modella il corpo o la GPD che stima la coda.

A questo fine si confronta ogni percentile simulato U^{si}_j con il valore che assume la probabilità cumulata $F(u)$, dove u è la soglia selezionata per separare il corpo dalla coda della distribuzione. Tutti i percentili inferiori al valore $F(u)$ appartengono al corpo della distribuzione e per ottenere la corrispondente perdita simulata devono essere invertiti utilizzando la funzione che modella il corpo. Quelli maggiori di $F(u)$ invece appartengono alla coda e sono invertiti utilizzando la funzione corrispondente, la GPD.

Simulazione Corpo: $U^{si}_j \leq F(u)$

1	U^{s1}_1	...	$U^{s1}_{\lambda_1}$	
2	U^{s2}_1	...		
...				
n	U^{sn}_1	U^{sn}_3	...	$U^{sn}_{\lambda_n}$

Coda: $U^{si}_j > F(u)$

U^{s1}_2	
...	$U^{s2}_{\lambda_2}$
U^{sn}_2	...

Avendo scelto di modellizzare il corpo della severity tramite la distribuzione empirica, $F(u)$ deriva dal rapporto fra il numero di realizzazioni campionarie inferiori alla soglia e la numerosità complessiva del campione; stima la probabilità empirica di avere una perdita minore o uguale alla soglia stessa.

La probabilità $F(u)$ è quindi determinante per il valore finale del VaR, in quanto serve a stimare la massa di probabilità per la coda, cioè per le perdite più estreme, pari ad $1-F(u)$: al decrescere di $F(u)$ una percentuale maggiore di perdite sono estratte dalla coda, con la conseguente simulazione di perdite aggregate maggiori.

Data la rilevanza del parametro, si è scelto di stimare il valore $F(u)$ sul solo set di dati interni, senza considerare i dati esterni, anche se per la stima della distribuzione empirica si utilizza l'intero set di dati DIPO a disposizione. Conseguenza di questa scelta è che la proporzione fra le perdite simulate superiori alla soglia e quelle inferiori rispetta quella storica desumibile dalla raccolta di dati interni di perdita di MPS e non quella desumibile dal set di dati esterno DIPO, quindi:

$$F(u) = \frac{\# \text{eventi interni} < u}{\# \text{totale eventi interni}}$$

Per le classi che includono dati esterni, ma per le quali non viene effettuato scaling, dato che l'analisi per la determinazione della soglia è effettuata sul dataset dei dati DIPO ed MPS, teoricamente potrebbe essere selezionata una soglia monetaria che, nel dataset interno, non lascia alcun dato sopra soglia, ovvero potrebbe esistere il caso per cui $F(u)=100\%$. Per scongiurare questa eventualità, per tali classi, oltre al continuo monitoraggio del numero di dati interni sopra

soglia individuati, un metodo, a valle del processo, è quello di vincolare il valore di $F(u)$ calcolata sui dati interni ad un valore massimo del 99%¹⁵.

4.1.5.1.3 Step 3: Inversione dei percentili e simulazione delle singole perdite

Per simulare le perdite estratte dal corpo della distribuzione si devono invertire, per ogni vettore U^{si} , i percentili inferiori ad $F(u)$. Prima di essere invertiti, i percentili sono riscaldati nell'intervallo 0-1 dividendoli per il valore $F(u)$. L'inversione avviene poi considerando solo la distribuzione empirica del corpo, vale a dire eliminando dal set di severità tutte le perdite superiori alla soglia u stessa.

Il rescaling della probabilità è reso necessario dalla scelta di misurare la massa di probabilità sul solo set di dati interni. In questo modo si ottiene, infatti, un $F(u)$ diverso rispetto a quello che si sarebbe stimato sull'intero set di dati, inclusi quelli esterni.

$$X_j^i = F_{emp}^{-1} \left(\frac{U_j^{si}}{F(u)} \right) \quad \forall U_j^{si} \leq F(u)$$

Equazione 2

Per simulare le perdite estratte dalla coda si devono invertire, per ogni vettore U^{si} , i percentili superiori ad $F(u)$. Anche in questo caso, per poter essere invertiti correttamente, i percentili devono essere riscaldati riportandoli allo spazio 0-1.

$$X_j^i = GPD_{\xi, \beta}^{-1} \left(\frac{U_j^{si} - F(u)}{1 - F(u)} \right) \quad \forall U_j^{si} > F(u)$$

Equazione 3

Alla fine si ottiene per ognuno degli n vettori di perdite simulati i valori delle perdite dal corpo e dalla coda.

Perdite Simulate dal corpo:

1	X_1^1	...	$X_{\lambda_1}^1$			
2	X_1^2	X_2^2	...	$X_{\lambda_2}^2$		
...
n	X_1^n	X_2^n	X_3^n	$X_{\lambda_n}^n$

Perdite Simulate dalla coda:

X_1^{s1}			
X_1^{s2}	X_2^{s2}	$X_{\lambda_2}^{s2}$	
...	
X_1^{sn}	X_1^{sn}	...	$X_{\lambda_n}^{sn}$

4.1.5.1.4 Step 4: Perdite aggregate annue simulate e stima del VaR

Per ognuna delle n simulazioni la perdita aggregata annua S^i è data dalla somma delle singole perdite simulate, sia dal corpo sia dalla coda.

¹⁵ Precedentemente, in caso di superamento della barriera del 99%, si richiedeva che $F(u)$ venisse calcolata sul dataset utilizzato per modellare la severity. Tuttavia questa soluzione comportava una rilevante instabilità nelle stime.

$$S^i = \sum_{j=1}^{\lambda_j} X_j^i \quad \forall i = 1, 2, \dots, n$$

Equazione 4

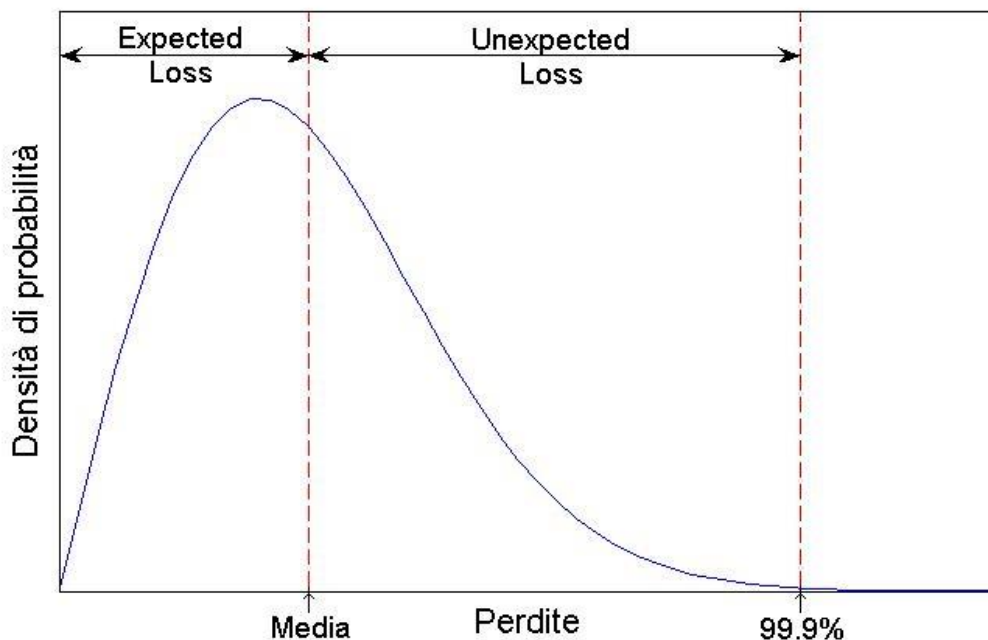
Il vettore S^i delle n perdite aggregate è ordinato per importo e ad ogni osservazione è attribuita una probabilità pari ad $1/n$, similmente ad una distribuzione empirica. Il VaR stimato è pari al quantile al 99,9% della distribuzione di perdite annue simulata.

4.1.5.2 Informazione contenuta nella distribuzione di perdita

La distribuzione delle perdite aggregate per ET contiene tutte le informazioni utili a valutare l'esposizione al rischio per la tipologia di evento analizzato. Volendo condensare in maniera sintetica l'informazione necessaria alle scelte gestionali, si prendono in considerazione due parametri della distribuzione:

- la perdita attesa (Expected Loss, EL);
- la perdita inattesa (Unexpected Loss, UL).

La prima è una misura della perdita media prevista dalla simulazione nell'arco temporale considerato, pari dunque alla media della distribuzione delle perdite aggregate, mentre la seconda vuole quantificare quanto capitale dovrebbe essere allocato per far fronte ad ulteriori perdite possibili ma con una minore probabilità di accadimento, e viene stimata come la differenza fra il valore di VaR (Value at Risk, fissato al 99,9%) e la perdita attesa.



4.1.5.2.1 Perdita attesa (EL)

La convoluzione fra severity e frequency gode della proprietà notevole per cui la media della convoluzione è pari al prodotto delle medie delle distribuzioni di partenza:

$$E(\text{Loss}) = E(\text{Severity}) \cdot E(\text{Frequency})$$

I processi che più influiscono sulla determinazione di queste medie sono:

- Frequency: la media della frequency annuale dipende dalla distribuzione delle perdite per data di rilevazione.
- Severity: la media della severity dipende principalmente dal fit del corpo. Per le scelte metodologiche fatte, è fortemente influenzata dalla media empirica del set di dati DIPO considerati, ma può essere notevolmente influenzata anche dalla dispersione che si registra nelle code dopo l'applicazione della EVT.

4.1.5.2.2 Perdita inattesa (UL)

I quantili più elevati della distribuzione delle perdite aggregate dipendono in maniera più sensibile dai quantili alti della severity e della frequency. Inoltre la UL è in un certo qual modo legata anche alla varianza della distribuzione (S.A. Klugman, H.H. Panjer, G.E. Willmot, 2004):

$$\sigma^2(\text{Loss}) = \sigma^2(\text{Severity}) \cdot E(\text{Frequency}) + E(\text{Severity})^2 \cdot \sigma^2(\text{Frequency})$$

I processi che influenzano il valore della UL sono principalmente:

- Frequency: la forma della frequency annuale non varia in modo sensibile a seconda del tipo di analisi (empirica o parametrica) attraverso la quale viene costruita. Si evidenzia invece, come già discusso, un'influenza sensibile sulla media, che porta ad una traslazione piuttosto rigida dell'intera distribuzione. L'effetto sui quantili più alti delle perdite aggregate è comunque contenuto, dato che solo il primo termine della varianza aumenta in maniera lineare con la media della frequency.
- Severity: i quantili più elevati della severity sono dipendenti dalla stima della EVT. Si ha una dipendenza molto spiccata del valore di perdita inattesa dalla EVT nel caso in cui il parametro di shape vari in maniera sostanziale al variare della soglia, dato che questo comporta una variazione sia sulla varianza della distribuzione di severity sia sulla sua media.

4.1.6 Intervalli di Confidenza del VaR

4.1.6.1 Valutazione dell'errore di simulazione

Una volta determinata la stima di VaR tramite convoluzione si determinano gli intervalli di confidenza entro i quali, con un certo grado di affidabilità, ricadono le stime.

La legge dei grandi numeri garantisce che la media campionaria è uno stimatore consistente della media di una popolazione; vale a dire che grazie alla legge dei grandi numeri possiamo fidarci che la media che calcoliamo a partire da un numero sufficiente di campioni sia sufficientemente vicina alla media vera, quindi al crescere del numero di VaR stimati la media del campione dovrebbe convergere al valore "vero".

4.1.6.2 Analisi di sensitivity del VaR

Un test di stabilità del valore ottenuto e la sua non eccessiva dipendenza dalla scelta effettuata per il posizionamento della soglia di applicazione della EVT (che è il momento in cui maggiore è l'apporto del valutatore e l'eventuale errore) è quella di testare la stabilità dei valori di requisito all'interno di un certo range di significatività della soglia monetaria.

In altre parole si riesegue una simulazione di VaR per diverse soglie monetarie (come si fa per individuare un intervallo di stabilità del parametro di shape in cui individuare la soglia) e si controlla se la particolare soglia selezionata produca o meno un requisito stabile nell'intorno considerato.

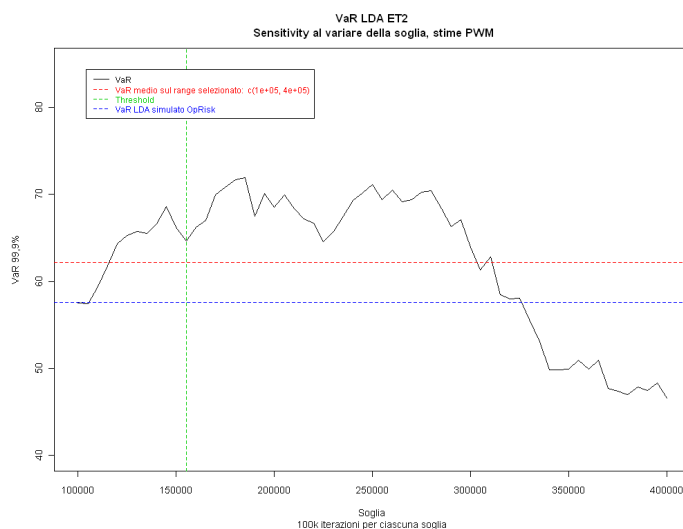


Figura 17 Analisi di sensitivity del VaR al variare della soglia monetaria

4.1.7 Assunzioni

Le assunzioni fatte nell'implementazione del metodo utilizzato, discusse di volta in volta all'interno dell'analisi, sono le seguenti¹⁶:

- le perdite avvengono in qualunque giorno dell'anno (durata gestionale dell'anno: 365 giorni);
- la distribuzione delle frequenze di accadimento non cambia nel corso del tempo e non influenza la distribuzione della frequenza delle perdite future;
- la distribuzione di frequency annuale non dipende in maniera sensibile dal sampling rate scelto;
- la distribuzione della severity non cambia nel corso del tempo;

¹⁶ Si è cercato di limitare il più possibile il numero e l'entità delle scelte arbitrarie all'interno dell'analisi. Per le assunzioni che si sono dovute accettare, sono stati valutati sempre gli impatti sui risultati.

- la distribuzione di frequency e quella di severity non sono correlate.

4.2 Scenario Analysis

4.2.1 Introduzione

Scopo di questa sezione è descrivere la metodologia statistico-quantitativa per:

- il trattamento delle stime soggettive raccolte attraverso l'Analisi di Scenario;
- la produzione di un output espresso in termini di capitale: perdita attesa, inattesa e valore a rischio.

Le modalità di preparazione dei questionari di Analisi di Scenario (strumento di raccolta delle stime soggettive), così come tutte le considerazioni di natura gestionale che seguono all'analisi, sono oggetto di trattazione specifica nel manuale di identificazione e nella direttiva sui rischi operativi. Questa sezione è focalizzata solo sugli aspetti statistici alla base del modello quantitativo che si pone come obiettivi specifici:

- supportare la raccolta delle stime soggettive con una struttura quantitativa basata su una modellistica rigorosa, riducendo così l'ampia discrezionalità lasciata da sistemi basati su scale di tipo puramente qualitativo;
- giungere ad una stima della perdita attesa (Expected Loss, EL), perdita inattesa (Unexpected Loss, UL) e del VaR per tipologia di rischio / fattore di rischio / unità organizzativa.

4.2.2 Stime soggettive

L'analisi di Scenario è un processo di autovalutazione da parte del Top Management del profilo di rischio operativo derivante dalle criticità individuate nel tavolo "Rischi Operativi".

L'analisi di Scenario, per come strutturata nel Gruppo Montepaschi, consente di ottenere i seguenti risultati:

- Esprimere in termini di rischio (Capitale) le opinioni raccolte dal top management.
- Valorizzare l'ottica forward-looking intrinseca nei giudizi soggettivi espressi.
- Ottenere risultati aggregati per differenti ET o BU.
- Integrare più fonti di informazione (soggettiva e quantitativa).
- Assicurare il commitment del management attraverso il suo diretto coinvolgimento.
- Rappresentare uno strumento di base per il *decision making* (strategie mitigative).

Dall'analisi di Scenario si ottengono quindi stime soggettive:

- espresse dal Top Management aziendale,
- con riferimento ai rischi operativi specifici individuati dalle varie unità organizzative, identificate anche come *risk owner*;
- espresse in termini di frequenza e impatto caratterizzanti le perdite operative potenzialmente riconducibili ai suddetti rischi rilevati.

Le stime espresse dal Top Management sono:

- stima soggettiva della **frequenza media**, definita come il numero medio atteso di eventi nell'arco temporale di riferimento (un anno);
- stima soggettiva dell'**impatto tipico**, definito come il valore di perdita più frequente per ogni classe di rischio;
- stima soggettiva dell'**impatto nel caso peggiore** ("*worst case*"), definito come l'impatto del singolo evento qualora esso si manifesti nel peggior modo (ragionevolmente) concepibile: quest'ultima indicazione, fondamentale per raccogliere informazione relativamente alla dispersione dell'impatto rispetto al proprio valor medio (leggi rischiosità dell'evento considerato), consente di giungere ad un output espresso in termini di perdita inattesa;
- individuazione del **fattore di rischio** prevalente, definito come il fattore di rischio tra Risorse Umane, Processi e Sistemi, sulla cui gestione è possibile intervenire per mitigare il rischio.

Il top management (o esperto) è libero di indicare i valori che ritiene più opportuni, senza che le proprie previsioni debbano sottostare ad alcun vincolo o livello di granularità stabilito a priori; in questo modo sono privilegiati e responsabilizzati gli esperti in grado di fornire stime più precise e dettagliate.

Dato che i valori puntuali stimati dall'esperto compaiono nelle equazioni che determinano i parametri delle distribuzioni, senza alcuna approssimazione, nel passaggio dalle stime ai parametri non viene perso alcun dettaglio dell'informazione fornita dall'esperto.

4.2.3 Analisi quantitativa: approccio attuariale

Le ipotesi alla base dell'approccio attuariale adottato prevedono:

- di considerare separatamente la distribuzione del numero di eventi in un determinato arco temporale (frequenza) e quella dell'impatto dei singoli eventi nel medesimo intervallo di tempo (impatto);
- di procedere quindi alla loro "convoluzione" in una unica distribuzione delle perdite aggregate e alla determinazione del VaR tagliando la coda di tale distribuzione aggregata al livello di significatività desiderato.

Seguendo l'impostazione attuariale, quindi, la perdita inattesa viene determinata sulla distribuzione delle perdite, ottenuta per convoluzione di una distribuzione di frequenza ed una di impatto. Occorre perciò formulare opportune ipotesi in merito alle distribuzioni che descrivono le variabili casuali frequenza e impatto. Nell'analisi dei dati effettivi di perdita, le ipotesi parametriche possono essere testate attraverso prove di adattamento della distribuzione ipotizzata sul campione a disposizione. Nell'ambito dell'analisi condotta su stime soggettive, invece, trattandosi di un'analisi ex ante (ancora non ci sono state manifestazioni di perdita, l'orizzonte temporale di riferimento è l'anno successivo), occorre assumere come valide a priori alcune ipotesi. In particolare sono state adottate le seguenti ipotesi parametriche:

- frequenza: distribuzione di Poisson (1 parametro);
- impatto: distribuzione Lognormale (2 parametri).

La distribuzione di Poisson, a dominio discreto (interi non negativi) e con funzione di ripartizione

$$f_N(n) = \frac{\lambda^n}{n!} e^{-\lambda}, \quad \lambda > 0, \quad n = 0, 1, 2, \dots,$$

è definita da un unico parametro $\lambda > 0$ che ne rappresenta sia la media (numero medio di eventi per unità di tempo), sia la varianza. La funzione di densità della Lognormale è data da:

$$g_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{x} \exp\left\{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right\}, \quad \sigma > 0,$$

che dipende dai due parametri μ, σ . Valore atteso, moda e funzione di ripartizione inversa sono dati da:

$$\bar{x} = \exp\left(\mu + \frac{\sigma^2}{2}\right), \quad x_{\max} = \exp(\mu - \sigma^2), \quad x_p = \exp\{\mu + \sigma G^{-1}(P)\},$$

dove G^{-1} indica la funzione di ripartizione inversa della Normale standard. La terna di valori forniti dagli esperti (frequenza media, impatto tipico, impatto caso peggiore) viene tradotta nei tre parametri distribuzionali λ, μ, σ (si veda la sottosezione successiva per i dettagli).

Una volta definiti i valori di λ, μ, σ , la distribuzione delle perdite aggregate (sul periodo temporale considerato) della classe di rischio in esame viene ottenuta mediante convoluzione delle distribuzioni di frequenza e impatto. È su questa distribuzione che, prescelto un opportuno percentile, si procede alla determinazione della perdita inattesa associata alla tipologia di rischio, come di seguito raffigurato.

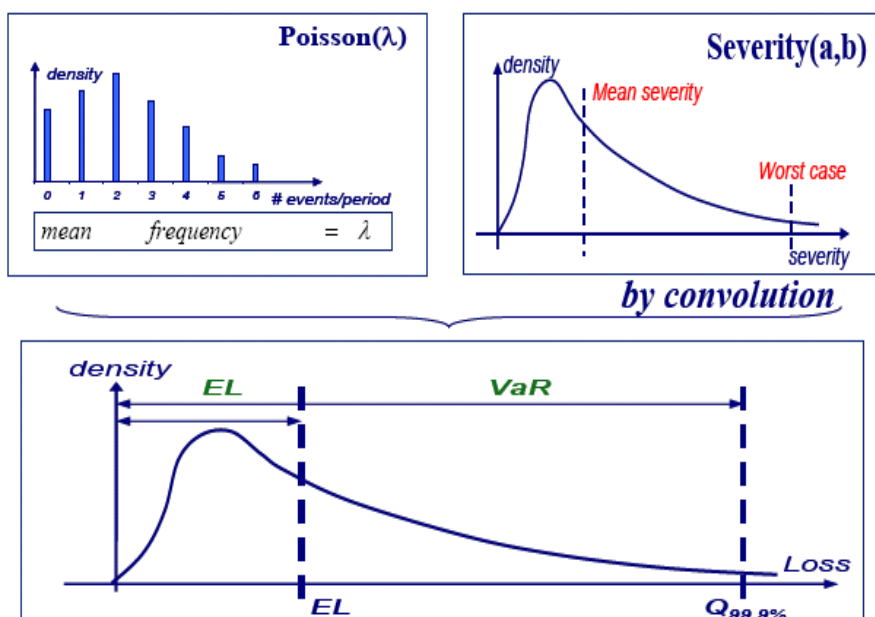


Figura 18 Scenario: Convoluzione di frequenza e impatto

Ricapitolando, a partire da una qualsiasi terna (freq, sev, wcase) di valori forniti dagli esperti per frequenza media, impatto medio e impatto peggiore, è sempre possibile ottenere un determinato ammontare di perdita attesa e inattesa.

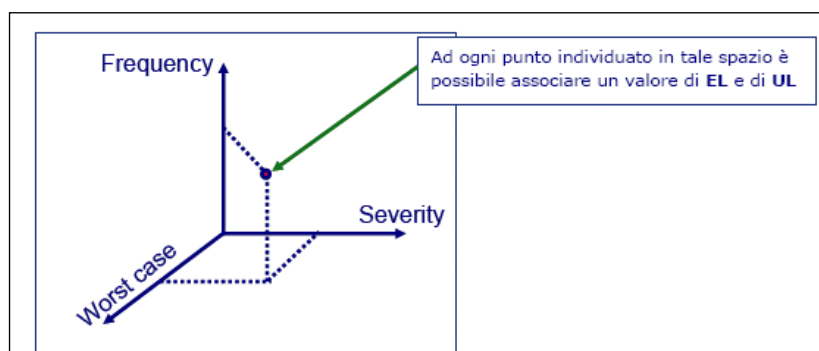


Figura 19 Spazio Frequenza/Impatti

4.2.4 Calcolo dei parametri distribuzionali a partire dalle stime soggettive

L'esecuzione dello Scenario comporta l'acquisizione di tre valori numerici: frequenza media, impatto medio e impatto peggiore; essi vengono utilizzati per determinare i parametri delle distribuzioni coinvolte.

- La frequenza media è il numero atteso di eventi nell'orizzonte temporale di un anno: questo valore viene quindi preso come stima del parametro λ della distribuzione di Poisson.

- L'impatto tipico è inteso come il valore di impatto che si manifesta con maggiore frequenza: viene quindi identificato con la moda della distribuzione (punto di massimo della densità di probabilità)¹⁷.
- L'impatto peggiore è inteso come il più alto valore di perdita del singolo evento che si prevede si manifesti entro un dato orizzonte temporale di analisi.

Pertanto è necessario fissare l'orizzonte temporale di confidenza. L'interpretazione adottata lega il quantile sulla stima di impatto peggiore al numero totale di eventi entro l'orizzonte temporale di confidenza considerato (ovvero al numero di eventi tra cui quello stimato risulta il peggiore). In coerenza con la previsione di frequenza media f_{av} e con un orizzonte temporale di confidenza T_{conf} il livello di probabilità con il quale la stima raccolta è la peggiore è determinato come:

$$P_{wc} = 1 - \frac{1}{T_{conf} * f_{av}}$$

Per ottenere tale quantile associato all'impatto *worst case* si utilizzano tre parametri:

- Il rapporto minimo δ_{th} tra impatto tipico e caso peggiore;
- Il quantile minimo Q_{min} associabile all'impatto peggiore;
- Il quantile massimo Q_{max} associabile all'impatto peggiore.

La determinazione dei valori ragionevoli per questi parametri è stata ottenuta in maniera empirica attraverso analisi di sensitività e ragionevolezza delle scelte effettuate. Sono stati selezionati i valori:

$$\delta_{th} = 5, \quad Q_{min} = 0.90, \quad Q_{max} = 0.999.$$

Il valore $\delta_{th} = 5$ è stato inizialmente ottenuto osservando il valore mediana della serie ottenuta dai rapporti fra caso peggiore ed impatto tipico stimati dal management durante le sessioni di scenario 2009 all'interno del Gruppo Montepaschi. L'appropriatezza di tale valore viene sottoposta a monitoraggio nel corso del tempo.

Nell'approccio adottato si procede calcolando il valore empirico del rapporto impatto peggiore su tipico :

$$\delta_{emp} = \frac{\text{Impatto peggiore}}{\text{Impatto tipico}}$$

Si definisce quindi la variabile δ :

¹⁷ Si ritiene che l'esperto sia in grado di fornire la stima dell'impatto tipico con un maggiore grado di affidabilità rispetto ad una stima dell'impatto medio (valore atteso della distribuzione): per fornire quest'ultima, l'esperto dovrebbe essere in grado di tenere in debita considerazione anche la coda della distribuzione, ovvero gli eventi a bassa frequenza ed alto impatto, cosa complessa proprio per la natura di tali eventi.

$$\delta = \max \left[\frac{\delta_{emp}}{\delta_{th}}, 1 \right]$$

Se il rapporto caso peggiore/impatto tipico è minore o uguale a δ_{th} , si associa al caso peggiore il quantile minimo, altrimenti si determina un quantile in base al rapporto reale tra caso peggiore ed impatto tipico. Tale quantile non può comunque superare il quantile massimo fissato (99,9%):

$$Q_1 = \min \left[\left(1 - \frac{1 - Q_{min}}{\delta} \right), Q_{max} \right]$$

Si determina inoltre l'orizzonte di confidenza relativo al *worst case*, derivante dal quantile così calcolato e dalla frequenza media indicata durante la sessione di Scenario. E' stato inoltre definito un vincolo di significatività di tale orizzonte, per cui si prevede di subire un evento pari al caso peggiore al massimo una volta l'anno (coerentemente con l'orizzonte temporale dell'esperto):

$$T_{conf} = \max \left\{ \frac{1}{1 - Q_1} * \frac{1}{f_{av}}, 1 \right\}$$

questa valutazione può essere utile all'esperto per calibrare le proprie valutazioni. Invertendo quest'ultima relazione otteniamo la formula iniziale che lega il quantile all'orizzonte di confidenza e alla frequenza media:

$$P_{wc} = 1 - \frac{1}{T_{conf} * f_{av}}.$$

Questa metodologia permette di associare al caso peggiore un quantile variabile della distribuzione di impatto, coerente con le valutazioni di frequenza di accadimento e quindi di capacità previsionale dell'esperto (maggiore è la frequenza attesa, maggiore è la probabilità di valutare correttamente il caso peggiore, viceversa la stima basata su poca esperienza può scontare una distorsione implicita). La stima del *worst case*, inoltre, non è sempre associata al quantile 99,9% della distribuzione (che equivale al peggiore tra 1000 eventi), ma ad un quantile più realistico e coerente con le previsioni di frequenza.

La tabella sottostante riporta alcuni esempi di applicazione del metodo proposto.

Frequenza	Impatto Tipico	Caso peggiore	delta emp	delta	Quantile associato al worst case	Orizzonte
5	2,000	3,000	1.50	1.00	0.9000	2
0.5	1,000	5,000	5.00	1.00	0.9000	20
2	10,000	80,000	8.00	1.60	0.9375	8
25	1,000	1,000,000	1,000.00	200.00	0.9990	40

Tabella 1 Esempio di determinazione dell'orizzonte di confidenza associato al worst case

Osservazioni.

1. La scelta di un quantile minimo del 90% da associare al caso peggiore presuppone un orizzonte di esperienza del manager di circa 10 anni; inoltre, l'introduzione del quantile massimo associabile alla stima permette di evitare riscalamenti al ribasso durante l'integrazione quali-quantitativa; il modello di integrazione considera infatti il valore traslato della risposta dell'esperto al 99,9% della distribuzione per ottenere i parametri di una GPD qualitativa¹⁸.
2. La scelta di richiedere le stime di impatto tipico ed impatto peggiore rispetto ad un dato orizzonte temporale – anziché di impatto medio e peggiore al 99,9% – è motivata dal maggior grado di attendibilità con cui l'esperto è in grado di fornire tali stime; inoltre tale scelta permette di ottenere statistiche di rischio più prudentiali a parità di stime raccolte.
3. Per la distribuzione Lognormale la moda è sempre inferiore alla media (asimmetria destra); inoltre per valori tipici di f_{av} e T_{conf} il quantile stimato è inferiore al 99,9%; si determina perciò uno spostamento a destra della distribuzione di impatto, verso valori più alti.
4. Nell'esecuzione dello Scenario, è fondamentale che l'esperto sia consapevole dell'interpretazione delle stime. In particolare, è condiviso con l'esperto l'orizzonte temporale di confidenza per la stima di impatto peggiore.
5. Durante l'esecuzione dello scenario è possibile che il manager chiamato a valutare uno specifico rischio non sia in grado di fornire un valore puntuale; in tal caso possono essere utilizzate delle classi di valori che permettono di avere un'idea dimensionale del fenomeno. In questi casi nelle successive elaborazioni si considera il valore medio della classe.
6. I due parametri della distribuzione di impatto sono determinati risolvendo rispetto ad essi le equazioni per la moda della distribuzione e per la funzione di ripartizione inversa (valutata al quantile stimato per il caso peggiore); per una Lognormale si hanno soluzioni in forma esplicita¹⁹:

$$\begin{cases} \sigma = \frac{1}{2} \left\{ -G^{-1}(P_{wc}) + \sqrt{[G^{-1}(P_{wc})]^2 + 4 \ln \frac{x_{wc}}{x_{max}}} \right\}, \\ \mu = \sigma^2 + \ln x_{max}, \end{cases}$$

dove x_{max} e x_{wc} sono le stime di impatto tipico ed impatto peggiore.

¹⁸ Per i dettagli si veda il paragrafo 4.3 sull'integrazione quali-quantitativa.

¹⁹ Un'equazione di secondo grado del tipo: $ax^2 + bx + c = 0$ con $a \neq 0$ ha soluzione $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

4.2.5 Calcolo del VaR qualitativo per classe di rischio interna

A partire da dicembre 2014 il Gruppo Montepaschi ha adottato un nuovo metodo di calcolo del requisito qualitativo di Gruppo per classe di rischio interna del modello AMA.²⁰ L'obiettivo del nuovo metodo è rendere il processo più robusto ed accrescere la capacità del modello di catturare il profilo di rischio operativo prospettico della banca.

La nuova metodologia adottata per il calcolo del VaR qualitativo per classe di rischio interna prevede tre passi:

1. Simulazione per ciascuna domanda di scenario;
2. Simulazione per ciascuna sottoclasse di ciascuna classe di rischio interna del modello;
3. Simulazione per ciascuna classe di rischio interna del modello.

Passo 1: il VaR qualitativo per ciascuna domanda è calcolato come il percentile al livello 99.9% della distribuzione di perdita di tale domanda. La distribuzione di perdita si ottiene mediante convoluzione delle distribuzioni di frequenza e severità, come descritto nella sezione precedente. Le simulazioni di ciascuna domanda vengono salvate, per essere poi riutilizzate nei passi successivi.

Passo 2: data una classe di rischio interna, le domande afferenti a tale classe vengono suddivise in sottogruppi, a seconda del "Fattore di rischio prevalente" ad esse attribuito. Poiché i fattori di rischio considerati sono:

4. Risorse Umane,
5. Processo,
6. Sistemi,

l'insieme delle domande afferenti ad una classe interna di rischio viene suddiviso in tanti gruppi (sottoclassi) quanti sono i fattori di rischio di tale insieme di domande. Al massimo le sottoclassi possono essere 3. Si calcola quindi una simulazione congiunta assumendo prudenzialmente correlazione perfetta tra le domande di ciascuna sottoclasse. Ad ogni passo,

- a. si estrae casualmente un livello di probabilità (ovvero un numero tra 0 ed 1);
- b. vengono estratti dalle simulazioni di ciascuna domanda (salvate al punto precedente), percentili corrispondenti al livello di probabilità estratto al punto a. (l'utilizzo del medesimo percentile per tutte le domande è l'aspetto che garantisce la correlazione perfetta);
- c. tali percentili vengono quindi sommati, fornendo un singolo valore che rappresenta la perdita annuale simulata per la sottoclasse in oggetto.

²⁰ Autorizzazione comunicata con lettera n.0050310/15 del 19/01/2015, a valere sulle segnalazioni riferite al 31.12.2014. Precedentemente il VaR qualitativo veniva calcolato assumendo correlazione nulla all'interno delle classi di rischio e correlazione perfetta tra le classi.

Passo 3: al termine del passo 2 sono state prodotte le simulazioni di tutte le sottoclassi di ciascuna classe di rischio interna. Il VaR qualitativo aggregato per ciascuna classe di rischio interna viene calcolato come segue:

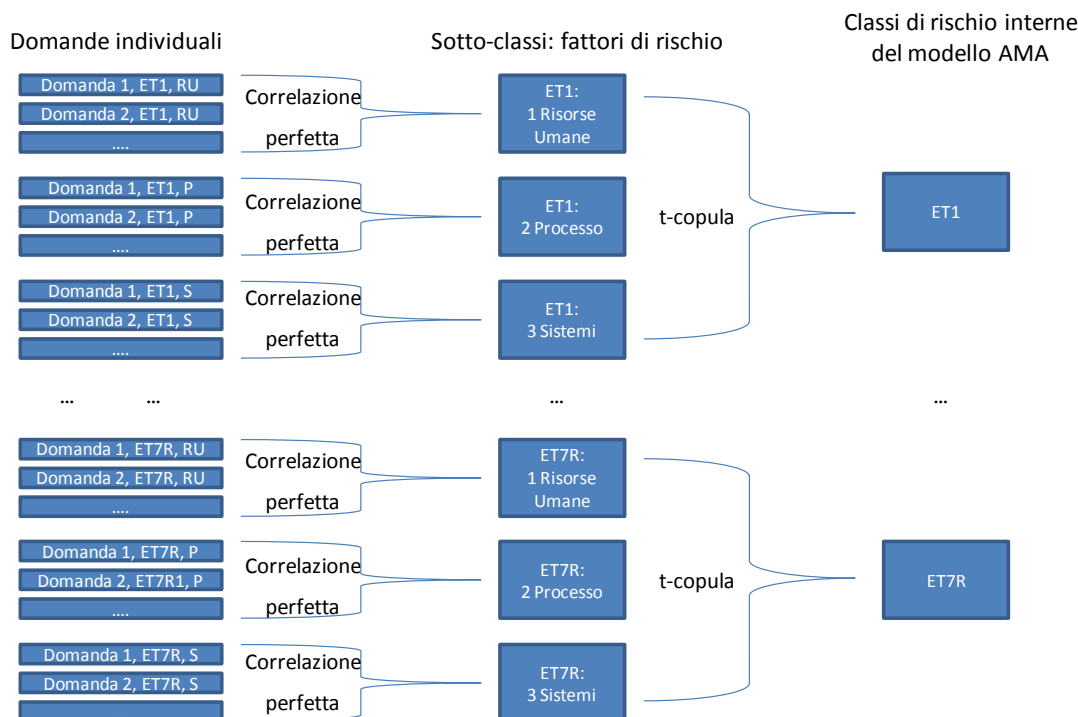
- se la classe contiene una sola sottoclasse, allora la distribuzione simulata della classe viene posta uguale a quella della sottoclasse;
- se la classe contiene due o tre sottoclassi, si effettua una aggregazione tra le sottoclassi, basata su approccio a copula t-Student con $v=3$ gradi di libertà e matrice di correlazione lineare data da

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \quad \text{oppure} \quad \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}$$

dove $\rho = 20\%$ è identico per tutte le coppie di sottoclassi. Osserviamo che tale metodo di aggregazione tra classi è del tutto analogo a quello utilizzato per l'integrazione tra i diversi event type (vedi Sezione 4.4). In questo caso la scelta del valore della correlazione tra le sottoclassi (ρ) è stata guidata dai livelli di correlazione tipici osservati in letteratura tra classi di rischio²¹; la scelta dei gradi di libertà utilizzati è dettata da motivi prudenziali.

La figura sottostante riporta in forma schematica le tre fasi del processo di calcolo del VaR qualitativo per classe di rischio interna.

²¹ Si veda ad esempio l'articolo "*Observed Correlations and Dependencies Among Operational Losses in the ORX Consortium Database*", di Eric Cope e Gianluca Antonini, del 27 novembre 2008, disponibile alla pagina <http://www.orx.org/Pages/ORXResearch.aspx>, con riferimento in particolare alle Figure 1 e 2 e relativa discussione nel testo.



Il VaR qualitativo complessivo viene calcolato sommando semplicemente i VaR di ciascuna classe di rischio. Questo equivale ad assumere correlazione perfetta tra le diverse classi di rischio.

Le assunzioni sui parametri di correlazione utilizzati nella copula sono oggetto di verifica annuale, in termini di tenuta e prudenzialità, dopo il completamento delle analisi di Scenario di ciascun anno. La verifica mira a valutare la sensitività del VaR di scenario rispetto al parametro di correlazione ρ utilizzato per il calcolo delle distribuzioni qualitative per ciascuna classe interna di rischio. L'analisi di sensitività viene effettuata ricalcolando il VaR qualitativo sotto varie ipotesi per ρ . Oltre ai casi limite di correlazione nulla e perfetta ($\rho = 0$ e $\rho = 1$), si esaminano i valori $\rho = 10\%$, 30% , 50% , 75% , mettendoli a confronto con il valore $\rho = 20\%$ utilizzato per il calcolo del requisito patrimoniale. L'analisi mira a confermare che il VaR qualitativo ottenuto per $\rho = 20\%$ assuma un valore intermedio tra i casi estremi di ρ corrispondenti a correlazione nulla e perfetta e che la sensibilità del VaR attorno a tale valore sia contenuta.

Durante la fase di definizione e ripartizione delle domande di scenario per classe di rischio, viene verificata l'assenza di aggregazioni o frammentazioni artificiali di eventi di rischio simili e delle relative domande del questionario. Si procede suddividendo le domande del questionario per Event Type di livello 2, Ambito²² e Tema²³ (si veda il Manuale di Identificazione per i dettagli); ciascuna domanda relativa ad una tripletta ("Event Type" - "Ambito" - "Tema") viene poi

²² Si tratta dell'ambito di rischio operativo, circostanziato sul business, che consente di raggruppare insieme domande diverse che riguarda no gli stessi specifici processi/attività, ad esempio "Ambito Finanza";

²³ Consente di descrivere in modo più dettagliato le possibili criticità relative ai singoli Ambiti. Ad esempio per l'ambito Finanza un Tema è "monitoraggio dei limiti".

associata ad una Società del Gruppo Montepaschi e ad un'Unità Organizzativa all'interno di tale Società. La verifica consiste nel controllare che esista una sola domanda per ciascuna tripletta "Event Type", "Ambito", "Tema" abbinata alla singola coppia "Società" ed "Unità Organizzativa". Inoltre, allo scopo di evitare variazioni immotivate nel fattore di rischio delle domande ed evitare instabilità nelle risposte nei casi dubbi, viene fornito al risk owner il fattore di rischio assegnato alla domanda di scenario l'anno precedente.

4.2.6 Calcolo del VaR qualitativo per Società

Il VaR qualitativo per ciascuna Società del Gruppo Montepaschi appartenente al perimetro AMA viene calcolato sia a correlazione nulla, sia a correlazione perfetta a partire dalle distribuzioni di perdita di ciascuna domanda.

- La simulazione congiunta a correlazione nulla non fa uso di funzioni copula o matrici di correlazione: semplicemente, per ciascuna Società le domande vengono suddivise in 7 gruppi a seconda del loro Event Type (ET) di primo livello di Basilea II, dopodiché per ciascun gruppo viene effettuata una simulazione simultanea ed indipendente delle distribuzioni marginali di ciascuna domanda, seguita da aggregazione annuale (somma per ciascun anno simulato). Per ciascuna Società vengono in tal modo calcolati sette VaR a correlazione nulla, uno per ciascun ET di livello uno. La somma di tali VaR fornisce il VaR della Società considerata: si assume quindi correlazione nulla tra le domande di ciascun ET di livello uno e correlazione perfetta tra i diversi ET.
- La simulazione congiunta a correlazione perfetta viene realizzata in modo analogo: per ogni gruppo di domande ad ogni passo vengono estratti, da ciascuna domanda del gruppo, i percentili corrispondenti al medesimo livello di probabilità, il tutto seguito da aggregazione annuale. Questo fornisce sette VaR a correlazione perfetta, uno per ciascun ET di livello 1, che vengono poi sommati (assumendo nuovamente correlazione perfetta tra i diversi ET) per fornire il VaR per Società a correlazione perfetta.

4.2.7 Archiviazione dei risultati

A partire dal 31 dicembre 2014, tutte le procedure di calcolo del VaR qualitativo:

- per domanda
- per classe di rischio
- per società

sono state migrate nell'ambiente ASIA.

Il calcolo viene eseguito una volta l'anno, in corrispondenza della trimestrale di dicembre; i risultati (incluse le simulazioni delle varie distribuzioni coinvolte) vengono archiviati in una cartella di ASIA che è accessibile in sola lettura ed a cui le esecuzioni in corso d'anno poi fanno riferimento.

Il confronto fra la rischiosità prospettica, risultato delle opinioni espresse nelle analisi di scenario dai Top Manager, e la rischiosità storica, risultato dell'elaborazione quantitativa delle perdite di LDC, è facilmente ottenibile a livello aggregato, attraverso un confronto diretto delle relative stime di VaR.

4.3 Integrazione quali-quantitativa

A partire da dicembre 2014²⁴ il Gruppo Montepaschi ha adottato un nuovo metodo di integrazione tra componente storica e componente di scenario, con l'obiettivo di incrementare l'incidenza delle valutazioni qualitative nel calcolo del requisito, alla luce di una maggiore consapevolezza del management sulla rilevanza dello svolgimento delle analisi di scenario ed in modo da allineare le logiche del modello AMA alle *best practices* internazionali.

Il metodo si basa sulla costruzione di una distribuzione integrata come mistura delle distribuzioni di perdita storica e qualitativa di scenario: la distribuzione integrata può essere vista come una combinazione della distribuzione qualitativa con un peso pari a w e di quella storica con un peso pari a $1 - w$. La distribuzione integrata dipende quindi dal valore fissato per il peso w .

L'approccio utilizzato prevede comunque che la componente storica abbia il peso maggiore nel calcolo del requisito: si impone quindi che il peso w possa variare solamente nell'intervallo 10%-40%.

Il metodo presuppone di assegnare alla componente qualitativa di scenario un peso maggiore rispetto all'approccio utilizzato fino al 2014: il peso è determinato sulla base di quanto i dati storici riflettano correttamente il profilo di rischio attuale e prospettico della banca. L'idea è quella incrementare il peso (all'interno dell'intervallo prefissato) della componente di scenario nel momento in cui si ritenga che le attività ed il business della banca o la situazione di contesto evidenzino dei “*break*” strutturali rispetto al passato. In tale situazione i dati osservati storicamente potrebbero, infatti, non risultare coerenti con il reale profilo di rischio del Gruppo, mentre le analisi di scenario, al contrario, esprimendo le valutazioni soggettive *forward looking*, dovrebbero tener conto del nuovo profilo e del nuovo contesto.

Il peso w da assegnare alla componente di scenario viene stimato sulla base di un indicatore sintetico, costruito per catturare situazioni di “*break*” strutturale rispetto al passato.

In sintesi, il metodo consiste dei seguenti passi, descritti nelle successive sotto-sezioni:

1. costruzione di un indicatore di continuità/discontinuità rispetto al passato che possa rilevare un “*break*” strutturale;
2. trasformazione dell'indicatore di continuità/discontinuità nel peso w da assegnare alla componente di scenario;

²⁴ Autorizzazione comunicata con lettera n.0050310/15 del 19/01/2015, a valere sulle segnalazioni riferite al 31.12.2014.

3. calcolo della distribuzione mistura per l'integrazione tra la componente storica e la componente di scenario, in base al peso w calcolato al passo precedente.

4.3.1 Indicatore di continuità/discontinuità

Al fine di costruire un indicatore di continuità/discontinuità rispetto al passato, sono stati individuati alcuni driver, trasversali alla banca e di natura eterogenea, che consentano di identificare situazioni di *break* strutturale in vari ambiti di operatività, sia in termini di business aziendale sia di rischio operativo. In analogia con i *Key Risk Indicators* calcolati dal Gruppo Montepaschi, sono stati quindi identificati i seguenti comparti:

- Credito,
- Mercato,
- Liquidità,
- Compliance,
- Wealth Management,
- Pianificazione,
- Risorse Umane,
- Legale,
- Macroeconomico.

Per ciascun comparto, sulla base del giudizio “esperto” delle singole funzioni competenti, sono stati individuati alcuni indicatori che sono in relazione con l'operatività ed il business della banca e rappresentano, allo stesso tempo, informazioni note e utilizzate gestionalmente dal management.

L'ultimo comparto non si riferisce a dati interni, ma ad indicatori esterni macroeconomici, in modo da avere un'indicazione anche relativamente al contesto di riferimento.

Per ciascun comparto, i criteri di scelta degli indicatori forniti dalle varie funzioni si sono basati su:

1. evitare informazioni ridondanti all'interno dei comparti e sovrapposizioni tra diversi comparti;
2. privilegiare gli indicatori con una maggiore profondità storica;
3. privilegiare l'uniformità dei dati all'interno del comparto.

Va comunque precisato che la scelta degli indicatori sarà oggetto di rivalutazioni periodiche, dal momento che deve essere funzionale ad identificare eventuali situazioni di discontinuità della banca e deve basarsi su indicatori gestionalmente utilizzati e monitorati. In particolare trimestralmente avviene l'aggiornamento delle serie storiche degli indicatori ed annualmente insieme alle funzioni competenti viene verificata la composizione del paniere e valutata la possibilità di modifica/sostituzione/cancellazione/inserimento degli indicatori. Le serie storiche vengono archiviate sull'ambiente ASIA.

Compliance	Pianificazione	Liquidità
Tempi concessione Credito	Raccolta a vista conti corr pass	Saldo liquidità operativa 1 mese
Masse adive % di adeguatezza	Raccolta a vista dep a risp	Limite liquidità operativa 1 mese
% proposte eseguite completamente	Raccolta a breve termine-certif dep<18m	Ratio 1 liquidità strutturale
% clienti rossi non contattati	Raccolta a breve termine-gof raccolta	Limite Ratio 1 liquidità strutturale
% clienti nuovi	Raccolta a breve termine-pct	
% clienti persi	Raccolta a breve estero	
Compendio tra acquisition ed erosion	Raccolta a m. lungo termine-certif dep>18m	
Clienti a rischio Abbandono	Raccolta a m. lungo termine-obbligaz senior	
% gestori con n. clienti superiore al valore soglia	Raccolta a m. lungo termine-prestito subordinato	
% gestori con n. clienti inferiore al valore soglia	Impieghi a vista	
Turn over medio gestori	Impieghi a scadenza Italia	
Numero Reclami pervenuti	Impieghi a scadenza estero	
Numero Richieste di rimborso carte	Impieghi a m. lungo termine -cred spec	
Tempi medi di rimborso al cliente	Impieghi a m. lungo termine -mutui	
Tempi medi di riscontro Isola della rete	Impieghi a m. lungo termine -altri impieghi	
% di personale a zero ore di formazione		
Soddisfazione complessiva della clientela		
Credito	Risorse Umane	Legale
Nuove sofferenze/ non sofferenze anno prec.(importi)	N.assunzioni/organico	Nuove cause per anatocismo/numero clienti
Nuove sofferenze/ non sofferenze anno prec.(teste)	N.adesione fondo di solidarietà/organico	Nuove cause per piani finanziari/numero clienti
Incagli/tot.crediti(teste)	N.esodo incentivato/organico	Nuove cause per credito/numero clienti
Incagli/tot. crediti (importi)	N.dimissioni/organico	Nuove cause per titoli in default/numero clienti
Sofferenze /tot crediti(importi)	N.licenziam/organico	Nuove cause per altro/numero clienti
Sofferenze/tot. crediti(teste)	N.pensionamento/organico	Nuovi accantonamenti per anatocismo/num. nuove cause tot.
Nuovi flussi/importi in bonis anno precedente(importi)	Ore formazione/organico	Nuovi accantonamenti per piani finanziari/num.nuove cause tot.
Nuovi flussi/import in bonis anno precedente(teste)	Tasso di assenza (gg assenza/gg lavorati nell'anno)	Nuovi accantonamenti per credito/num.nuove cause tot.
Nuovi flussi/tot.crediti anno precedente(importi)	N.dirigenti/organico	Nuovi accantonamenti per titoli in default/num. nuove cause tot.
Nuovi flussi/tot.crediti anno precedente(teste)	N. quadri direttivi/organico	Nuovi accantonamenti per altro/num. nuove cause tot.
Wealth Management	N. aree professionali/organico	
Volumi GP Azionarie	N. strutture centrali/organico	
Volumi GP Bilanciate	Nuove cause contenz passivo/organico	
Volumi GP Obbligazionarie	Spese del personale ind.per infl/organico	
Volumi GP Total Return		
Volumi GP Liquidità		
	Mercato	Macroeconomico
	VaR PNV	Pil
	Deviazione standard VaR PNV	Disoccupazione
	Volumi Hedge Fund	Retribuzione
		Business confidence
		Consumer confidence
		Prezzi al consumo
		Immobiliare
		Produzione industriale
		Tasso di cambio Euro-Dollaro
		Tasso Zero coupon-3m
		Vendite al consumo
		Fatturato industriale vendite
		Ordini industriali
		Tasso Zero coupon-10y
		CDS Italia

Figura 20 informazioni individuate nei vari comparti

Può accadere che comparti diversi possono includere indicatori insistenti sullo stesso fenomeno. Non esiste, tuttavia, sovrapposizione tra informazioni raccolte nei diversi comparti, in quanto ogni indicatore riporta informazioni specifiche del comparto e differenti dagli altri.

Nell'approccio utilizzato si verifica che nel comparto Compliance abbiamo l'indicatore *percentuale di personale a zero ore di formazione* e nel comparto Risorse Umane *Ore formazione/organico* che pur insistendo sullo stesso fenomeno (la formazione del personale) esaminano aspetti diversi. Il primo rileva i dipendenti che non sono stati raggiunti dalla formazione nell'anno in analisi, che può essere quindi considerato un indicatore della quota parte di dipendenti che non hanno ricevuto nessun intervento formativo nell'anno. Il secondo indice, invece, indica le ore di formazione procapite fruite nell'anno, fornendo quindi un altro tipo di informazione. Potrebbe accadere ad esempio che, pur aumentando le ore di formazione media procapite, tuttavia l'incremento sia relativo solo ad un particolare business della banca, ma la quota parte di dipendenti che non ha ricevuto alcun tipo di formazione rimanga invariata.

Per quanto riguarda il comparto Mercato va precisato che due indicatori insistono sullo stesso fenomeno: il VaR medio trimestrale e la deviazione standard trimestrale del VaR. Tuttavia si tratta di indicatori che forniscono informazioni molto diverse relative all'andamento del VaR di mercato. Da una parte viene misurato il livello di Value at Risk, dall'altra viene misurata la sua variazione in un dato arco temporale: per stessi livelli di VaR medio può infatti accadere che la variabilità osservata nel periodo possa essere molto diversa.

4.3.1.1 La costruzione degli indicatori

Per individuare un indicatore “qualitativo” (*score*) che indichi il “grado” di discontinuità/continuità rispetto al passato, l’analisi si è basata sull’osservazione dell’andamento storico dei dati. A partire da questi si è cercato di utilizzare una metodologia comune per i vari comparti, con lo scopo di individuare la “presenza” di discontinuità dell’ultimo anno osservato rispetto alla storia passata.

Per ciascun indicatore si calcola quindi la variazione rispetto all’anno precedente e si valuta se la variazione osservata nell’ultimo anno risulta in linea con le variazioni annuali osservate in passato. Nel caso in cui la variazione non risulti in linea con la storia passata, l’indicatore sta segnalando una discontinuità.

Per stabilire se siamo in presenza di un *break* strutturale o meno, la variazione è stata confrontata con la deviazione standard calcolata sull’intera serie storica delle variazioni attribuendo un valore dicotomico (0,1) ad ogni indicatore, a seconda che si verifichi o meno una discontinuità.

Di seguito si descrive nel dettaglio la metodologia adottata.

Bisogna innanzitutto precisare che i dati forniti per il calcolo degli indicatori sono misurati su orizzonti temporali diversi (annuali, trimestrali, mensili e giornalieri). È stato deciso di riportare tutti gli indicatori su base trimestrale, in modo da rendere il trattamento omogeneo. Unica eccezione è rappresentata dagli indicatori del comparto Risorse Umane, misurati su base annuale.

Per gli indicatori misurati su base giornaliera (come i dati di mercato e di liquidità), o su base mensile (come i dati della pianificazione e del wealth management) i dati vengono quindi resi su base trimestrale, semplicemente facendo una media dei dati di origine.

Si calcola quindi la variazione degli indicatori misurati in istanti temporali diversi utilizzando la differenza. Nel dettaglio, la metodologia adottata consiste nella costruzione di una variabile z , che rappresenta il confronto tra il valore che il dato assume ad un trimestre t e quello assunto nello stesso trimestre dell’anno precedente²⁵:

$$z = y(t) - y(t-4)$$

Una volta calcolata la variabile z , deve essere definito se la variazione osservata tra il valore assunto al tempo t e quello al tempo precedente risulta significativa, ossia se la variazione osservata sui dati riflette effettivamente una situazione di discontinuità con il passato.

A tal fine la variazione viene confrontata con la deviazione standard di tutte le differenze osservate e viene considerata significativa se la variabile z , data dalla differenza $y(t)-y(t-4)$, supera, in valore assoluto, il valore della deviazione standard delle differenze.

Per gli indicatori in cui la profondità della serie storica non è sufficientemente estesa per poterne utilizzare la deviazione standard, è stata considerata una sua proxy. Facendo una semplice

²⁵ Per quanto riguarda gli indicatori del comparto Risorse Umane, unica eccezione espressa su base annuale, la variabile z è definita come differenza tra l’indicatore all’anno T e il valore assunto all’anno precedente: $z = y(T) - y(T-1)$.

assunzione, si ipotizza che la deviazione standard sia pari al 20% della media dei valori della variabile osservata²⁶.

Applicando la metodologia descritta, per ogni indicatore, confrontando la variabile *z* con la deviazione standard, sono stati quindi ottenuti valori dicotomici, che assumono valore 1 se la variazione risulta significativa o valore 0 nel caso contrario.

Trattandosi di dati trimestrali, per ogni anno si dispone quindi di 4 variabili dicotomiche.

L'indicatore di continuità/discontinuità dell'anno in esame è ottenuto come media delle 4 variabili dicotomiche trimestrali. Un esempio esplicativo è riportato in appendice.

Per il comparto Risorse Umane, unica eccezione di dati misurati direttamente su base annuale, l'indicatore di continuità/discontinuità è direttamente la variabile dicotomica che assume valore 1 nel caso in cui la variazione in valore assoluto è superiore alla deviazione standard.

4.3.1.2 Gli indicatori per comparto

In base al paragrafo precedente, per ciascuna variabile in esame è stato calcolato l'indicatore di continuità/discontinuità.

È possibile quindi calcolare un indicatore di continuità/discontinuità sintetico per ciascun comparto, considerando tutti gli indicatori individuati all'interno del comparto.

Il numero di discontinuità evidenziate rispetto al numero totale degli indicatori del comparto fornisce un indicatore sintetico o “*score*” per il comparto in analisi. In pratica lo score è calcolato come media dei singoli indicatori di discontinuità del comparto.

Lo score è una valutazione qualitativa espressa su scala ordinale che assume valori compresi tra 0 ed 1, in tabella sono rappresentati gli score moltiplicati per 100.

COMPARTO	Indicatori 2014
COMPLIANCE	18%
WEALTH	69%
LIQUIDITA'	31%
MERCATO	15%
PIANIFICAZIONE	30%
RISORSE UMANE	13%
CREDITO	58%
LEGALE	80%
MACROECONOMICI	21%

Figura 21 Indicatori di continuità/discontinuità (dicembre 2014)

²⁶ La proxy è stata utilizzata per gli indicatori della Compliance, del Personale, del Wealth Management e del Legale, per un indicatore della Liquidità e uno del Mercato, e verrà utilizzata fin quando la serie storica non presenti una adeguata profondità.

4.3.2 La costruzione dei pesi per la componente di scenario

4.3.2.1 Indicatori di continuità/discontinuità per event type

Una volta calcolato l' indicatore di continuità/discontinuità per ciascun comparto, vengono identificati quali sono le classi di rischio per le quali ogni comparto può dare informazioni utili circa la continuità/discontinuità dell'*event type* (ET) in esame.

Per ogni classe di rischio quindi, su base esperienziale, vengono assegnati dei pesi per ciascun comparto informativo, tali che la somma dei pesi per ogni *event type* sia uguale a 100%. Nel caso in cui tutti i comparti siano ritenuti informativi per una data classe di rischio, si assume un uguale peso per ciascun comparto. Per alcuni *event type*, tuttavia, sempre su base *judgemental*, si ritiene corretto considerare informativi solo alcuni comparti, a cui possono essere assegnati pesi uguali oppure pesi diversi esclusivamente sulla base di un giudizio "esperto". Inoltre, in accordo con l'Autorità di Vigilanza, si è ritenuto corretto valorizzare la distinzione tra comparti conferenti in termini di manifestazioni del rischio (ad esempio il comparto Legale) rispetto a quelli con maggiore vocazione prospettica (ad esempio il comparto Macroeconomico). Tale distinzione si evidenzia nei pesi attribuiti alle classi di rischio legate a cause o reclami (ET4C, ET4R, ET7C, ET7R).

ET	compliance	wealth	liquidità	mercato	pianificazione	risorse umane	credito	legale	macroeconomici
ET1	0.17				0.17	0.17	0.17	0.17	0.17
ET2R			0.15			0.15			0.70
ET2CC					0.25			0.25	0.50
ET2AFE				0.20	0.20		0.20	0.20	0.20
ET2STP								0.50	0.50
ET3					0.20	0.40		0.20	0.20
ET4A	0.25						0.25	0.25	0.25
ET4R	0.60							0.20	0.20
ET4C	0.15							0.70	0.15
ET5	0.25		0.25						0.50
ET6	0.20				0.20	0.20		0.20	0.20
ET7C	0.15							0.70	0.15
ET7P	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11
ET7R	0.60							0.20	0.20

Figura 22 Pesi assegnati ai comparti per ciascun event type (dicembre 2014)

In particolare, per l'*event type* ET7 Partite diverse si ritengono informativi tutti i comparti ed a ciascuno viene assegnato un peso uguale (11%).

Relativamente ad alcune classi si ritengono informativi solo alcuni comparti, ma tutti con lo stesso peso:

- ✓ ET1: informazioni relative ai *break* strutturali sulle perdite operative connesse ad eventi di frodi possono essere contenute nel comparto Compliance o Legale, in quanto indicatori di non adeguatezza e di litigiosità, e quindi indirettamente possono essere indicatori di atteggiamenti fraudolenti; variazioni nel comparto Pianificazione e Credito possono essere dovute a variazioni nell'attività di business e nella struttura organizzativa che potrebbero rendere più agevole eventuale attività illecita; variazioni significative nella composizione del personale (comparto Risorse Umane) possono avere un impatto sulla

soddisfazione dei dipendenti e sul relativo comportamento; i cicli economici possono condizionare i trend di criminalità.

- ✓ ET2 Altre Frodi Esterne: le informazioni possono essere contenute nel comparto Mercato e Pianificazione, variazioni nel business, monitorate in modo non opportuno potrebbero esporre la banca a frodi; variazioni nel comparto Credito possono essere dovute a un peggioramento del merito creditizio dei clienti con eventuale incremento nelle attività fraudolente; il comparto Legale può essere informativo relativamente alla litigiosità dei clienti; i cicli economici possono condizionare i trend di criminalità.
- ✓ ET2 Sistemi Tradizionali di Pagamento: il comparto Legale può essere informativo relativamente alla litigiosità dei clienti mentre i cicli economici possono condizionare i trend di criminalità.
- ✓ ET4 Anatocismo: i comparti Compliance e Legale, possono essere indicatori di contestazioni sul tema. Considerando che le azioni legali relativamente all'anatocismo si evidenziano su clientela in sofferenza, il comparto Credito può risultare informativo, allo stesso modo del comparto Macroeconomico.
- ✓ ET6: i comparti Compliance e Legale, possono essere indicatori di contestazioni sul tema per non adeguatezza nei processi legate a problematiche sui sistemi; variazioni nel comparto Pianificazione possono essere dovute a variazioni nell'attività di business e nella struttura organizzativa che potrebbero essere soggette a problematiche di tipo IT; variazioni significative nella composizione del personale (comparto Risorse Umane) possono avere un impatto sull'operatività ed i processi IT; i cicli economici possono eventualmente essere informativi di eventuali politiche di cost cutting nei sistemi/controlli.

Per quanto riguarda altre classi di rischio, si ritengono informativi solo alcuni comparti, ai quali si attribuiscono pesi diversi, sulla base di un giudizio "esperto". In questo caso valgono le seguenti considerazioni:

- ✓ ET2 Rapine: informazioni relative ai *break* strutturali sulle perdite operative connesse ad eventi di rapina possono essere contenute principalmente nel comparto Macroeconomico. I cicli economici possono infatti condizionare i trend di criminalità. Inoltre, si può ritenere plausibile che variazioni significative nella composizione del personale (comparto Risorse Umane) o nella situazione di liquidità della banca possano essere incentivi/disincentivi alla delinquenza.
Per tale motivo, si ritiene assegnare in modo judgemental un peso del 70% al comparto Macroeconomico ed in misura marginale ai comparti Risorse Umane (15%) e Liquidità (15%).
- ✓ ET2 Carte Clonate: come sopra, i cicli economici possono condizionare i trend di criminalità e quindi il comparto Macroeconomico può essere ritenuto come maggiormente informativo soprattutto in ottica prospettica (peso 50%). Inoltre si può ritenere plausibile che variazioni significative nel comparto Legale (come indicatore di maggiore litigiosità della clientela) e nel comparto Pianificazione (come indicatore di variazione di operatività della banca) possano risultare informative in modo marginale. Pertanto si attribuisce ad entrambe un peso pari al 25%.

- ✓ ET3: per quanto riguarda i rapporti di impiego e sicurezza sul lavoro è banale considerare il comparto Risorse Umane come maggiormente informativo (peso 40%), allo stesso tempo si possono considerare informativi della litigiosità sia il ciclo economico (comparto Macroeconomico, con peso 20%) che il comparto Legale (peso 20%), mentre il comparto Pianificazione (peso 20%) può fornire informazioni su variazioni significative del business aziendale.
- ✓ ET4Reclami e ET7Reclami: il comparto Compliance è considerato principalmente informativo (con peso del 60%) relativamente ai reclami di clienti e ai reclami inerenti l'esecuzione, consegna e gestione dei processi mentre Legale e Macroeconomico forniscono informazioni sulla litigiosità con uguale peso, pari al 20%.
- ✓ ET4C e ET7C: il comparto Legale è considerato particolarmente informativo (con peso del 70%) relativamente alle cause di clienti e alle cause inerenti l'esecuzione, consegna e gestione dei processi mentre Compliance e Macroeconomico forniscono informazioni di litigiosità con uguale peso, pari al 15%.
- ✓ ET5: il comparto Macroeconomico, in quanto indicatore del ciclo economico, può essere considerato informativo (peso 50%) relativamente ai danni ai beni materiali, ad esempio nel caso di vandalismi; problemi di liquidità (peso 25%) potrebbero essere visti come anticipatori di criticità nella manutenzione dei beni, così come il comparto Compliance, con peso 25%.

Lo schema dei pesi assegnati ai vari comparti per ciascuna classe di rischio viene archiviato sull'ambiente ASIA insieme alle giustificazioni relative alle singole scelte effettuate per la sua definizione. Eventuali modifiche nei pesi possono essere valutate in corrispondenza della trimestrale di dicembre, dandone opportuna informativa all'Autorità di Vigilanza.

In base al valore assunto dall'indicatore di comparto e dal peso judgemental assegnato, viene quindi calcolato uno *score*, indicatore di discontinuità/continuità per ogni *event type*:

$$score_{ETj} = \sum_i^q p_{ij} score_i$$

dove p_{ij} è il peso assegnato all'i-esimo comparto relativamente al j-esimo *event type*, mentre $score_i$ è lo score assegnato all'i-esimo comparto. Lo *score* può assumere valori compresi tra 0 e 100 (vedi Figura 23).

EVENT TYPE		score
ET1	Frode interna	37
ET2R	Frode esterna-Rapine	21
ET2CC	Frode esterna-Carte clonate	38
ET2AFE	Frode esterna-Altre Frodi Esterne	41
ET2STP	Frode esterna-Sistemi Tradizionali di Pagamento	51
ET3	Rapporto di impiego e sicurezza sul lavoro	31
ET4A	Clienti-Anatocismo	44
ET4R	Clienti-Reclami	31
ET4C	Clienti-Cause	62
ET5	Danni a beni materiali	23
ET6	Perdite dovute a interr operatività o a disf. sist. inf.	32
ET7C	Esec, consegna e gest dei processi-Cause	62
ET7P	Esec, consegna e gest dei processi-Partite diverse	37
ET7R	Esec, consegna e gest dei processi-Reclami	31

Figura 23 Score per event type (dicembre 2014)

4.3.2.2 Assegnazione dei pesi per event type

Una volta calcolato un indicatore di continuità/discontinuità (*score*) per ciascun *event type*, è necessario trasformare questo valore nel peso *w* che deve assumere la componente di scenario. Lo *score* infatti può assumere valori compresi tra 0 e 100, mentre il peso della componente qualitativa può variare in un *range* prefissato, pari a 10%-40%.

Per motivi di stabilità di modello, si fissano quindi 7 possibili valori che il peso della componente qualitativa può assumere:

10% - 15% - 20% - 25% - 30% - 35% - 40%

In pratica il peso *w* può assumere solo questi 7 valori.

Per attribuire allo *score* uno di questi possibili valori si procede con una mappatura (“*mapping*”) secondo quanto indicato in Figura 24:

indicatore		peso
0	30	10%
30	40	15%
40	45	20%
45	50	25%
50	55	30%
55	65	35%
65	100	40%

Figura 24 mapping

A ciascuna classe di rischio viene attribuito così un *peso* (Figura 25):

EVENT TYPE		score	peso
ET1	Frode interna	37	15%
ET2R	Frode esterna-Rapine	21	10%
ET2CC	Frode esterna-Carte clonate	38	15%
ET2AFE	Frode esterna-Altre Frodi Esterne	41	20%
ET2STP	Frode esterna-Sistemi Tradizionali di Pagamento	51	30%
ET3	Rapporto di impiego e sicurezza sul lavoro	31	15%
ET4A	Clienti-Anatocismo	44	20%
ET4R	Clienti-Reclami	31	15%
ET4C	Clienti-Cause	62	35%
ET5	Danni a beni materiali	23	10%
ET6	Perdite dovute a interr operatività o a disf. sist. inf.	32	15%
ET7C	Esec, consegna e gest dei processi-Cause	62	35%
ET7P	Esec, consegna e gest dei processi-Partite diverse	37	15%
ET7R	Esec, consegna e gest dei processi-Reclami	31	15%

Figura 25 Attribuzione peso alle classi di rischio (dicembre 2014)

Al fine di incrementare la reattività del peso della componente qualitativa, il peso così determinato viene rettificato per tenere conto di eventuali variazioni significative osservate sulla componente di scenario sia rispetto allo scenario dell'anno precedente sia rispetto alla componente quantitativa.

In particolare, viene aggiunto un add-on al peso nel caso in cui il VaR della componente qualitativa presenti una variazione maggiore del 50% rispetto allo scenario dell'anno precedente o rispetto alla componente storica.

L'add-on è definito semplicemente aggiungendo 5% al peso precedentemente calcolato.

In Figura 26 sono riportati i pesi post add on, w , calcolati a partire dai pesi post mapping per ogni event type:

EVENT TYPE	peso	>50%	peso post add on
ET1	15%	SI	20%
ET2 Altre Frodi Esterne	20%	NO	20%
ET2 Rapine	10%	SI	15%
ET2 Carte clonate	15%	NO	15%
ET2 Sistemi Trad. di Pagamento	30%	NO	30%
ET3	15%	NO	15%
ET4 Anatocismo	20%	SI	25%
ET4 Cause	35%	SI	40%
ET4 Reclami	15%	NO	15%
ET5	10%	NO	10%
ET6	15%	SI	20%
ET7 Cause	35%	NO	35%
ET7 Partite diverse	15%	NO	15%
ET7 Reclami	15%	SI	20%

Figura 26 Attribuzione peso w post add on (dicembre 2014)

A titolo esemplificativo riportiamo di seguito il processo con il quale avviene l'attribuzione del peso per l'event type 2 rapine (ET2R).

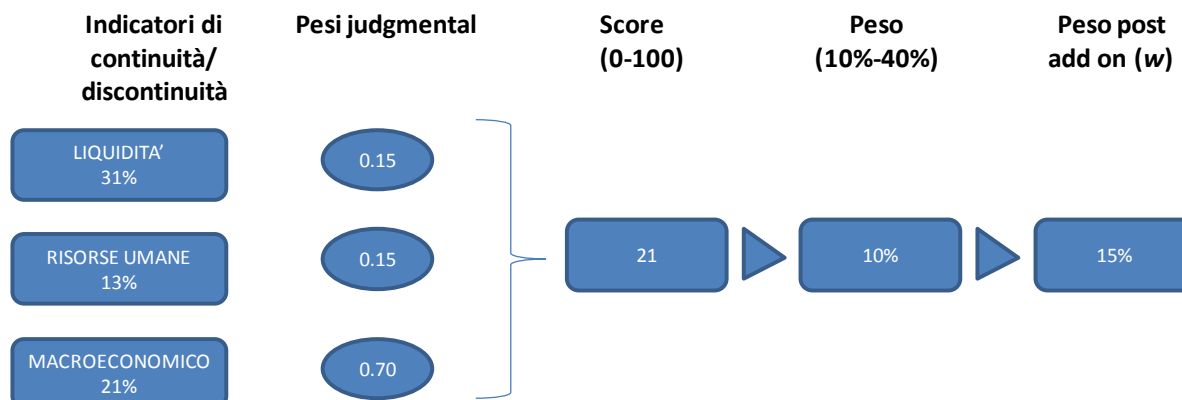


Figura 27 Attribuzione del peso w all'ET2R

Nell'esempio riportato per l'ET2R, al peso del 15% si arriva considerando i 3 comparti informativi per tale event type (Liquidità con peso 15%, Risorse Umane con peso 15% e Macroeconomico con peso 70%) che producono uno score pari a 21. In base al mapping lo score si traduce in un peso del 10%, che viene rettificato per tenere in considerazione la variazione significativa osservata sulla componente di scenario. Al peso del 10% si aggiunge pertanto il 5% dell'add-on. In conclusione il peso da utilizzare per la mistura per tale *event type* sarà pari al 15%.

4.3.3 Calcolo della distribuzione mistura per l'integrazione tra la componente storica e la componente di scenario

L'obiettivo del metodo descritto in questa sezione è di produrre una stima del requisito, che chiameremo "requisito integrato", che combini le informazioni contenute nei VaR qualitativo e storico. Il metodo è costruito in modo da rispettare il vincolo che il requisito integrato sia compreso tra il VaR storico ed il VaR qualitativo:

1. sia individualmente per ciascuna classe di rischio del modello AMA;
2. sia collettivamente per il VaR aggregato (ottenuto come somma dei VaR delle singole classi).

Per ciascuna classe di rischio i VaR storico e qualitativo vengono calcolati a partire da due distribuzioni di perdite operative annuali:

- la distribuzione storica o LDA, denotata con L;
- la distribuzione qualitativa, denotata Q;

Si denota inoltre con X la distribuzione integrata, determinata a partire da L, Q, e con $V(L)$, $V(Q)$, $V(X)$ i VaR delle tre distribuzioni. Il metodo adottato è stato costruito in modo da rispettare il vincolo che

$$V(Q) \leq V(X) \leq V(L), \quad \text{se } V(Q) \leq V(L),$$

oppure

$$V(Q) \geq V(X) \geq V(L), \quad \text{se } V(Q) \geq V(L).$$

Inoltre, la medesima proprietà vale anche per le somme dei VaR delle singole classi di rischio. Si propone pertanto di costruire la distribuzione integrata come mistura delle distribuzioni storica e qualitativa:

$$X_p = p * Q + (1 - p) * L$$

Questa formula esprime la convinzione che la distribuzione delle perdite future venga ben descritta da una combinazione delle perdite di scenario in proporzione di p e delle perdite storiche in proporzione di $1 - p$. Pertanto il parametro p , che chiamiamo peso di mistura, è il peso della distribuzione qualitativa sulla distribuzione integrata X_p ; quest'ultima dipende quindi dal valore fissato per tale parametro. Con tale definizione viene rispettato il vincolo che il requisito integrato sia compreso tra i VaR qualitativo e quantitativo. Per pesi piccoli (p prossimo a zero) il VaR integrato è vicino al VaR storico, mentre per pesi grandi (p prossimo ad 1) è vicino al VaR qualitativo.

Nella simulazione numerica si procede come segue: denotiamo le distribuzioni empiriche storica e qualitativa, simulate mediante Monte Carlo, con

$$L_{\text{camp}} = \{l_k, k = 1, \dots, n\}, \quad Q_{\text{camp}} = \{q_k, k = 1, \dots, n\};$$

un campione della distribuzione integrata viene generato unendo

- $n * p$ elementi estratti casualmente da Q_{camp} con
- $n * (1 - p)$ elementi estratti casualmente da L_{camp} .

Le proprietà del metodo garantiscono che per ogni p fissato il VaR integrato sia una media pesata dei VaR storico e qualitativo, ossia risulta:

$$V(X_p) = w_{\text{eff}} * V(Q) + (1 - w_{\text{eff}}) * V(L)$$

dove w_{eff} , definito come peso effettivo (del VaR qualitativo sull'integrato), è dato da:

$$w_{\text{eff}} = \frac{V(X_p) - V(L)}{V(Q) - V(L)}.$$

Si osserva che:

1. il peso effettivo sul VaR è ben definito se $V(Q) \neq V(L)$, ovvero i VaR qualitativo e storico sono diversi; in tal caso w_{eff} è un numero compreso tra 0 ed 1 e dipende dal peso di mistura p ;
2. il peso effettivo $w_{\text{eff}} = w_{\text{eff}}(p)$ non è necessariamente uguale al peso di mistura p ;
3. per p fissato w_{eff} non è necessariamente uguale per due classi di rischio distinte.

Si procede quindi fissando il peso effettivo sul requisito w_{eff} e determinando, classe per classe, il peso di mistura p a cui corrisponde il peso effettivo prefissato: in termini matematici, fissiamo w_{eff} e calcoliamo p invertendo la funzione $w_{\text{eff}} = w_{\text{eff}}(p)$. Per ciascuna classe ETi:

- si fissa $\overline{w_{\text{eff},i}}$ al valore ottenuto per l'indicatore di continuità di Sezione 4.3.2.;
- si calcola il peso di mistura p_i tale che il corrispondente peso effettivo sia approssimativamente pari a $\overline{w_{\text{eff},i}}$ ovvero:

$$V(X_{p_i}, \text{ETi}) \approx \overline{w_{\text{eff},i}} * V(Q, \text{ETi}) + (1 - \overline{w_{\text{eff},i}}) * V(L, \text{ETi}).$$

Questa equazione non ha in generale una soluzione analitica: la soluzione viene approssimata numericamente con un metodo iterativo di bisezione. Pertanto, il peso effettivo finale può risultare non esattamente identico al valore $\overline{w_{eff,i}}$ prefissato. In conclusione, con il metodo di integrazione presentato il requisito integrato di ciascuna classe, $V(X_{p_i}, ET_i)$, si ottiene come media ponderata dei requisiti qualitativo e storico con un peso effettivo $\overline{w_{eff,i}}$ molto vicino a quello prefissato sulla base dell'indicatore di continuità presentato in Sezione 4.3.2.

Nel caso che si manifestino problemi di convergenza dell'algoritmo iterativo di bisezione, si procede come segue, basandosi su un principio di prudenzialità. L'algoritmo parte da due valori $p_{inf} = 0$ e $p_{sup} = 1$ del peso di mistura, sapendo che il valore ricercato giace tra 0 ed 1; ad ogni passo si aumenta p_{inf} oppure si diminuisce p_{sup} fino a soddisfare l'equazione sopra riportata a meno di una tolleranza specificata in partenza. Se l'algoritmo non riesce a raggiungere la tolleranza prefissata, si calcolano i due valori del VaR integrato corrispondenti a p_{inf} e p_{sup} e si pone il peso di mistura pari a p_{inf} se questo genera il VaR massimo tra i due e p_{sup} in caso contrario. In questo modo la scelta risulta prudentiale.

4.4 Integrazione tra event type: approccio t-copula

Il requisito regolamentare della Banca fino al 2013 è stato calcolato come somma dei VaR dei singoli Event Type, ipotizzando quindi una perfetta correlazione tra le classi di rischio.

La presenza di fattori comuni idiosincratici o di contesto esterno possono determinare fenomeni di correlazione tra le perdite, ma risulta difficile sostenere l'ipotesi della presenza di correlazione perfetta.

A partire da dicembre 2013, pertanto, il Gruppo Montepaschi ha adottato, coerentemente con quanto stabilito dalla normativa prudenziale e successivamente all'autorizzazione di Banca d'Italia²⁷, un modello di integrazione tra i diversi *event type* che tiene conto degli effetti di diversificazione dovuti alla non perfetta correlazione tra le classi di rischio. L'approccio di integrazione adottato utilizza una copula *t*-Student²⁸.

Le funzioni copula sono utilizzate per descrivere la struttura di dipendenza tra diverse variabili casuali. Il principale vantaggio di utilizzare una funzione copula è quello di poter modellare la struttura di dipendenza tra variabili casuali senza dover fare assunzioni sulle distribuzioni marginali.

Per catturare in maniera adeguata la dipendenza tra variabili a quantili alti, nota come dipendenza di coda o *tail dependence*, l'approccio prevalente in letteratura si basa sull'utilizzo della copula *t*-Student con n gradi di libertà.

²⁷ Autorizzazione comunicata con lettera n.0083034/14 del 24/01/2014, a valere sulle segnalazioni riferite al 31.12.2013.

²⁸ La metodologia adottata è analoga a quanto applicato in ambito di integrazione rischi per il calcolo del capitale economico.

L'uso della copula *t*-Student per l'integrazione dei diversi *event type* basa la propria fondatezza sulla capacità di cogliere non solo la correlazione lineare ma anche la correlazione di coda, in quanto il modello stima la misura di VaR tenendo in considerazione non solo i fenomeni lineari ma anche i fenomeni correlativi a quantili elevati. Il manifestarsi di eventi estremi, infatti, può modificare la natura delle correlazioni a quantili elevati e condurre ad una diversa calibrazione della matrice di correlazione (input della funzione copula) in modo che la funzione copula tenga adeguatamente conto delle correlazioni di coda verificatesi. Si rinvia al paragrafo 5.4 per una presentazione concisa della teoria e delle proprietà della copula *t*-Student, che per semplicità definiremo *t*-copula nel seguito, nonché per la notazione utilizzata nelle sottosezioni seguenti.

Si precisa, inoltre, che il Gruppo Montepaschi utilizza una cosiddetta *meta-copula t-Student*.

La copula *t*-Student per definizione descrive infatti la struttura di dipendenza di variabili casuali che presentano distribuzioni marginali *t*-Student esse stesse. Nell'approccio Montepaschi, invece, le marginali non sono distribuite come *t*-Student ma si utilizzando le distribuzioni empiriche ottenute dalle simulazioni MonteCarlo delle perdite annue per ciascuna classe di rischio, utilizzate per il calcolo del VaR per *event type*.

L'algoritmo adottato dal Gruppo Montepaschi per la stima della *t*-copula è strutturato nei seguenti passi:

- 1) Input alla meta-*t*-copula;
- 2) Stima empiriche dei parametri di interesse: tau di Kendall e dipendenza di coda;
- 3) Prima esecuzione del "metodo di discesa" per la perturbazione del tau di Kendall;
- 4) Seconda esecuzione del "metodo di discesa" con matrice di dipendenza di coda perturbata;
- 5) Simulazione congiunta e calcolo del VaR complessivo diversificato;
- 6) Riallocazione del beneficio di diversificazione sui singoli Event Types: approccio Expected Conditional Loss (EcL).

4.4.1 Input alla meta-*t*-copula

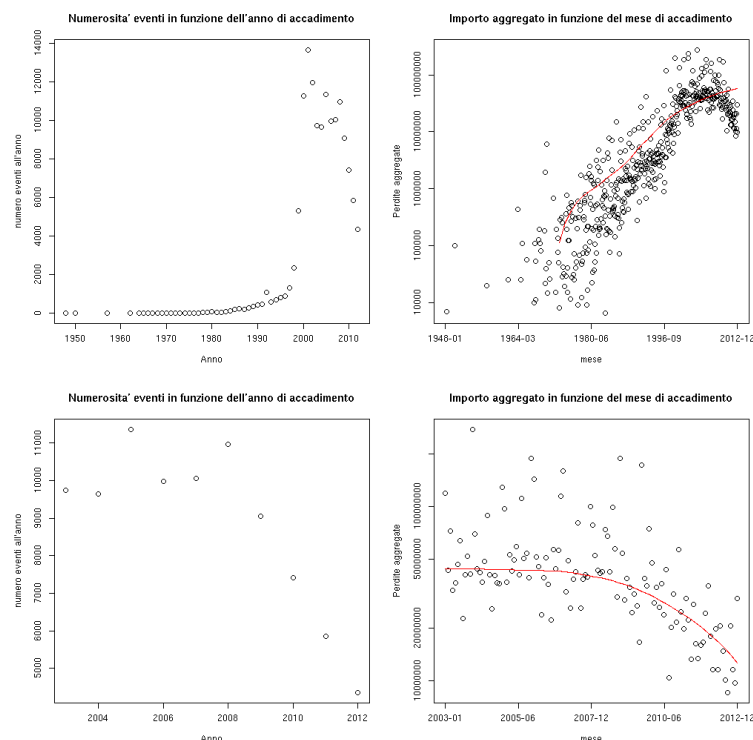
Per seguire un approccio *t*-copula è necessario stimare:

- la matrice di correlazione
- il numero di gradi di libertà
- le distribuzioni marginali

Relativamente alla stima dei parametri della *t*-copula è stato scelto di utilizzare l'intera base dati costituita dai dati interni ed esterni. Le ragioni di tale scelta si basano sul fatto che:

- la base dati utilizzata è coerente con il dataset su cui si basa la metodologia di calcolo del requisito patrimoniale,
- una serie storica più ricca di dati, ottenuta includendo anche i dati esterni, fornisce maggiore robustezza al modello;
- le stime della matrice di correlazione (da utilizzare nella copula) non risultano molto diverse da quelle ottenute utilizzando i soli dati interni, come verificato durante le analisi preliminari.

Relativamente al tipo di data da utilizzare (e.g. data di accadimento anziché di rilevazione), la scelta si è basata sulle analisi delle caratteristiche delle serie storiche. Utilizzando la data di accadimento è risultato infatti evidente che sia gli importi aggregati mensilmente, sia il numero annuale degli eventi mostrano una marcata non-stazionarietà, con pochi dati prima del 2000, seguiti da un rapido incremento con un picco ed una forte riduzione a partire dal 2009, come mostrato nella seguente figura:



In ragione della marcata non-stazionarietà, è stato deciso di basare la stima dei parametri della copula t-Student sulla serie storica costruita sulla data di rilevazione, coerentemente con le metodologie adottate per il modello di calcolo.

I dati vengono classificati in base ai 7 Event Type di primo livello di Basilea II e non in base alle classi di rischio del modello interno. In questo modo otteniamo una copula di dimensione più bassa²⁹. Si ha inoltre il vantaggio aggiuntivo che la struttura e la dimensione della copula non cambiano nel caso che venga modificata la granularità del modello AMA.

²⁹ Al 31 dicembre 2013 otteniamo una copula di dimensione 7 al posto di 14.

Come distribuzioni marginali si considerano le singole distribuzioni empiriche simulate mediante Monte Carlo per il calcolo del VaR di ciascun event type. Si utilizza infatti un approccio a meta-*t*-copula, dove le marginali non sono distribuite come *t*-Student, ma mantengono le loro distribuzioni originarie, ottenute dalle simulazioni Monte Carlo.

Per ottenere le 7 distribuzioni marginali per ciascun event type di primo livello, nel caso in cui una classe di rischio presenti più sottoclassi (come ad esempio l'ET2, con sottoclassi Carte Clonate, Rapine, Sistemi tradizionali di pagamento, Altre Frodi Esterne) si procede aggregando le serie di perdite simulate per ciascuna sottoclasse sotto ipotesi di correlazione perfetta. Si ottengono quindi 7 distribuzioni di perdita in corrispondenza di ciascuna classe di rischio.

4.4.2 Stima empiriche dei parametri di interesse

La copula *t*-Student viene calibrata in funzione della matrice di correlazione lineare, del numero di gradi di libertà e della matrice di tail dependence.

La stima dei parametri ρ_{ij} (matrice di correlazione) ed n (numero di gradi di libertà) della copula *t*-Student possono essere stimati con il metodo di massima verosimiglianza, tuttavia l'applicazione diretta di tale metodo può essere resa problematica dall'elevato numero di parametri liberi nella matrice di correlazione lineare, pari a $d(d-1)/2$, dove d è la dimensione della *t*-copula.

Per questo motivo si ricorre ad un approccio semi-parametrico.

Si stima innanzitutto la matrice di correlazione a ranghi τ di Kendall. Da tale matrice di correlazione a ranghi si ricava quindi la matrice di correlazione lineare implicita mediante lo stimatore di Linskov:

$$\rho_{ij} = \sin\left(\frac{\pi}{2} \cdot \tau_{ij}\right)$$

Si stima inoltre la matrice di dipendenza di coda empirica, cioè la dipendenza a quantili elevati osservata direttamente nelle serie storiche delle perdite effettive dei diversi event type, mediante il seguente stimatore consistente³⁰:

$${}_k\lambda_{xy}^{empirica} = \frac{1}{k} \sum_{h=1}^m 1(rango(x_h) > m-k, rango(y_h) > m-k),$$

dove:

- x_h e y_h sono le perdite osservate per le tipologie di event type X e Y;
- k rappresenta la soglia che divide le perdite estreme da quelle non estreme;

³⁰ R. Schmidt e U. Stadtmüller, Non parametric estimation of tail dependence. The Scandinavian Journal of Statistics 33 (2006).

- m rappresenta il numero di osservazioni;
- $1(A, B)$ rappresenta la funzione che restituisce 1 solo nel caso in cui A e B siano verificate contemporaneamente, altrimenti la funzione assume valore nullo.

L'utilità di tale stimatore si basa sulla sua indipendenza da ipotesi sulle distribuzioni multivariate o dalle copule adottate e dal fatto che esso può essere ottenuto direttamente dai dati empirici di perdita. Seguendo un approccio già utilizzato in ambito integrazione rischi³¹, la matrice di dipendenza di coda viene calcolata come

$$\lambda_{xy} = \frac{1}{\Delta} \sum_{k=1}^{\Delta} \lambda_{xy}^k, \quad \Delta = \sqrt{m}$$

per ciascuna coppia di classi di rischio.

4.4.3 Prima esecuzione del "metodo di discesa"

Al fine di tenere nella massima considerazione la dipendenza di coda presente nei dati, adottiamo una procedura prudentiale. L'idea è quella di utilizzare una t -copula che possa spiegare la correlazione di coda osservata sui dati, pertanto lo scopo è quello di individuare i parametri della t -copula in modo tale che la matrice di dipendenza di coda implicita alla copula minimizzi la distanza dalla matrice di coda empirica.

Una volta stimate le matrici di correlazione a ranghi e di coda, si applica una procedura di perturbazione progressiva della matrice τ di Kendall, che chiameremo metodo di discesa, mirata ad individuare il numero di gradi di libertà e la matrice di correlazione lineare (definita positiva) che meglio possano spiegare la dipendenza di coda nei dati osservati.

Si procede secondo i seguenti passi:

- a) Si fissa innanzitutto una griglia di valori per il numero di gradi di libertà n ;
- b) per un dato valore di n si effettua una serie di perturbazioni Monte Carlo della matrice τ di Kendall. Tali perturbazioni vengono estratte casualmente da una matrice varianza-covarianza con correlazione fissa al 20% per tutti i coefficienti.
- c) Ad ogni passo, data la matrice attuale (perturbata) di correlazione a ranghi, si stima la corrispondente matrice di correlazione lineare implicita mediante la formula di Lindskog ed, utilizzando il numero di gradi di libertà n fissato all'inizio, si stima una matrice di dipendenza di coda implicita mediante la formula di Embrechts:

³¹ Vedasi anche G.Frahm, M.Junker e R.Schmidt, *Estimating the tail-dependence coefficient: properties and pitfalls*, marzo 2006 (par. 4.4, pag.15).

$$\lambda_{ij}^{implicita} = 2t_{n+1} \left(-\sqrt{(n+1) \cdot \left(\frac{1-\rho_{ij}}{1+\rho_{ij}} \right)} \right)$$

dove t_n è la funzione di distribuzione t-Student.

- d) Si calcola la distanza tra la matrice di dipendenza di coda implicita ottenuta al passo precedente e la matrice di dipendenza di coda empirica: la nuova matrice perturbata viene accettata se la distanza diminuisce.

Al termine del processo si ottiene una t-copula la cui matrice di dipendenza di coda implicita ha distanza minima dalla matrice di dipendenza di coda empirica, dato il numero di gradi di libertà n considerato.

Questo processo viene ripetuto variando n in una griglia di valori e si seleziona infine la t-copula (con corrispondente matrice di correlazione e numero di gradi di libertà) per cui la distanza tra le matrici di dipendenza di coda implicita ed empirica è minima.

4.4.4 Seconda esecuzione del "metodo di discesa"

A fini prudenziali, si applica una perturbazione alla matrice di dipendenza di coda empirica λ stimata nella sezione precedente, usando l'incertezza delle simulazioni Monte Carlo effettuate durante il metodo di discesa. Sia σ la media aritmetica delle deviazioni standard esibite dai coefficienti di correlazione lungo il metodo di discesa: si considera la matrice di dipendenza di coda empirica perturbata data da

$$\lambda + 1.96 \sigma.$$

Si effettua infine un seconda applicazione del metodo di discesa, minimizzando la distanza delle matrici di dipendenza di coda implicite dalla matrice perturbata così definita, seguendo l'approccio descritto nel paragrafo precedente.

Al termine di questo processo, disponiamo di una t-copula data da una matrice di correlazione lineare ed un numero di gradi di libertà per i quali è minima la distanza della corrispondente matrice di dipendenza implicita di coda dalla matrice di dipendenza di coda empirica perturbata.

4.4.5 Calcolo del VaR complessivo diversificato

Una volta stimata la matrice di correlazione lineare ed il numero di gradi di libertà della t-copula, si procede alla simulazione della distribuzione multivariata delle perdite.

Questo approccio consente di generare congiuntamente 7 nuove serie che esibiscano le medesime caratteristiche delle distribuzioni marginali originali e che allo stesso tempo posseggano la struttura di dipendenza implicata dalla t-copula.

In particolare, al passo i-esimo della simulazione:

- Si estrae casualmente un valore dalla t-copula: tale valore è un vettore (u_1^i, \dots, u_7^i) di sette valori compresi tra 0 ed 1, che sono interpretati come probabilità;
- Dalle distribuzioni marginali simulate si calcolano i 7 percentili q_1^i, \dots, q_7^i corrispondenti alle probabilità (u_1^i, \dots, u_7^i) ;
- Si sommano i 7 percentili: $S_i = \sum_{k=1}^7 q_k^i$.

In altre parole, per ogni passo di simulazione si ottengono 7 valori simulati congiuntamente che vengono poi sommati, producendo una singola simulazione S^i di perdita aggregata. Dalla distribuzione

$$S = \{S_1, S_2, \dots, S_i, \dots, S_N\}$$

delle perdite aggregate viene estratto il percentile di interesse (99.9%), che costituisce il VaR complessivo diversificato.

Il beneficio di diversificazione è dato dalla differenza tra il VaR complessivo lordo (somma dei VaR delle singole classi di rischio interne del modello AMA) ed il VaR complessivo diversificato.

4.4.6 Allocazione del beneficio di diversificazione: *Expected Conditional Loss*

Una volta determinato il beneficio di diversificazione, è necessario attribuirne una quota a ciascuna tipologia di evento. La metodologia di allocazione del beneficio di diversificazione adottata dal Gruppo Montepaschi è ispirata alla nota teoria della *Conditional Expectation* ed è analoga a quella già utilizzata nell'ambito dell'integrazione rischi per il calcolo del capitale economico. In tale ambito, l'approccio è stato definito come *Expected Conditional Loss (EcL)*. La metodologia di riallocazione secondo EcL consente di determinare le percentuali di allocazione del requisito in maniera stabile, consistente e coerente.

Sia α il percentile di riferimento coerente con il *risk appetite* del Gruppo bancario. La misura di *Expected conditional Loss* per la k -esima classe di rischio è definita come la perdita attesa, condizionata al fatto che la perdita complessiva si trovi in un intorno del percentile α di semiampiezza δ :

$$EcL_k^\delta = E[L_k | VaR(\alpha - \delta) \leq L \leq VaR(\alpha + \delta)],$$

dove:

- L rappresenta la distribuzione congiunta aggregata delle perdite per tutti gli *event types*;
- L_k rappresenta la distribuzione di perdita relativa al k -esimo *event type*;

- $VaR(\alpha - \delta)$ è il percentile al livello di probabilità $\alpha - \delta$ estratto dalla distribuzione congiunta aggregata delle perdite.

La misura di EcL complessiva per tutti gli event types è data da:

$$EcL^\delta = E[L|VaR(\alpha - \delta) \leq L \leq VaR(\alpha + \delta)].$$

Fissata, dunque, la semiampiezza δ , e tenuta presente la proprietà additiva seguente:

$$EcL^\delta = \sum_k^7 EcL_k^\delta,$$

si definisce un vettore di 7 pesi di attribuzione del requisito:

$$\hat{\omega}_k^\delta = \frac{EcL_k^\delta}{EcL^\delta}, k = 1, \dots, 7.$$

Prudenzialmente, si fa variare δ in una griglia di valori $\delta_1, \delta_2, \dots, \delta_d$, ottenendo un insieme di vettori di pesi del quale si prende la media:

$$\hat{\omega}_k = \frac{\sum_{j=1}^d \hat{\omega}_k^{\delta_j}}{d}, k = 1, \dots, 7.$$

Questo vettore di pesi permette di riallocare il beneficio di diversificazione su ciascuna classe di rischio, ottenendo dei requisiti diversificati per ciascuna classe, la cui somma è il requisito diversificato complessivo.

4.5 Calcolo della detrazione per le perdite attese

A partire da dicembre 2013, successivamente all'autorizzazione di Banca d'Italia³², il Gruppo Montepaschi utilizza una tecnica di riduzione del requisito patrimoniale per i rischi operativi associata alla detrazione delle perdite attese.

La normativa prudenziale prevede che possano essere portate in detrazione dal requisito patrimoniale le perdite operative che rispettano simultaneamente le seguenti condizioni:

- ✓ *devono essere state coperte da accantonamenti specifici a conto economico*
- ✓ *devono essere individuate sulla base di criteri documentati e censite nel dataset di calcolo*

È stato quindi identificato un perimetro di “accantonamenti” fatti a bilancio, a fronte di eventi di rischio operativo, che risponde ai requisiti previsti dalla normativa prudenziale per la deduzione

³² Autorizzazione comunicata con lettera n.0083034/14 del 24/01/2014, a valere sulle segnalazioni riferite al 31.12.2013.

delle perdite attese: il censimento nel data set di calcolo, la copertura da accantonamenti specifici a conto economico e l'individuazione sulla base di criteri documentati.

4.5.1 Il perimetro di interesse

Le Fonti che rispondono ai requisiti sono la fonte Cause Legali e la fonte Risorse Umane.



Il processo di gestione delle cause è strutturato nelle seguenti fasi:

- a) censimento della causa nell'applicativo MICRA;
- b) determinazione degli accantonamenti;
- c) passaggio dei dati a Fondo Rischi ed Oneri;
- d) censimento nel data base dell'Operational Risk (OpRiskEv).

Nella fase di censimento le cause legali di Banca Montepaschi vengono gestite con l'ausilio del DB gestionale MICRA, che viene utilizzato per la gestione sia del contenzioso con la clientela e le controparti (Micra "Contenzioso" a partire dal 2009, che contiene anche le cause seguite dall'Area Compliance dal 2012), sia del contenzioso con il personale (Micra "Giuslavoristico", a partire dal 2010).

La determinazione degli accantonamenti segue metodologie diverse a seconda che si tratti di cause legali con clienti/controparti piuttosto che di cause con il personale.

Per quanto riguarda le cause legali con i clienti, l'Area Legale utilizza una metodologia specifica a partire dal 2010³³ su tutto il perimetro delle cause aperte e che è ampiamente descritta nelle Linee Guida³⁴, emanate all'interno dell'area con lo scopo di normare ed uniformare il processo di determinazione degli accantonamenti.

Per quanto riguarda le cause giuslavoristiche, seguite dalla Direzione Risorse Umane, non viene adottata una metodologia statistica. Gli accantonamenti vengono valutati sulla base delle

³³ La metodologia è stata definita nel 2010 ma è stata poi applicata retroattivamente su tutte le cause aperte.

³⁴ Il documento è stato emanato dal Servizio Assistenza Giudiziale nel dicembre 2010 e definisce più in generale il processo di gestione dei procedimenti giudiziari.

evidenze della causa e in riferimento all'esperienze maturate sul tema oggetto della causa. Infine, per quanto riguarda le controversie in materia di antiriciclaggio, seguite dall'Area Compliance, gli accantonamenti vengano impostati sulla base di una percentuale di perdita stimata, ad oggi è pari al 5%-10% dell'importo contestato. Tale quantificazione si basa sull'osservazione della serie storica e dell'andamento delle sanzioni passate e viene monitorato nel tempo per verificarne la coerenza.

Gli accantonamenti a Fondo Rischi e Oneri, effettuati in ossequio ai criteri previsti dai vigenti principi contabili, vengono operati con periodicità mensile. A livello contabile le registrazioni avvengono mediante scritture di sintesi, che sono il risultato delle variazioni analitiche ottenute mediante l'utilizzo dell'applicativo MICRA. La procedura MICRA, per la quale è prevista una chiusura periodica mensile, fornisce tutte le informazioni necessarie all'esecuzione delle scritture contabili.

Infine, i dati di perdita operativa sono raccolti dal Servizio Rischi Operativi e Reputazionali a partire dalle informazioni in possesso delle c.d. Fonti Informative, presso le quali è possibile reperire le informazioni necessarie per catalogare l'evento in modo completo ai fini dell'Operational Risk Management.

Pertanto il perimetro di interesse che rispetta i requisiti normativi è stato individuato nell'insieme delle posizioni che presentano congiuntamente le seguenti caratteristiche:

- sono censite in OpRiskEv e con i requisiti di ingresso nel dataset di calcolo per il requisito regolamentare;
- derivano da MICRA e sono in MICRA riscontrabili, pertanto sono gestite con un processo strutturato e documentato, che permette il seguimiento completo di ogni causa e la registrazione del relativo accantonamento;
- hanno accantonamento specifico a Fondo Rischi e Oneri, pertanto si possono individuare in modo univoco gli accantonamenti.

4.5.2 Il calcolo della detrazione

Una volta definito il perimetro di interesse per individuare le perdite operative che rispettino i requisiti normativi si procede al calcolo del valore da detrarre dal requisito.

Si deve considerare che la normativa regolamentare stabilisce che la detrazione delle perdite attese non può eccedere il valore della perdita attesa complessiva stimata nell'ambito del modello di calcolo.

Inoltre, al fine di considerare una definizione di perdita attesa che risponda a requisiti di stabilità e robustezza, la normativa suggerisce di evitare l'uso della media aritmetica dei dati campionari, non idonea nel caso di distribuzioni fortemente asimmetriche.

Il Gruppo Montepaschi ha pertanto ritenuto opportuno utilizzare la mediana come definizione di perdita attesa.

Pertanto, per ciascuna classe di rischio la detrazione per le perdite attese viene calcolata come il minimo tra le seguenti due quantità:

- l'ammontare complessivo di accantonamento presente nell'ultimo anno a Fondo Rischi e Oneri sul perimetro individuato, dove per accantonamento si intende lo stock di accantonamento complessivo presente a fondo
- la mediana della distribuzione delle perdite simulata dal modello interno per la classe in oggetto.

Periodicamente viene inoltre effettuato un'analisi di "backtesting" al fine di verificare la coerenza degli accantonamenti nel perimetro identificato con le perdite effettivamente subite dal Gruppo, si veda Sezione 4.7.5.

4.6 Allocazione del Capitale regolamentare

Dal 2008 per le società che ricadono nel perimetro AMA viene determinato un requisito patrimoniale individuale ripartendo quello determinato a livello di Gruppo sulla base dei seguenti criteri:

- Perdita media annua; ottenuta come media su tre anni delle perdite operative raccolte per data di rilevazione tramite LDC. In questo modo si tiene conto del dato di perdita storica;
- VaR qualitativo ottenuto tramite le analisi di scenario, che è determinato dagli esperti e rappresenta la componente qualitativa dell'analisi. In questo modo si tiene conto della componente forward looking;
- Requisito da metodo Standard che rispecchia gli aspetti dimensionali e reddituali della società.

I tre valori di capitale regolamentare ottenuti vengono successivamente sintetizzati in un unico indicatore attraverso la ponderazione:

- Media LDC sugli ultimi 3 anni³⁵: 20%
- VaR qualitativo: 20%
- Requisito da metodo Standard: 60%

³⁵ Si considerano gli eventi con data di rilevazione a partire dal primo gennaio del terzo anno di calendario precedente l'ultima trimestrale di chiusura anno. Ad esempio, al 30/09/2015 sono stati selezionati i dati con data di rilevazione a partire dal 1/1/2012, mentre al 31/12/2015 i dati a partire dal 1/1/2013.

Si ritiene che tale suddivisione garantisca una ripartizione equa del capitale senza svantaggiare troppo Società con eventi storici superati e senza favorire visioni ottimistiche nel management durante lo scenario.

Si riporta un esempio di allocazione del capitale secondo quanto descritto:

Si ipotizzi di considerare unicamente 3 banche: A, B, C e si ipotizzi un requisito regolamentare computato attraverso il metodo AMA di 800. La tabella mostra l'allocazione del requisito AMA che si avrebbe secondo ciascuno dei tre criteri:

Legal Entity	Requisito Standard	%	Requisito AMA	Perdita realizzata	%	Requisito AMA	Stima Perdita Esperti	%	Requisito AMA
A	500	50%	400	60000	34%	274	48000	49%	396
B	300	30%	240	100000	57%	457	35000	36%	289
C	200	20%	160	15000	9%	69	14000	14%	115
	1000		800	175000		800	97000		800

- Il criterio del requisito standard, riflette solo parzialmente il profilo di rischio della banca;
- la proporzionalità alle perdite realizzate storicamente penalizzerebbe le entità che sono incorse in perdite molto elevate anche nel caso in cui esse siano imputabili a fenomeni che non possono più accadere;
- la proporzionalità alle stime di perdita degli esperti creerebbe un conflitto di interessi in capo agli stessi, i quali tenderebbero a formulare ipotesi troppo ottimistiche al solo scopo di ottenere un'allocazione inferiore.

La soluzione adottata porta alla seguente allocazione:

$$A \quad (0,6*400 + 0,2*274 + 0,2*396) = 374$$

$$B \quad (0,6*240 + 0,2*457 + 0,2*289) = 293$$

$$C \quad (0,6*160 + 0,2*69 + 0,2*115) = 133.$$

La deduzione delle perdite attese si alloca sulle società AMA che hanno una gestione attiva e strutturata degli accantonamenti e quindi contribuiscono al calcolo della deduzione stessa, vedi Sezione 4.5.

4.7 Prove di Stress, analisi di Sensitivity e backtesting del modello AMA.

Scopo delle attività di stress testing è quello di stimare il cambiamento del valore di un portafoglio nel momento in cui si verificano ampie variazioni in un set di variabili finanziarie e/o macroeconomiche. Tali variabili condizionano negativamente i fattori di rischio alla base della misurazione.

Lo stress testing nei rischi operativi si differenzia da quello effettuato nei rischi di credito e di mercato per la mancanza di una correlazione diretta e dimostrata fra variabili finanziarie e/o macroeconomiche (tassi o indici ad esempio) ed eventi di rischio operativo.

Naturalmente la situazione dei mercati può influire (ed influisce) su alcune specifiche tipologie di eventi, ad esempio i reclami sui prodotti finanziari. Per questa classe si registra spesso un incremento di litigiosità (reclami e cause legali) in situazioni di mercato particolarmente negative. Le variabili di mercato possono portare ad esempio al fallimento di una società che a sua volta farà aumentare l'insorgenza di reclami da parte della clientela. Questa situazione si è presentata storicamente dopo il default dell'Argentina, o i crack di Cirio e Parmalat. Il caso Italease è stato generato dalla crisi dei derivati, che ha portato ad un eccezionale numero di ricorsi e reclami, mentre nel caso di Società Generale, la frode è stata scoperta a causa dell'andamento avverso dei mercati finanziari su cui investiva il trader.

Tuttavia la relazione fra variabili finanziarie ed il verificarsi degli eventi non è dimostrabile. Si pensi ad esempio alla classe frode esterna dove confluiscono le rapine; in questo caso il fenomeno è correlato ad aspetti sociali (ad esempio l'indulto) ma non sconta nessun tipo di relazione con i parametri di rischio di credito e di mercato.

Di seguito vengono considerati i possibili approcci alla materia.

Si precisa che in corrispondenza della predisposizione del resoconto annuale ICAAP il Servizio Rischi Operativi effettua analisi di stress del modello per verificarne la robustezza e la sensibilità ad eventuali shock esterni.

4.7.1 Stress di modello

La base dati interna dei rischi operativi può essere stressata con l'inserimento di alcuni punti di perdita che vanno ad impattare su due aspetti:

- stress sulla distribuzione di severity, mediante perdite ritenute estreme;
- stress sulla distribuzione di frequency, mediante perdite di valore inferiore ma di numero consistente.

Lo stress può avvenire sia sulla base dati quantitativa (derivante da Loss Data Collection, sia su quella qualitativa, derivante dalle Analisi di Scenario).

Questo tipo di approccio presenta il vantaggio di essere facilmente “formalizzabile” dal punto di vista metodologico, nel senso che i parametri di stress possono essere definiti sulla base di alcune grandezze tipiche della base dati non stressata.

4.7.2 Scenario Stress

Il modello di misurazione dei rischi operativi può essere stressato simulando situazioni specifiche registrate nel Sistema oppure le previsioni degli esperti sugli sviluppi futuri di criticità emerse nella fase di assessment dei rischi operativi.

Il primo approccio utilizza la logica secondo cui alcune situazioni, sicuramente di carattere estremo, registratesi nel contesto operativo (Sistema), possano ripetersi all'interno del Gruppo Montepaschi, o in condizioni tali da generare un impatto diretto sul modello Gruppo Montepaschi.

Scopo di questo approccio è quello di contestualizzare gli scenari di stress su condizioni reali, con il duplice scopo di testare il modello, ma allo stesso tempo fornire al management indicazioni di carattere gestionale. Vengono selezionati a questo fine, eventi di rischio operativo di cui sono noti sufficienti dettagli informativi (sostanzialmente da informazioni di stampa, di pubblico dominio), e viene simulato un set di dati che possano in qualche modo avere un'attinenza diretta con quanto verificatosi in altro istituto finanziario. Il set di dati viene successivamente inserito nella base dati storica, e viene effettuato il calcolo del requisito patrimoniale sulla base dati stressata.

La logica del secondo approccio è quella di valutare le problematiche percepite come critiche per il Gruppo Montepaschi dagli “addetti ai lavori”.

4.7.3 Reverse stress

Il Reverse stress consiste nell'analisi delle code della distribuzione di perdita per individuare scenari critici per il Gruppo Montepaschi.

Partendo dalle distribuzioni stimate per il requisito vengono determinati i valori di capitale economico “target” al 99,96° percentile (coerente con una situazione di miglioramento del rating del Gruppo). Analizzando tali valori di coda, si possono individuare quali scenari possono comportare una situazione critica per la banca. Per determinare gli scenari di reverse stress vengono analizzati sia l'andamento storico delle perdite interne e di Sistema (analisi storica) sia le risposte ai questionari di scenario (analisi prospettica).

4.7.4 Stress congiunto con gli altri rischi

Lo stress congiunto con gli altri rischi (mercato, credito, ecc) è stato costruito utilizzando come driver comune l'andamento del PIL italiano.

La tipologia di stress in questo caso si basa sulla logica di utilizzare i periodi storici identificati dall'andamento del PIL (espansione, recessione lieve e recessione forte) per generare una serie storica di pseudo-dati da utilizzare all'interno del modello AMA, interpretandola come pseudo serie dell'anno futuro.

4.7.5 *Backtesting del modello AMA*

Il presente documento descrive il metodo utilizzato per effettuare il *backtesting* del modello AMA per i rischi operativi, ovvero il confronto tra le previsioni fornite dal modello e le perdite effettivamente osservate.

L'esercizio si è basato su un *backtesting in-sample*: sono stati confrontati i valori predetti dal modello con le perdite interne effettivamente osservate, seguendo la procedura di seguito descritta:

1. Si calcolano le perdite trimestrali osservate, ossia si aggregano le perdite per trimestre. Vengono estratti gli eventi dalle tre tabelle EV, SC e DE (eventi, sospesi contabili e dubbi esiti): non si estraggono i macroeventi, ma se ne prendono gli effetti dalla tabella EV.
2. Si selezionano eventi con data compresa tra il primo gennaio 2009 e l'ultima data disponibile, considerando la data di accadimento oppure di rilevazione. Non si applicano i filtri usati per il calcolo del requisito.
3. I valori predetti dal modello vengono quindi confrontati con le perdite osservate aggregate per trimestre. Si calcola il VaR del modello AMA su un orizzonte temporale trimestrale anziché annuale. Si calcola il VaR integrato, per diversi valori del livello di confidenza: $p=50\%$, 75% , 90% , 95% , 99% , sotto due ipotesi:
 - a. dipendenza perfetta tra le classi;
 - b. dipendenza non-perfetta, includendo pertanto un beneficio di diversificazione.

In ciascun caso, non abbiamo applicato la deduzione per perdite attese.

A titolo di esempio, riportiamo la Figura 28, che mette a confronto le perdite trimestrali osservate con i percentili della distribuzione simulata non diversificata prodotta dal modello AMA al 30/09/2014 con orizzonte temporale trimestrale (ricordiamo che nel modello abbiamo usato i parametri ottenuti a seguito di integrazione quali-quantitativa).

Tre sole perdite aggregate trimestrali sono superiori al 90% percentile di modello:

1. primo trimestre del 2013: per effetto degli eventi legati al restatement di bilancio (Alexandria e Santorini);
2. terzo trimestre del 2011: per effetto dei noti eventi di abuso di diritto;
3. primo trimestre 2010: per effetto della frode interna da 24 EUR/mln dell'Area Territoriale Sud.

Tuttavia soltanto la perdita legata al restatement di bilancio risulta superare il 99%.

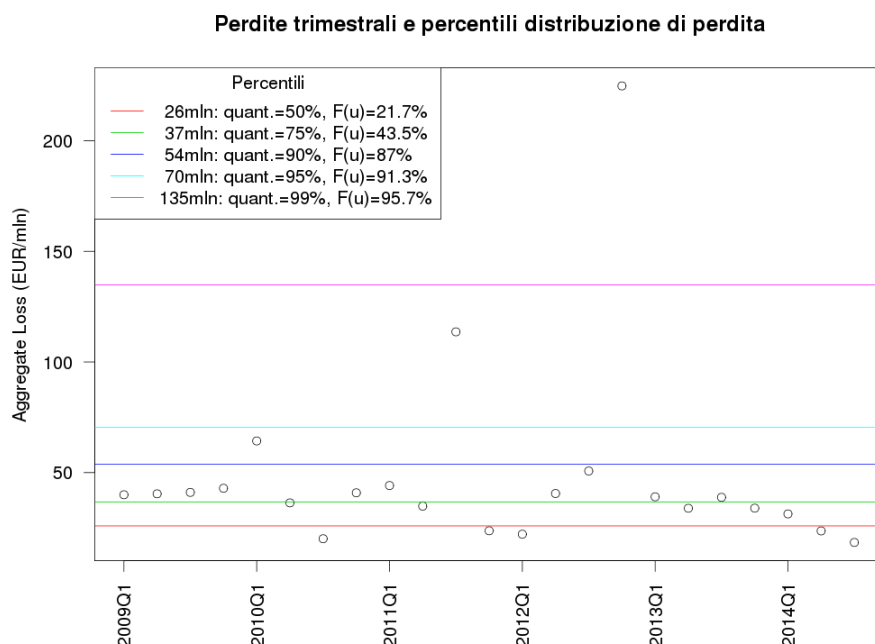


Figura 28 Perdite trimestrali e percentili distribuzione di perdita

4.8 Benchmarking delle perdite interne rispetto alle perdite di sistema

Ai fini di analizzare le perdite interne sostenute dal Gruppo Montepaschi in ottica di *benchmarking*, viene effettuata semestralmente una analisi di confronto tra i dati di perdita interni e quelli del sistema bancario italiano.

A tale fine vengono utilizzati i dati del consorzio ABI DIPO (Database Italiano delle Perdite Operative), cui partecipano circa 33 membri tra banche e gruppi bancari (circa 230 entità segnalanti in totale). I dati sono gli stessi utilizzati ai fini del calcolo del requisito.

Ogni aderente, compreso Montepaschi, invia al DIPO semestralmente i dati di perdita, ricevendone in ritorno i flussi di dati di tutto il consorzio in forma anonima. Gli aderenti sono divisi in gruppi, definiti *Peer groups*, sulla base dei costi operativi e dell'indicatore rilevante. Il Gruppo Montepaschi appartiene al Peer 1 insieme ad altre 8 principali banche italiane.

Il confronto viene effettuato comparando l'andamento delle perdite operative di Montepaschi, rapportate al Total Asset del Gruppo, all'andamento delle perdite operative delle altre banche italiane appartenenti allo stesso Peer, anch'esse rapportate al Total Asset complessivo.

Il confronto viene fatto per anno di rilevazione, sia a livello complessivo che per event type.

I dati interni non contengono le perdite relative alle revocatorie fallimentari, poichè incluse nel calcolo del requisito del rischio di credito, pertanto i dati relativi ai peers vengono depurati di questa componente in modo da renderli confrontabili.

5 Approfondimenti teorici

5.1 Analisi Dati Storici

5.1.1 Pre-analisi

5.1.1.1 Autocorrelazione Seriale

L'autocorrelazione seriale è tecnicamente definita come una dipendenza di ordine k tra ciascun i -esimo elemento della serie e l'elemento $(i-k)$ -esimo dove il termine k è detto ritardo (lag).

Per poter affrontare il concetto di analisi della correlazione si deve fare riferimento ai momenti teorici dei processi stocastici, in particolare all'autocovarianza che si identifica come la covarianza tra valori della serie Z in istanti temporali diversi. Normalmente la covarianza misura la tendenza di due grandezze a variare nello stesso senso, in questo caso si utilizza un'unica variabile misurata in due istanti temporali diversi. In formule si ottiene:

$$\gamma(t, t+k) = E[(Z_t - \mu(t))(Z_{t+k} - \mu(t))] = \text{Cov}(Z_t, Z_{t+k})$$

Equazione 5

Da notare che l'autocovarianza è funzione di due istanti temporali, t e $(t+k)$. In questo senso la varianza risulta essere un caso particolare dell'autocovarianza, ponendo $k=0$. L'autocorrelazione ha il vantaggio, rispetto all'autocovarianza, di non essere compresa fra limiti fissi, ma tra i valori estremi di -1 e $+1$. L'autocorrelazione si ottiene semplicemente dividendo l'autocovarianza per il prodotto degli scarti quadratici medi di Z_t e Z_{t-k} .

La funzione di autocorrelazione (ACF) è il coefficiente di correlazione lineare $\rho(k)$ tra le variabili casuali Z_t e Z_{t-k} calcolato al variare di k .

$$\rho(k) = \text{Corr}(Z_t Z_{t-k}) = \frac{\text{Cov}(Z_t Z_{t-k})}{\sqrt{\text{Var}(Z_t) \text{Var}(Z_{t-k})}}$$

Equazione 6 Autocorrelazione

5.1.1.2 Autocorrelation Plot

Le serie storiche possono essere esaminate tramite l'Autocorrelation Plot (ACP), il quale illustra graficamente e numericamente la funzione di autocorrelazione (ACF), elencando i coefficienti di correlazione in serie per ritardi consecutivi (asse delle ordinate) in un intervallo specificato di ritardi (Lags, asse delle ascisse).

Nell'esame dell'ACP si deve tenere presente che le autocorrelazioni per ritardi consecutivi sono formalmente dipendenti. Se il primo elemento è strettamente legato al secondo e questo ad un terzo, allora in qualche modo anche il primo ed il terzo sono collegati. Questo induce che lo schema di dipendenza seriale può cambiare sensibilmente eliminando l'autocorrelazione di primo livello (cioè differenziando la serie con un ritardo di 1).

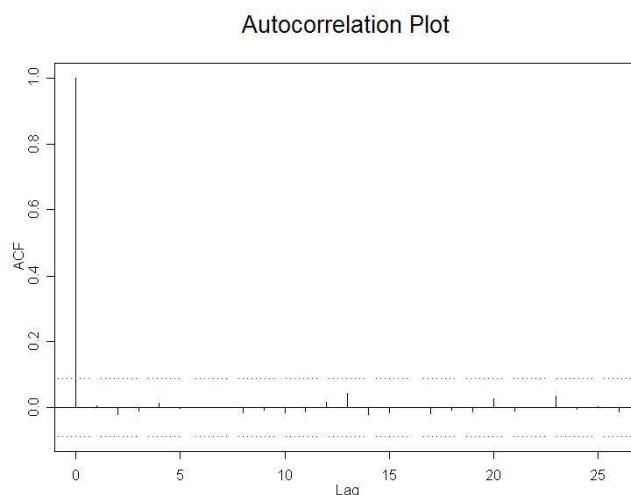


Figura 29 Autocorrelation plot

Le due linee tratteggiate orizzontali individuano la fascia entro cui, ad un determinato livello di confidenza (95% o 99%), non si ha significatività statistica per la correlazione. In caso di assenza di autocorrelazione la distribuzione asintotica della stima del coefficiente di autocorrelazione è di tipo normale e si può costruire una banda di confidenza di estremi:

$$\pm \frac{z_{1-\alpha/2}}{\sqrt{N}}$$

dove N rappresenta l'ampiezza del campione, z è il quantile della distribuzione normale standard ed α il livello di significatività prescelto.

5.1.1.3 Plot of Records Development

Il Plot of Record Development (PRD) è un altro strumento utilizzato per testare l'indipendenza di una serie storica X_1, \dots, X_n . Il PRD si basa sull'analisi della frequenza dei record all'interno della serie storica, dove la generica X_i è un record se:

$$X_i > \max_{j=1, \dots, i-1} (X_1, \dots, X_j, \dots, X_{i-1}) \text{ con } i = 1, \dots, n$$

Equazione 7

Sotto l'ipotesi di indipendenza dei dati campionari si dimostra (P. Embrechts, C. Kluppelberg, T. Mikosch, 1997)³⁶ che il numero di records in un intervallo $[1, \dots, n]$ è approssimativamente pari al $\ln(n)$. Pertanto man mano che ci sposta lungo la serie diventa sempre meno probabile il verificarsi di un nuovo record.

Il PRD è costruito riportando sull'asse delle ordinate il numero di records e su quello delle ascisse il logaritmo del numero delle osservazioni campionarie. Nel plot viene riportata sia la

³⁶pp. 247-260.

progressione dei record sia la funzione logaritmica e l'ipotesi di indipendenza delle osservazioni è tanto maggiore tanto più la progressione in esame si addensa lungo la spezzata logaritmica.

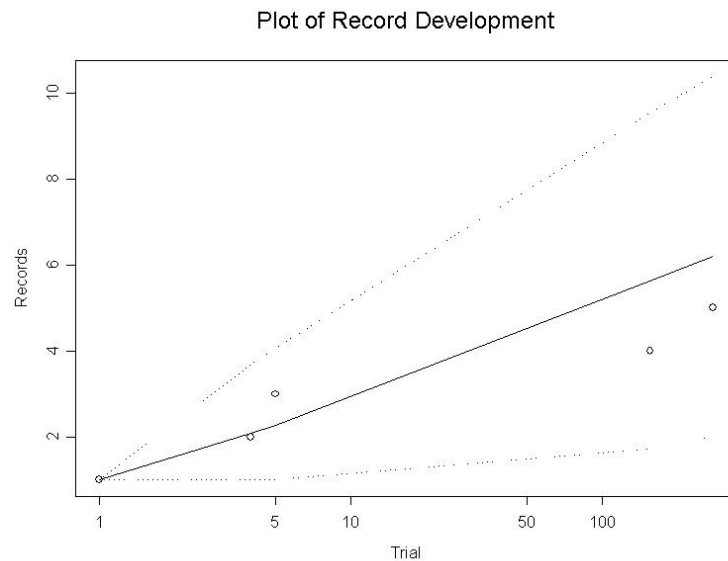


Figura 30 Plot dello sviluppo dei record

5.1.2 Strumenti d'analisi grafica

5.1.2.1 Box Plot

Il box plot serve per descrivere in modo compatto e grafico la distribuzione di una variabile basandosi sui quartili e sul campo di variazione.

È il disegno su un piano cartesiano di un rettangolo (blu nella figura che segue), i cui estremi sono il primo e terzo quartile (Q_1 e Q_3), tagliato da una linea all'altezza della mediana (Q_2). Il minimo della distribuzione viene indicato con (Q_0), mentre il massimo con (Q_4). Poiché tra Q_1 e Q_3 si trova il quantile che bipartisce le frequenze (o mediana) della distribuzione, se gli scarti tra i quantili estremi e il quantile centrale sono piccoli, vuol dire che la variabilità è contenuta; altrimenti la variabilità è elevata.

Il grafico si compone di ulteriori due righe corrispondenti ai valori distanti 1,5 volte la distanza rispettivamente tra il primo quartile e la mediana e il terzo quartile e la mediana. Con l'obiettivo di rappresentare la pesantezza della coda, inoltre, sono rappresentati nel grafico tutti i valori che fuoriescono dall'intervallo delimitato dalle due righe come righe isolate.

Il grafico esemplificativo che segue evidenzia la forte asimmetria della distribuzione della severity e una coda destra pesante.

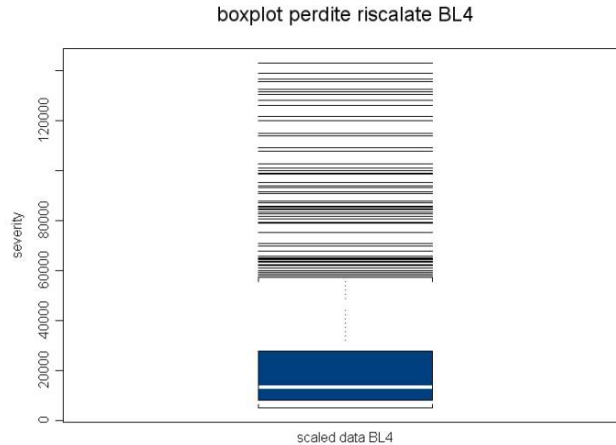


Figura 31 Boxplot delle perdite

5.1.2.2 Mean Excess Plot

Definiamo come *mean excess function* del campione di dati X la seguente quantità:

$$e(u) = E(X - u | X > u) = \frac{1}{1 - F_X(u)} \int_u^{\infty} (x - u) \cdot f_X(x) dx$$

Equazione 8

dove f_X è la funzione densità di probabilità associata a X . Dato un campione empirico di dati, il Mean Excess Graph è dunque il grafico della media dei punti oltre la soglia:

$$e(u) = \frac{1}{N(u)} \cdot \sum_{i=1}^{N(u)} (x_{i:N} - u)$$

Equazione 9 Mean Excess

dove N è il numero di punti oltre la soglia u , mentre x_i sono i punti del campione.

5.1.2.3 Hill Plot - stimatore di Hill

Lo stimatore di Hill è dato dalla seguente espressione:

$$\hat{\xi} = \frac{1}{k-1} \sum_{i=1}^{k-1} \ln X_{i,N} - \ln X_{k,N} \quad \text{con } k > 1$$

Equazione 10 Stimatore di Hill

dove k è il numero di eccedenze, N la dimensione del campione e $X_{1,N}, \dots, X_{k,N}$ i valori delle osservazioni oltre la soglia ordinate in modo decrescente.

5.1.2.4 Rapporto Massimo – Somma

Il metodo del Rapporto Massimo – Somma permette di analizzare in maniera qualitativa il valore dei momenti della distribuzione empirica, fornendo quindi la possibilità di capire se tale distribuzione è a coda pesante.

Date n estrazioni X_i IID, per un numero strettamente positivo p definiamo le seguenti quantità:

$$S_n(p) = |X_1|^p + \dots + |X_n|^p$$

$$M_n(p) = \max(|X_1|^p, \dots, |X_n|^p)$$

Equazione 11

Si può dimostrare che la quantità, funzione del numero n di estrazioni:

$$R_n(p) = \frac{M_n(p)}{S_n(p)} \text{ con } n \geq 1, \quad p > 0$$

Equazione 12 Massimo su somma.

è legata al valore del momento di ordine p , nel senso che se, dato p , al crescere di n il rapporto tende a zero, si può concludere che la distribuzione ha il momento di ordine p finito. Inoltre è vera anche l'implicazione inversa, se il momento di ordine p della distribuzione risulta finito, allora, al crescere di n , $R_n(p)$ tende a zero.

In termini matematici vale il seguente teorema (O'Brien, 1980):

$$\lim_{n \rightarrow \infty} P(R_n(p) \rightarrow 0) = 1 \Leftrightarrow E\left[|X|^p\right] < \infty$$

che afferma che la successione $R_n(p)$ converge con probabilità 1 a zero se e solo se il momento di ordine p della distribuzione risulta finito.

Supponiamo che X_1, \dots, X_n siano distribuite con una funzione $f(x)$ con supporto $(0, \infty)$. Dimostriamo che, se il momento di ordine p diverge, la coda della distribuzione di X ha un peso maggiore di quello della funzione $x^{-(p+1)}$.

Il momento di ordine p è: $M_p = E[X^p] = \int_0^\infty x^p f(x) dx$.

Se tale momento diverge significa che l'integrale di $x^p f(x)$ diverge. Data una funzione $g(x)$ limitata, positiva e decrescente da un certo punto in poi, si ha che

$$\int_0^\infty g(x) dx = \infty \Leftrightarrow g(x) \geq x^{-1}, \forall x > \bar{x}$$

Supponendo che anche nel nostro caso la densità di probabilità sia decrescente da un certo punto in poi³⁷, identificando $g(x) = x^p f(x)$, la condizione affinché il momento p -esimo diverga è che la coda della distribuzione soddisfi:

$$f(x) \geq x^{-(p+1)}.$$

Al contrario, se il momento di ordine p risulta finito, allora la coda della distribuzione, da un certo punto in poi, deve essere:

³⁷ In tutti i casi di nostro interesse, le densità di probabilità sono decrescenti da un certo punto in poi.

$$f(x) \leq x^{-(p+1)}$$

Avendo un'informazione sulla forma della coda della distribuzione, si può dunque risalire ad un valore del parametro della distribuzione che meglio modella tale andamento (ad esempio, il parametro di *shape* nella GPD).

ESEMPIO: Generalized Pareto Distribution

Siano $X_1, \dots, X_N > 0$ gli eccessi oltre la soglia delle perdite, supposti distribuiti secondo una Pareto Generalizzata di shape ξ e con funzione di densità $f_{\xi, \sigma}(x)$ data da:

$$f_{\xi, \sigma}(x) = \begin{cases} \frac{1}{\sigma} \left(1 + \frac{\xi}{\sigma} x\right)^{-\frac{1}{\xi}-1}, & \xi > 0, x \geq 0 \\ \frac{1}{\sigma} \left(1 + \frac{\xi}{\sigma} x\right)^{-\frac{1}{\xi}-1}, & \xi < 0, 0 \leq x \leq -\frac{1}{\xi} \end{cases}$$

Se per un certo k in poi $\frac{\xi}{\sigma} X_k \gg 1$, allora possiamo approssimare la $f_{\xi, \sigma}(x)$ nel seguente modo

$$f(x) \approx \sigma^{1/\xi} (\xi \cdot x)^{-(1/\xi+1)}$$

Per ogni p tale che la successione $\{R_n(p)\}_{n=k}^N$ non converge a zero, allora il momento di ordine p è infinito e risulta $x^{-(1/\xi+1)} \geq x^{-(p+1)}$ che equivale a $\xi \geq 1/p$.

D'altra parte, per ogni p' tale che $\{R_n(p')\}_{n=k}^N$ converge a zero, il momento di ordine p' è finito e

$$x^{-(1/\xi+1)} \leq x^{-(p'+1)}$$

che equivale a $\xi \leq 1/p'$.

Questo esempio, mostra come il metodo del Rapporto Massimo – Somma fornisca informazioni sul parametro di shape della GPD.

5.1.3 Metodi di Stima

5.1.3.1 Minimi Quadrati (Least Squares -LS)

Il metodo LS viene utilizzato per eseguire i fit di funzioni discrete o per i fit di istogrammi. Esso consiste nel minimizzare la differenza quadratica fra le ordinate della pdf empirica e di quella ipotizzata. Nel caso in cui gli scarti dalla distribuzione empirica siano di tipo gaussiano, la somma degli scarti quadratici segue una distribuzione Chi-quadro.

5.1.3.2 Maximum Likelihood Estimation (MLE)

Il metodo MLE si basa sul calcolo della funzione di verosimiglianza. La trattazione che segue ripercorre tutti gli step matematici fondamentali della teoria:

- funzione di verosimiglianza;
- stimatori di massima verosimiglianza.

La funzione di verosimiglianza. Siano X_1, \dots, X_n osservazioni campionarie estratte da una data distribuzione parametrica caratterizzata da una determinata forma funzionale (possono essere i dati di severity o quelli di frequency). Si indichi con θ il vettore di parametri della forma funzionale.

Sia $f(X_T; \theta)$ la funzione di probabilità congiunta del campione, la quale associa ad ogni campione e ad ogni valore parametrico un numero positivo. Al variare di X_T , dato un valore dei parametri θ , si ha la funzione di densità congiunta del campione.

Se $f(X_T; \theta)$ viene analizzata come funzione del vettore dei parametri, data una certa realizzazione campionaria, essa definisce la cosiddetta funzione di verosimiglianza, indicata come:

$$l(\theta; X_T) = f(X_T; \theta)$$

E' inoltre utile definire anche la cosiddetta funzione di log-verosimiglianza, che è data dal logaritmo della $l(\theta; X_T)$, trasformazione monotona ammissibile poiché la funzione di densità è sempre positiva:

$$L(\theta; X_T) = \ln l(\theta; X_T)$$

Il vettore delle derivate prime parziali (gradiente) della funzione di log-verosimiglianza viene definito *score*:

$$q(\theta) = \frac{\partial L(\theta)}{\partial \theta}$$

mentre la matrice delle derivate seconde (Hessiano) della log-verosimiglianza viene indicata come:

$$Q(\theta) = \frac{\partial q(\theta)}{\partial \theta} = \frac{\partial^2 L(\theta)}{\partial \theta \cdot \partial \theta'}$$

Nel caso di variabili indipendenti ed identicamente distribuite la log-verosimiglianza può essere scritta come:

$$L(\theta) = \sum_{t=1}^T \ln f(x_t; \theta) = \sum_{t=1}^T l_t(\theta),$$

Equazione 13

dove $l_t(\theta) = \ln f(x_t; \theta)$ rappresenta la log-verosimiglianza per una singola osservazione.

Gli stimatori di massima verosimiglianza. Dato un modello statistico parametrico ed un campione di numerosità T , su un vettore aleatorio X di dimensione n , il metodo della massima verosimiglianza è basato sulla ricerca, tra tutti i possibili valori di θ , di quello più coerente con il campione estratto.

Lo stimatore di massima verosimiglianza è quindi definito come il vettore dei parametri che rende massimo il valore della funzione di verosimiglianza calcolata sul campione dato, cioè:

$$\hat{\theta}_T : \ell(\hat{\theta}_T) \geq \ell(\theta) \quad \forall \theta \in \Theta$$

Considerando la funzione di log-verosimiglianza, lo stimatore MLE viene quindi ottenuto dalla soluzione del seguente problema di massimo:

$$\max_{\theta \in \Theta} L(\theta).$$

Supponendo che la log-verosimiglianza sia derivabile, il problema può essere risolto considerando i vincoli di ricerca del massimo di una funzione continua:

$$\frac{\partial L(\theta)}{\partial \theta} = 0 \quad \text{e} \quad \frac{\partial^2 L(\theta)}{\partial \theta^2} < 0.$$

Occorre osservare che:

- il metodo della MLE fornisce stimatori che godono di proprietà ottimali in senso asintotico;
- dalle condizioni del primo ordine non è spesso possibile ricavare gli stimatori in forma chiusa e si rende necessario ricorrere ai metodi di ottimizzazione numerica.

Il metodo di massima verosimiglianza fornisce il miglior stimatore a livello asintotico, ed è distribuito in maniera normale attorno al valore vero ricercato.

1. Intervalli di confidenza per lo stimatore di massima verosimiglianza.

Utilizzando il teorema asintotico (Cfr Par 5.5.3) è possibile definire intervalli di confidenza per le stime dei parametri della GPD.

Fissato il livello di confidenza α si definisce l'intervallo di confidenza al livello $(1-\alpha)\%$ per ξ :

$$\xi \pm \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \sqrt{\frac{\hat{v}_{1,1}}{N_t}}$$

Equazione 14

Analogamente per β si ha:

$$\beta \pm \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \sqrt{\frac{\hat{v}_{2,2}}{N_t}}$$

Dove $\hat{v}_{1,1}, \hat{v}_{2,2}$ sono, rispettivamente, il primo o il quarto elemento diagonale della matrice delle varianze e covarianze asintotica per l'MLE (definita in 7.3.3) empirica, ovvero ottenuta sostituendo ai parametri i loro stimatori, e N_t (dove t indica la soglia) è definito come:

$$N_t = \text{card}\{i : i = 1, \dots, n, X_i > t\}$$

Una misura più precisa che si può usare in alternativa è la profile likelihood definita genericamente come:

$$L_p(\xi) = \max_{\beta \in \xi} (L(\xi, \beta))$$

Gli intervalli di confidenza sono perciò definiti come:

$$CI_{\xi} = \left\{ \xi : \log L_p(\xi) \geq \log L_p(\hat{\xi}) - \frac{\chi_1^2(1-\alpha)}{2} \right\}$$

Equazione 15

Nel caso in cui i vincoli per l'esistenza della matrice di varianze e covarianze ($\xi > -1/2$) non siano soddisfatti è stato deciso di utilizzare un metodo numerico che consiste nel calcolo di intervalli basati sul campionamento (con ripetizione), e successivo fitting, dei dati sulla coda della distribuzione, in altre parole si approssima la distribuzione asintotica utilizzando un numero elevato ma finito di simulazioni (almeno 1000).

5.1.3.3 Maximum Penalized Likelihood Estimation (MPLE)

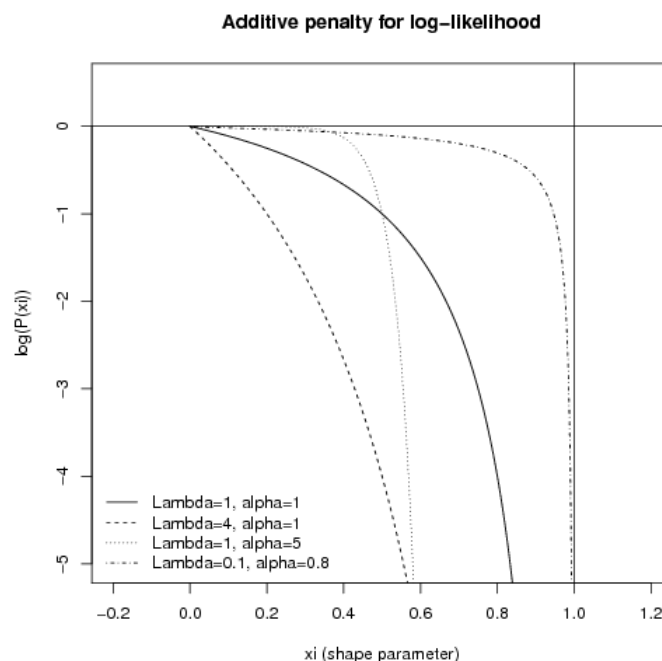
Questo metodo è stato introdotto da Coles e Dixon (articolo [18] della bibliografia) nella modellizzazione di eventi estremi (in particolare per l'approccio a massimi su blocchi). La verosimiglianza penalizzata è un metodo usato in generale per incorporare in un processo inferenziale informazioni che siano supplementari rispetto a quelle contenute nei dati. Un'applicazione standard è lo "smoothing" non parametrico, in cui la funzione di verosimiglianza è bilanciata da un termine che penalizza la discontinuità della funzione approssimante. Coles e Dixon utilizzano una funzione penalizzante al fine di complementare la funzione di verosimiglianza con l'informazione che il valore del parametro di coda ξ sia minore di uno e che valori vicini ad uno siano meno probabili di valori più piccoli. Questo approccio è analogo ad una formulazione Bayesiana nella quale informazione strutturale riguardo a ξ sia introdotta mediante una opportuna distribuzione a priori. Coles e Dixon propongono una funzione penalizzante della forma

$$P(\xi) = \begin{cases} 1, & \text{se } \xi \leq 0, \\ \exp \left[-\lambda \left(\frac{1}{1-\xi} - 1 \right)^{\alpha} \right], & \text{se } 0 < \xi < 1, \\ 0, & \text{se } \xi \geq 1, \end{cases}$$

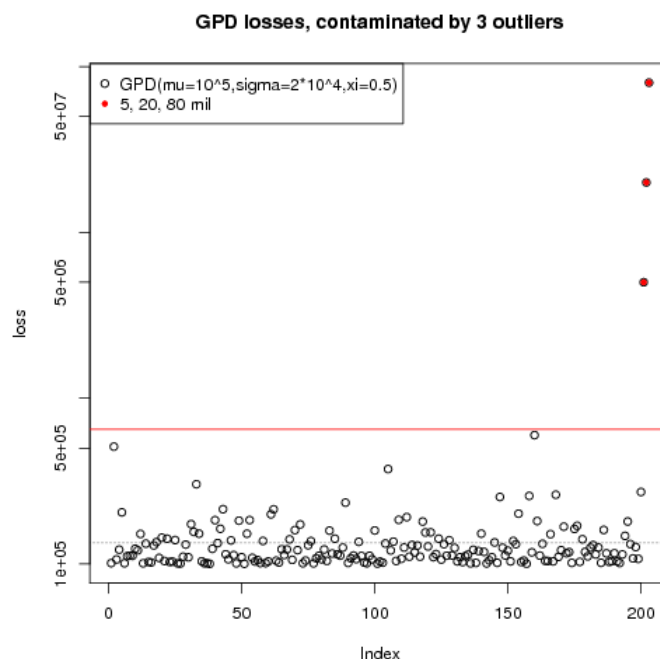
dove α e λ sono dei parametri non-negativi. Dato un campione di osservazioni indipendenti z_1, z_2, \dots, z_k estratte da una distribuzione GPD, la corrispondente funzione di verosimiglianza penalizzata è

$$L_{\text{pen}}(z_i; \xi, \mu, \beta) = P(\xi) L(z_i; \xi, \mu, \beta),$$

dove $L(z_i; \xi, \mu, \beta)$ denota la funzione di verosimiglianza della GPD valutata nelle osservazioni z_i . Il valore di (ξ, μ, β) che massimizza la funzione L_{pen} è chiamato lo *stimatore di massima verosimiglianza penalizzata* ("Maximum Penalized Likelihood Estimator", MPLE). Valori grandi per il parametro α nella funzione $P(\xi)$ corrispondono ad una maggiore penalizzazione (relativa) per valori di ξ vicini ad 1, mentre il parametro λ è un peso complessivo della penalizzazione. La figura sottostante mostra la funzione di penalizzazione per alcune combinazioni di valori dei parametri α e λ . Gli autori, a seguito di sperimentazione, consigliano i valori $\alpha = \lambda = 1$.



Nel contesto del modello AMA del Gruppo Montepaschi, questo stimatore è stato testato nella fase diagnostica del calcolo del requisito patrimoniale a partire dalla trimestrale del 31/12/2011, a fianco agli altri stimatori robusti (MDPD ed MGF-AD2). La procedura numerica utilizzata dal Gruppo Montepaschi è quella implementata nel pacchetto **R** POT di Mathieu Ribatet (<http://cran.r-project.org/web/packages/POT/index.html>). Come valori di partenza per la massimizzazione della verosimiglianza penalizzata, vengono utilizzati quelli ottenuti tramite stimatore PWM. Mostriamo ora un test di sensitività per lo stimatore, utilizzando una simulazione MonteCarlo. Consideriamo un insieme D_0 di 200 eventi di perdita, estratti da una distribuzione GPD con parametri $(\xi, \mu, \beta) = (0.5, 10^5, 2 * 10^4)$ ed un insieme “contaminato” D_c ottenuto aggiungendo a D_0 tre eventi di importo pari a 5, 20, 80 milioni. La figura sottostante illustra che queste tre perdite possono essere considerate degli “outlier” rispetto all’insieme D_0 . In effetti, la tabella sottostante mostra che il primo ed il terzo quartile sono scarsamente influenzati dall’introduzione delle tre nuove perdite, mentre la media aumenta di un fattore moltiplicativo 5 (le due linee orizzontali nella figura denotano la media di D_0 , con linea nera tratteggiata, e la media di D_c con linea rossa continua).



	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
D_0 : GPD losses	100.2	104.7	115.2	133.8	138.0	598.4
D_c : GPD + outliers	100.2	104.8	115.9	649.1	139.2	80000.0

Nella tabella sottostante vengono paragonate le stime GPD fornite dai cinque stimatori considerati dal Gruppo Montepaschi per le due basi dati D_0 e D_c . Osserviamo che, come prevedibile, tutte le stime sono vicine ai valori originali di $(\xi, \beta) = (0.5, 20000)$ per la base dati D_0 . Al contrario, per la base dati contaminata D_c il metodo MLE ed in particolar modo il metodo PWM forniscono stime notevolmente più alte rispetto a quelle fornite dai tre stimatori robusti MDPD, MGF ed MPLE: queste ultime sono tutte vicine a 0.7 e pertanto più vicine al valore corretto pari a 0.5, mentre lo stimatore PWM fornisce una stima distorta del parametro. Questo esempio illustra la maggiore affidabilità dello stimatore MPLE rispetto al PWM nel caso in cui un insieme di dati contenga perdite di entità notevolmente superiore rispetto al resto dei dati. Per questo motivo, lo stimatore MPLE è stato adottato a partire dalla trimestrale del 30 giugno 2013.

	β		ξ	
	D_0	D_c	D_0	D_c
MLE	19047	15702	0.46	0.88
MPLE	19577	16684	0.43	0.72
MGF-AD2	17839	17300	0.57	0.70
MDPD	18539	17246	0.50	0.68
PWM	19443	16683	0.43	0.97

I. Intervalli di confidenza per lo stimatore di massima verosimiglianza penalizzata (MPLE).

Utilizzando il teorema asintotico (Cfr Par 5.5.3) è possibile definire intervalli di confidenza per le stime dei parametri della GPD.

Fissato il livello di confidenza α si definisce l'intervallo di confidenza al livello $(1-\alpha)\%$ per ξ :

$$\xi \pm \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \sqrt{\frac{\hat{v}_{1,1}}{N_t}}$$

Equazione 16

Analogamente per β si ha:

$$\beta \pm \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \sqrt{\frac{\hat{v}_{2,2}}{N_t}}$$

Dove $\hat{v}_{1,1}, \hat{v}_{2,2}$ sono, rispettivamente, il primo o il quarto elemento diagonale della matrice asintotica delle varianze e covarianze per lo stimatore MPLE (definita in 6.3.3), ovvero ottenuta sostituendo ai parametri i loro stimatori, e N_t (dove t indica la soglia) è definito come:

$$N_t = \text{card}\{i : i = 1, \dots, n, X_i > t\}$$

Questo metodo viene affiancato da un metodo numerico basato su simulazione, che consiste nel calcolo di intervalli basati sul campionamento (con ripetizione), e successivo fitting, dei dati sulla coda della distribuzione, in altre parole si approssima la distribuzione asintotica utilizzando un numero elevato ma finito di simulazioni (almeno 1000).

5.1.3.4 Probability Weighted Moments (PWM)

Data una variabile casuale X con funzione di distribuzione $F(\cdot)$ i momenti pesati in probabilità sono definiti come:

$$M_{p,r,s} = E\{X^p [F(X)]^r [1-F(X)]^s\} \quad p, r, s \in \mathbb{R}$$

Il metodo PWM si basa sull'eguaglianza dei momenti pesati in probabilità della distribuzione con i momenti pesati in probabilità del campione.

Supponiamo di avere un campione X_1, \dots, X_n di variabili casuali iid distribuite con una GPD $F_{\xi\beta}(x)$ di parametri incogniti ξ, β .

Per stimare i due parametri (ξ, β) è sufficiente considerare i momenti pesati corrispondenti alle scelte $(p=1, r=0, s=0)$ e $(p=1, r=0, s=1)$ che risultano essere:

$$M_{1,0,0} = \frac{\beta}{(1-\xi)}$$

$$M_{1,0,1} = \frac{\beta}{2(2-\xi)}$$

Equazione 17

Le controparti empiriche a questi momenti corrispondono a:

$$\hat{M}_{1,0,0} = \frac{1}{n} \sum_{j=1}^n X_{j,n}$$

$$\hat{M}_{1,0,1} = \frac{1}{n} \sum_{j=1}^n X_{j,n} (1 - F_n(X_{j,n}))$$

Equazione 18

dove $X_{j,n}$ è il campione ordinato delle X_1, \dots, X_n tale che

$$X_{1,n} = \min(X_1, \dots, X_n), \quad X_{n,n} = \max(X_1, \dots, X_n)$$

e $F_n(X)$ è la funzione di probabilità empirica del campione. Eguagliando (2) e (3) otteniamo le stime dei due parametri:

$$\hat{\xi}_{PWM} = 2 - \frac{\hat{M}_{1,0,0}}{\hat{M}_{1,0,0} - 2\hat{M}_{1,0,1}}$$

$$\hat{\sigma}_{PWM} = \frac{2\hat{M}_{1,0,0}\hat{M}_{1,0,1}}{\hat{M}_{1,0,0} - 2\hat{M}_{1,0,1}}$$

Equazione 19 Parametri stimati con PWM

Come si vede dalla (3), $\hat{M}_{1,0,0}$ è la media campionaria delle X_1, \dots, X_n . Mentre la $\hat{M}_{1,0,1}$ si può riscrivere tenendo conto che per campioni ordinati di v.c. iid $X_{j,1}, \dots, X_{j,n}$

$$F_n(X_{j,n}) = \frac{j}{n}$$

Il metodo dei momenti lega il valore dei parametri della distribuzione ipotizzata ad alcuni valori caratteristici della distribuzione stessa. Questi valori vengono computati sul set di dati e quindi i parametri della distribuzione possono essere calcolati.

II. Intervalli di confidenza per lo stimatore dei momenti pesati in probabilità.

Utilizzando il teorema asintotico³⁸ è possibile definire intervalli di confidenza per le stime dei parametri della GPD.

Fissato il livello di confidenza α si definisce l'intervallo di confidenza al livello $(1-\alpha)\%$ per ξ :

$$\xi \pm \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \sqrt{\frac{\hat{v}_{1,1}}{N_t}}$$

Equazione 20

Analogamente per β si ha:

$$\beta \pm \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \sqrt{\frac{\hat{v}_{2,2}}{N_t}}$$

Equazione 21

Dove $\hat{v}_{1,1}, \hat{v}_{2,2}$ sono, rispettivamente, il primo o il quarto elemento diagonale della matrice delle varianze e covarianze asintotica per il PWM (definita in 7.3.3) empirica, ovvero ottenuta sostituendo ai parametri i loro stimatori, e N_t (dove t indica la soglia) è definito come:

$$N_t = \text{card}\{i : i = 1, \dots, n, X_i > t\}$$

Una stima più precisa si ottiene mediante l'utilizzo della profile likelihood definita come:

$$L_p(\xi) = \max_{\beta|\xi} (L(\xi, \beta))$$

Gli intervalli di confidenza sono perciò definiti come:

$$CI_\xi = \left\{ \xi : \log L_p(\xi) \geq \log L_p(\hat{\xi}) - \frac{\chi_1^2(1-\alpha)}{2} \right\}$$

Nel caso in cui i vincoli per l'esistenza della matrice di varianze e covarianze ($\xi < 1/2$) non siano soddisfatti è stato deciso di utilizzare un metodo numerico che consiste nel calcolo di intervalli

³⁸ Cfr. Paragrafo 5.5.3 Proprietà asintotiche degli stimatori MLE, PWM ed MPLE

basati sul campionamento (con ripetizione), e successivo fitting, dei dati sulla coda della distribuzione, in altre parole si approssima la distribuzione asintotica utilizzando un numero finito di simulazioni (almeno 1000).

5.1.3.5 Minimum Density Power Divergence Estimator (MDPDE)

Una delle definizioni di robustezza è dovuta a Tukey meglio nota come resistenza; uno stimatore è resistente se è scarsamente influenzato dalla presenza nel campione di pochi outlier o dalla presenza di un piccolo errore nelle osservazioni (esempio arrotondamenti).

Uno dei metodi per misurare il grado di robustezza è quello di utilizzare la funzione di influenza (Hampel 1974); questo indicatore è uno dei più importanti mezzi per misurare l'impatto di una piccola perturbazione del modello sull'output finale. La influence function (IF) di un funzionale multivariato T relativo ad una distribuzione F descrive l'effetto sulla stima di una contaminazione infinitesimale della funzione F nel punto x , ed è definita come:

$$IF(x; T, F) = \lim_{\varepsilon \downarrow 0} \frac{T(G) - T(F)}{\varepsilon}$$

Dove $G = (1 - \varepsilon)F + \varepsilon H(x)$ è una mistura del modello originale e di quello contaminato; $(1 - \varepsilon)$ è la probabilità di avere dei dati non contaminati e $H(x)$ è una distribuzione arbitraria "contaminata".

Uno stimatore robusto, nel modello teorico, non può essere più efficiente dello stimatore di massima verosimiglianza, perciò è comunque utile associare uno stimatore di massima verosimiglianza ad uno robusto, se le due stime non differiscono molto, il modello teorico è ben specificato, ed è preferibile l'utilizzo dello stimatore di massima verosimiglianza, altrimenti lo stimatore robusto è più affidabile.

Il pacchetto POT di R contiene una serie di stimatori più o meno robusti tra cui il Minimum Density Power Divergence Estimator. Questo stimatore è trattato approfonditamente in una tesi del 2003 (S.Juarez, 2003) ed è stato ripreso nel 2004 (S.Juarez and W.Schucany, 2004) con un'applicazione alla GPD nello studio dei dati sui millimetri di pioggia caduti.

Siano X_1, X_2, \dots, X_n , un campione casuale di eccessi estratto da una distribuzione non nota con densità g , il modello parametrico utilizzato per descrivere la distribuzione degli eccessi è quello della famiglia GPD:

$$f(x; \xi, \beta) = \frac{1}{\beta} \left(1 - \xi \frac{x}{\beta}\right)^{\frac{1}{\xi}-1} \quad x \in D(\xi, \beta) \quad (\xi, \beta) \in \theta \quad ^{39}$$

dove:

$$D(\xi, \beta) = [0, \infty) \text{ if } \xi \leq 0, \text{ e } D(\xi, \beta) = \left[0, \frac{\beta}{\xi}\right] \text{ if } \xi > 0; \quad \theta = \{(\xi, \beta) \in \mathbb{R}^2 : \xi \in \mathbb{R}, \beta > 0\}$$

A partire dallo stimatore power divergence (DPD) tra la densità f e g , definito, per α positivo, come:

³⁹ Si può ottenere la definizione classica (Embrechts et. al) ponendo nell'equazione ξ uguale a $(-\xi)$.

$$d_{\alpha}(g, f) = \int_{\mathcal{X}} \left\{ f^{1+\alpha}(x; \theta) - \left(1 + \frac{1}{\alpha}\right) g(x) f^{\alpha}(x; \theta) + \frac{1}{\alpha} g^{1+\alpha}(x) \right\} dx.$$

Si ottiene il minimo funzionale di DPD, $T_{\alpha}(G)$, come il punto in cui lo spazio del parametro corrispondente agli elementi di F (famiglia di densità) è il più vicino alla densità g :

$$d_{\alpha}(g, f(\cdot; T_{\alpha}(G))) = \inf_{\theta \in \Theta} d_{\alpha}(g, f(\cdot; \theta))$$

Siano (ξ_0, β_0) i parametri obiettivo. Per un α fissato $\alpha > 0$ (tuning constant) l' MDPDE per la GPD è il valore $(\hat{\xi}_{\alpha}, \hat{\beta}_{\alpha})$ che minimizza:

$$H_{\alpha}(\xi, \beta) = \frac{1}{\beta^{\alpha}(1 + \alpha - \alpha\xi)} - \left(1 + \frac{1}{\alpha}\right) \frac{1}{n} \sum_{i=1}^n \frac{1}{\beta^{\alpha}} \left(1 - \xi \frac{X_i}{\beta}\right)^{(\xi^{-1}-1)\alpha}$$

Sul dominio:

$$\left\{ (\xi, \beta) \in \Theta : \beta > 0, \quad \max_{1 \leq i \leq n} \{X_i\} \xi < \beta, \quad -\infty < \xi < 0, \quad 0 < \xi < \frac{1+\alpha}{\alpha} \right\}$$

Come l'MLE ed il PWM anche l'MDPDE soddisfa il teorema di normalità asintotica, come dimostra il seguente teorema:

Teorema: siano (ξ_0, β_0) i parametri target nel modello GPD; per ipotesi sia $\xi_0 < \frac{1+\alpha}{2+\alpha}$ per un $\alpha > 0$ fissato, e le condizioni di integrabilità siano soddisfatte. Allora esisterà una sequenza di stimatori MDPD $\{(\hat{\xi}_{\alpha,n}, \hat{\beta}_{\alpha,n})\}$ consistente con (ξ_0, β_0) per $n \rightarrow \infty$ e si avrà :

$$\sqrt{n}(\hat{\xi}_{\alpha,n} - \xi_0, \hat{\beta}_{\alpha,n} - \beta_0)^t \rightarrow N((0,0)^t, V(\xi_0, \beta_0))$$

dove $V(\xi_0, \beta_0) = J_{\alpha}^{-1}(\xi_0, \beta_0) K_{\alpha}(\xi_0, \beta_0) J_{\alpha}^{-1}(\xi_0, \beta_0)$.

Juarez e Schucany hanno definito in forma chiusa le due matrici simmetriche 2x2, K_{α} e J_{α} per la GPD⁴⁰:

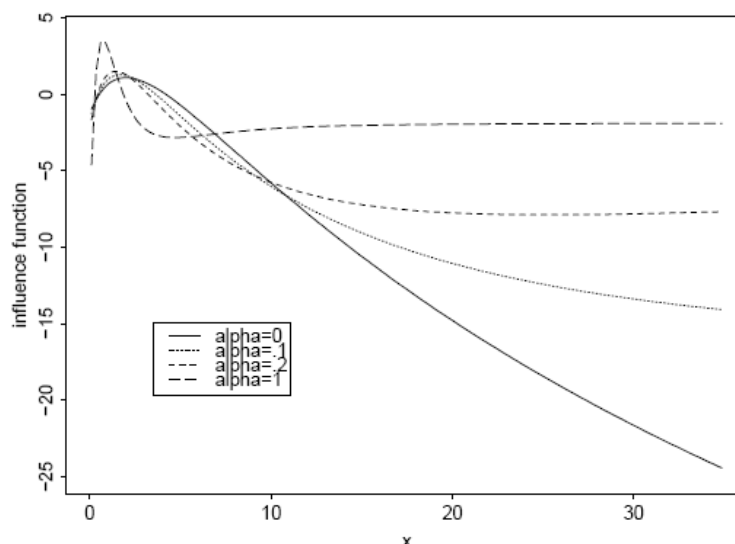
Sotto le condizioni di modello e fissato α la funzione di influenza dell'MDPDE per la GPD è data da:

$$IF(x; \alpha) = J_{\alpha}^{-1}(\xi, \beta) [S(x; \xi, \beta) f^{\alpha}(x; \xi, \beta) - U_{\alpha}(\xi, \beta)]$$

Per β fissato, la funzione di influenza per ξ è :

$$IF_{\xi}(x; \alpha) = \frac{S_{\xi}(x; \xi, \beta) f^{\alpha}(x; \xi, \beta) - \int_{D(\xi, \beta)} S_{\xi}(x; \xi, \beta) f^{1+\alpha}(x; \xi, \beta) dx}{\int_{D(\xi, \beta)} S_{\xi}^2(x; \xi, \beta) f^{1+\alpha}(x; \xi, \beta) dx}$$

⁴⁰ Per approfondimenti: Juarez, S. and Schucany, W. (2004). Le formule per il calcolo della varianza asintotica si trovano nell'appendice.



La figura precedente estratta dal lavoro di Juarez mostra l'assenza di limite inferiore per la funzione di influenza dell'MLE (linea continua) quando il parametro di scale è fissato. Un incremento arbitrario di una sola osservazione fa diminuire il valore del parametro di shape (nel grafico è riportato $-\text{csi}$) senza che questo raggiunga un limite inferiore. Analogamente la IF dell'MDPDE raggiunge un asintoto orizzontale più lentamente quanto più α è vicina allo 0 (per $\alpha = 0$ MLE e MDPDE coincidono).

Sempre nel lavoro di Juarez si dimostra che l'elemento al posto 1-1 di $V(\xi_0, \beta_0)$, (pari a $(1 - \xi_0)^2$ per MLE) dimostra che l'MDPDE ha una efficienza relativa più alta per valori piccoli di α , mentre per valori più elevati l'efficienza diventa inaccettabile, pertanto il valore generalmente assunto da questo parametro è 0,1.

5.1.3.6 Maximum Goodness of Fit Estimator (MGF)

Un altro stimatore robusto che è possibile affiancare agli stimatori classici è lo stimatore di massimo accostamento (Maximum Goodness of Fit) descritto da Luceno [12] in un articolo del 2005.

Questo metodo è basato sulla minimizzazione della distanza fra la funzione di distribuzione empirica e la teorica e si può usare anche nei casi in cui i metodi di massima verosimiglianza e/o dei momenti non sono applicabili.

La duttilità di questo metodo risiede nel fatto che, a seconda del tipo di definizione si dà alla distanza tra le distribuzioni, si ha un valore di stima diverso.

Ad esempio una statistica può permettere di dare maggiore peso alla coda sinistra della distribuzione, una alla coda destra oppure alla parte centrale, o ancora si può assegnare lo stesso peso a ogni punto.

Sia (x_1, \dots, x_n) un campione di n osservazioni IID di una variabile continua X con CDF $F(x)$. Siano $x(1), \dots, x(n)$ le statistiche ordinate corrispondenti e $S_n(x)$ la funzione di distribuzione empirica, si hanno le seguenti definizioni delle statistiche:

Table 1
Three classical EDF statistics

Statistic	Acronym	Formula
Kolmogorov distance	KS	$D_n = \sup_x F(x) - S_n(x) $
Cramer-von Mises	CM	$W_n^2 = n \int_{-\infty}^{\infty} \{F(x) - S_n(x)\}^2 dF(x)$
Anderson-Darling	AD	$A_n^2 = n \int_{-\infty}^{\infty} \frac{\{F(x) - S_n(x)\}^2}{F(x) \{1 - F(x)\}} dF(x)$

Table 2
Modified Anderson-Darling statistics

Statistic	Acronym	Formula
Right-tail AD	ADR	$R_n^2 = n \int_{-\infty}^{\infty} \frac{\{F(x) - S_n(x)\}^2}{1 - F(x)} dF(x)$
Left-tail AD	ADL	$L_n^2 = n \int_{-\infty}^{\infty} \frac{\{F(x) - S_n(x)\}^2}{F(x)} dF(x)$
Right-tail AD of second degree	AD2R	$r_n^2 = n \int_{-\infty}^{\infty} \frac{\{F(x) - S_n(x)\}^2}{\{1 - F(x)\}^2} dF(x)$
Left-tail AD of second degree	AD2L	$l_n^2 = n \int_{-\infty}^{\infty} \frac{\{F(x) - S_n(x)\}^2}{\{F(x)\}^2} dF(x)$
AD of second degree	AD2	$a_n^2 = r_n^2 + l_n^2$

La statistica AD dà più peso alle code della CDF rispetto al CM, analogamente le statistiche ADR e ADL assegnano un maggiore peso alla coda selezionata rispetto al CM.

Usando la notazione $z_i = F(x(i))$ e tenendo presente che $S_n(x)$ è una funzione a tratti con salti corrispondenti alle statistiche ordinate, le statistiche EDF descritte nelle tabelle 1 and 2 possono essere riscritte come segue:

Table B.1
Computational forms for the EDF statistics

Acronym	Formula
KS	$D_n = \frac{1}{2n} + \max_{1 \leq i \leq n} \left z_i - \frac{i - 1/2}{n} \right $
CM	$W_n^2 = \frac{1}{12n} + \sum_{i=1}^n \left(z_i - \frac{i - 1/2}{n} \right)^2$
AD	$A_n^2 = -n - \frac{1}{n} \sum_{i=1}^n (2i - 1) \{ \ln z_i + \ln (1 - z_{n+1-i}) \}$
ADR	$R_n^2 = \frac{n}{2} - 2 \sum_{i=1}^n z_i - \frac{1}{n} \sum_{i=1}^n (2i - 1) \ln (1 - z_{n+1-i})$
ADL	$L_n^2 = -\frac{3n}{2} + 2 \sum_{i=1}^n z_i - \frac{1}{n} \sum_{i=1}^n (2i - 1) \ln z_i$
AD2R	$r_n^2 = 2 \sum_{i=1}^n \ln (1 - z_i) + \frac{1}{n} \sum_{i=1}^n \frac{2i - 1}{1 - z_{n+1-i}}$
AD2L	$l_n^2 = 2 \sum_{i=1}^n \ln z_i + \frac{1}{n} \sum_{i=1}^n \frac{2i - 1}{z_i}$
AD2	$a_n^2 = 2 \sum_{i=1}^n \{ \ln z_i + \ln (1 - z_i) \} + \frac{1}{n} \sum_{i=1}^n \left(\frac{2i - 1}{z_i} + \frac{2i - 1}{1 - z_{n+1-i}} \right)$

La società di consulenza Quantide ha eseguito uno studio sui dati Montepaschi per scegliere la misura di accostamento più adatta per la stima dei parametri della GPD. L'analisi tecnica è disponibile nei report allegati.

I criteri utilizzati per stabilire se una misura di accostamento è più o meno preferibile rispetto alle altre sono principalmente: la vicinanza dei quantili stimati con i quantili empirici, e il fatto

che le stime dei parametri non portino a risultati in qualche modo degeneri. Se uno stimatore (o una misura di GOF) produce stime degeneri o non sensate, questo viene scartato.

Non esiste un metodo che renda preferibile o meno in assoluto una statistica rispetto alle altre, pertanto è stata eseguita un'analisi utilizzando i dati Montepaschi ed inserendo perturbazioni nelle serie storiche. Dalle analisi emerge come le misure di GOF più adatte a descrivere i dati Montepaschi siano quella di Anderson-Darling Superiore su coda destra, la misura di Anderson-Darling quadratica e la misura di Cramer-Von Mises. Le ultime due ottengono stime piuttosto stabili anche in presenza di perturbazioni nelle serie storiche.

Per effettuare una ulteriore scrematura è stata valutata la variabilità nella stima dei parametri calcolata utilizzando campioni estratti casualmente dalla GPD (di parametri stimati con MLE sui dati originali), di numerosità variabile da 40 a 200 dati; per ogni numerosità sono stati simulati 1000 campioni indipendenti e su di questi sono state calcolate media e varianza delle stime ed errore standard (per valutare l'efficienza degli stimatori).

La variabilità delle stime dei parametri ξ e β al variare dello scenario rappresentata come deviazione standard mostra come le stime MGF basate sulle misure AD2 e CVM risultino più stabili rispetto alle altre; nel caso AD2 la variabilità è abbastanza prossima a quella dell'MLE (con una riduzione in efficienza circa del 20%) che ci fa propendere per la scelta di quest'ultimo metodo.

5.1.4 Test Statistici

5.1.4.1 Test di autocorrelazione

1.1.1.1.1 Augmented Dickey Fueller test

Sia $X_t, t = 0, 1, \dots$ un processo definito da

$$\begin{aligned}X_0 &= x_0 \\X_t &= \delta + X_{t-1} + \varepsilon_t\end{aligned}$$

Con ε_t rumore bianco a varianza σ^2 . Questo processo è una passeggiata aleatoria con deriva (Random Walk with Drift) dove la deriva è rappresentata dal parametro δ . Se $\delta = 0$ si ha una passeggiata aleatoria. Procedendo per sostituzione si ottiene una formulazione equivalente del processo data da:

$$X_t = x_0 + t\delta + \sum_{i=1}^t \varepsilon_i$$

Ovvero la RWD è data dalla somma di una retta $x_0 + t\delta$, trend deterministico, nel senso che, conoscendo x_0 e δ è perfettamente prevedibile, e di una passeggiata aleatoria che è un *trend stocastico* dato che non è mai perfettamente prevedibile.

In una passeggiata aleatoria il valore iniziale x_0 e gli shock ε_t vengono per sempre “inglobati” nel processo. Infatti operando le stesse sostituzioni in un processo AR(1) stazionario, cioè con $|\phi| < 1$:

$$X_t = \delta + \phi X_{t-1} + \varepsilon_t$$

Otterremmo:

$$X_t = \phi^t x_0 + \delta \frac{1 - \phi^t}{1 - \phi} + \sum_{i=0}^{t-1} \phi^i \varepsilon_{t-i} \quad 41$$

che al crescere di t tende a “scordare” l’effetto del valore iniziale x_0 e di ogni shock ε_{t-i}

sufficientemente remoti nel tempo (per i grande). Per t sufficientemente grande $\delta \frac{1 - \phi^t}{1 - \phi}$ è

indistinguibile dalla media del processo AR(1) $\frac{\delta}{1 - \phi}$.

Il valore atteso, la varianza e l’autocovarianza di un RWD (random walk with drift) inizializzato dal valore $X_0 = x_0$ sono pari a

$$E[X_t] = x_0 + \delta t$$

$$Var[X_t] = t\sigma^2$$

$$Cov[X_t, X_s] = t\sigma^2 \quad s > t$$

I momenti di X_t dipendono da t quindi il processo non è stazionario.

Quando una serie è generata da un trend deterministico sommato ad un processo ARMA stazionario,

$$\phi(B)(X_t - \mu - \beta t) = \theta(B)\varepsilon_t$$

si dice che il processo X_t è *stazionario attorno a un trend* (trend stationary) o che ha un *trend puramente deterministico*.

Quando un processo ARMA ha $d \geq 1$ radici unitarie pari a 1, cioè

$$\varphi_{p+d}(B)(X_t - \delta) = \theta_q(B)\varepsilon_t$$

$$\text{con } \varphi_{p+d}(B) = \phi_p(B)(1 - B)^d$$

si dice integrato di ordine d , indicato con $I(d)$, o a *differenza stazionaria* (difference stationary), oppure dotato di *trend stocastico*.

Il test classico parte dal caso più semplice in cui si voglia testare l’ipotesi che una serie storica sia stata generata da un RW contro l’ipotesi essa provenga da un processo AR(1) stazionario con media nulla:

$$X_t = X_{t-1} + \varepsilon_t$$

$$X_t = \phi X_{t-1} + \varepsilon_t \quad |\phi| < 1$$

⁴¹ $\sum_{j=0}^{t-1} \phi^j = \frac{1 - \phi^t}{1 - \phi}$

Si dimostra che la quantità $n(\hat{\phi}_n - 1)$, dove $\hat{\phi}_n$ è lo stimatore dei minimi quadrati di ϕ , si distribuisce come il rapporto di un chi quadro centrata e riscalata e una distribuzione non standard. Questa distribuzione prende il nome di Dickey Fuller di primo tipo e i valori critici sono tabulati in diversi testi.

Solitamente invece di stimare il modello $X_t = \phi X_{t-1} + \varepsilon_t$ si utilizza :

$$\Delta X_t = \rho X_{t-1} + \varepsilon_t \text{ con } \rho = \phi - 1.$$

Equazione 22

Il test di Dickey-Fuller pone in alternativa due processi dalla dinamica piuttosto limitata: escludendo le componenti deterministiche si confronta un RW contro un AR(1) stazionario. Sarebbe più utile testare la presenza di una radice unitaria all'interno di un generico ARMA(p ; q) o almeno AR(p). Il test DF può essere generalizzato solamente a processi AR(p), che tuttavia per p sufficientemente grande possono approssimare ragionevolmente bene processi ARMA.

Un processo AR(p) ha (almeno) una radice unitaria quando l'equazione caratteristica $\phi_p(z) = 0$ è verificata per $z = 1$, da cui $1 - \phi_1 - \dots - \phi_p = 0$.

Quindi, se si riuscisse a ri-parametrizzare il modello AR(p) in modo da avere un parametro $\rho = \phi_1 + \dots + \phi_p - 1$, si potrebbe costruire un test di significatività per l'ipotesi $\rho = 0$.

L'operatore auto regressivo $\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ può essere riscritto come

$$\phi_p(B) = (1 - \phi_1 - \dots - \phi_p)B + (1 - \alpha_1 B - \dots - \alpha_{p-1} B^{p-1})(1 - B)^{42}$$

eguagliando i termini di medesimo grado:

$$\alpha_{p-1} = -\phi_p$$

$$\alpha_{p-2} - \alpha_{p-1} = -\phi_{p-1}$$

$$\alpha_{p-3} - \alpha_{p-2} = -\phi_{p-2}$$

...

Scrivendo il processo AR(p) per mezzo della rappresentazione precedente del polinomio autoregressivo otteniamo

$$\Delta X_t = \rho X_{t-1} + \alpha_1 \Delta X_{t-1} + \dots + \alpha_{p-1} \Delta X_{t-(p-1)} + \varepsilon_t$$

Equazione 23

dove si è posto $\rho = \phi_1 + \dots + \phi_p - 1$.

Questo test prende il nome di Dickey-Fuller aumentato (ADF da Augmented Dickey-Fuller) e può essere generalizzato con la presenza di componenti deterministiche. Le distribuzioni asintotiche sono tabulate in diversi testi.

⁴² Svolgendo l'equazione ed eguagliando i termini di medesimo grado si riottiene la formulazione originale.

La selezione del numero dei ritardi p viene normalmente fatta per mezzo del criterio di informazione di Schwarz: si stimano diversi modelli auto regressivi $AR(k)$, per $k = 1, \dots, K$, nella forma definita e si sceglie il valore di k per cui lo Schwarz è minimo. I test t sui coefficienti α_i hanno la usuale distribuzione gaussiana standard.

1.1.1.1.2 *t-test nell'analisi di regressione*

Nel caso si voglia testare se lo slope della retta di regressione $Y_i = \alpha + \beta x_i + \varepsilon_i$ sia diverso da zero si può usare questa statistica.

Supponiamo che le x_i siano note, α e β siano da stimare, le ε_i siano indipendenti e distribuite normalmente (errori casuali), con valore atteso nullo e varianza σ^2 .

Si vuole testare se il valore di β è uguale ad un valore β_0 specifico, spesso posto uguale a 0 per verificare l'indipendenza tra x ed y .

Utilizzando gli stimatori dei minimi quadrati per α e β , definendo lo stimatore dei residui

$\hat{\varepsilon} = Y_i - (\hat{\alpha} + \hat{\beta}x_i)$ e definendo SSE (somma dei quadrati dei residui) = $\sum_{i=1}^n \hat{\varepsilon}^2$ si ottiene la statistica:

$$\frac{(\hat{\beta} - \beta_0)\sqrt{n-2}}{\sqrt{SSE \sum_{i=1}^n (x_i - \bar{x})^2}}$$

Equazione 24

Che si distribuisce come una t di Student con $n-2$ gradi di libertà se l'ipotesi nulla è vera.

1.1.1.1.3 *Durbin Watson*

Sebbene un andamento decrescente del correlogramma sia un buon modo di percepire la componente di trend, per accertare formalmente la stazionarietà si deve ricorrere a dei test formali, utili a verificare congiuntamente l'assenza di autocorrelazione seriale per più lag. Il test DW si usa nei modelli di regressione dei minimi quadrati per verificare l'ipotesi nulla di indipendenza tra i residui del modello ($\rho = 0$), contro l'ipotesi alternativa di autocorrelazione del primo ordine ($\rho \neq 0$).

Serve dunque per verificare l'assenza di un legame lineare tra ciascun residuo ed il precedente. La statistica di DW ha la seguente forma

$$DW = \frac{\sum_{t=2}^n (u_t - u_{t-1})^2}{\sum_{t=1}^n u_t^2}$$

Equazione 25

In cui u_t è il residuo relativo al periodo t ed n è il numero delle osservazioni. Il DW è più efficace se espresso come test su

$$u_t = \rho u_{t-1} + \varepsilon_t$$

In questo caso se si accetta l'ipotesi $\rho = 0$, il test dà luogo ad un processo white noise.

Il valore della statistica di Durbin-Watson è sempre compreso tra 0 e 4.

Un valore di 2 indica che non appare presente alcuna autocorrelazione. Valori piccoli di DW indicano che i residui successivi sono, in media, vicini in valore l'uno all'altro, o correlati positivamente. Valori grandi di DW indicano che i residui successivi sono, in media, molto differenti in valore l'uno dall'altro, o correlati negativamente.

La distribuzione teorica della statistica di Durbin-Watson non è nota; tuttavia gli stessi Durbin e Watson hanno tabulato, con un esercizio di simulazione condotto col metodo Monte Carlo, i valori critici della statistica. Per verificare la presenza di **autocorrelazione positiva** al livello di significatività α , la statistica test DW viene confrontata con dei valori critici inferiori e superiori ($d_{L,\alpha}$ and $d_{U,\alpha}$):

- Se $DW < d_{L,\alpha}$ si ha una prova statistica di autocorrelazione positiva degli errori.
- Se $DW > d_{U,\alpha}$, si ha una prova statistica di **non** autocorrelazione positiva degli errori.
- Se $d_{L,\alpha} < DW < d_{U,\alpha}$ il test non è conclusivo.

Per verificare la presenza di **autocorrelazione negativa** al livello di significatività α , la statistica test d viene confrontata con dei valori critici inferiori e superiori ($d_{L,\alpha}$ and $d_{U,\alpha}$):

- Se $(4 - DW) < d_{L,\alpha}$ si ha una prova statistica di autocorrelazione negativa degli errori.
- Se $(4 - DW) > d_{U,\alpha}$, si ha una prova statistica di **non** autocorrelazione negativa degli errori.
- Se $d_{L,\alpha} < (4 - DW) < d_{U,\alpha}$ il test non è conclusivo.

I valori critici $d_{L,\alpha}$ e $d_{U,\alpha}$ variano secondo il livello di significatività (α), secondo il numero di osservazioni e il numero di parametri nell'equazione di regressione e vengono generalmente ottenuti da apposite tavole⁴³.

1.1.1.1.4 Il Q-Test di Ljung-Box

La stazionarietà di una serie comporta l'assenza di autocorrelazione seriale al variare del lag k , questo fa sì che il valore assunto dalla serie in t non influenza il valore assunto in un istante successivo. Le ipotesi che si prendono di riferimento sono dunque: $\rho_k = 0$ per ogni valore di k e $\rho_k \neq 0$ per almeno un valore di k .

La statistica Q di Ljung-Box formalizza tali ipotesi nella seguente forma:

⁴³ L'università di Stanford ha pubblicato valori critici della statistica fino al lag 21 sul sito <http://www.stanford.edu/~clint/bench/dwcrit.htm>.

Il software R utilizza invece il metodo bootstrap per la stima del p value.

$$Q = n(n+2) \sum_{i=1}^k \frac{\rho_i^2}{n-i}$$

Equazione 26

in cui n è la numerosità del campione⁴⁴. Q si distribuisce asintoticamente come un χ^2 con k gradi di libertà. L'ipotesi di assenza di autocorrelazione è rifiutata se $Q > \chi_{(1-\alpha),k}^2$

5.1.4.2 Test sulla bontà di adattamento

I test statistici per verificare la bontà di adattamento di un fitting alla distribuzione empirica dei dati vengono solitamente divisi in tre classi, a seconda del tipo di statistica su cui si basano:

- test parametrici;
- test non parametrici;
- criteri informativi.

1.1.1.1.5 Test parametrici

1. Chi-Quadro

Siano x_i e y_i ($i = 1 \dots N$) i valori e le frequenze registrate nei dati. Sia $F(x_i)$ l'altezza del bin dell'istogramma della funzione F con cui si vuole approssimare la distribuzione empirica nel punto x_i . F sia funzione di v parametri. Definiamo come numero di gradi di libertà del problema la quantità $d = N - v$. Il valore del Chi-quadro per gli N punti è pari a:

$$\chi_N^2 = \sum_{i=1}^N \frac{(y_i - F(x_i))^2}{F(x_i)}.$$

Equazione 27

Tale valore viene usualmente riportato in forma ridotta per rendere immediatamente confrontabili test su problemi diversi. Si definisce perciò il Chi-quadro ridotto per d gradi di libertà:

$$\chi_d^2 \equiv \frac{\chi_N^2}{d} = \frac{\sum_{i=1}^N \frac{(y_i - F(x_i))^2}{F(x_i)}}{d}.$$

Equazione 28

Nel caso l'ipotesi nulla fornisca un valore di Chi-quadro ridotto con probabilità cumulata pari ad α , si può affermare che tale ipotesi non può essere rifiutata ad un livello di confidenza di $1-\alpha$.

⁴⁴ Nel nostro caso ρ indica il coefficiente di autocorrelazione della serie al lag i .

1.1.1.1.6 Test non parametrici

1. Kolmogorov – Smirnov.

Indichiamo con F la CDF teorica e con \hat{F}_n la CDF empirica ottenuta con un campione di n valori. La statistica che si studia è quella legata alla variabile

$$D_n = \sup_x |F(x) - \hat{F}_n(x)|.$$

Equazione 29

Questo valore va confrontato con la statistica della variabile di Kolmogorov – Smirnov corrispondente. Se il valore è troppo alto il test non viene superato e viene rigettata l'ipotesi che il campione segua la statistica legata alla funzione F .

2. Cramer – von Mises

Il test di Cramer – von Mises valuta la “distanza” tra due funzioni di ripartizione tramite l'integrale

$$I = \int_{-\infty}^{+\infty} [\hat{F}(x) - F(x)]^2 dF(x).$$

Nel caso di un campione x_1, \dots, x_n , ordinato per valori crescenti, questa quantità viene stimata tramite la formula

$$W_n^2 = \frac{1}{12n} + \sum_{i=1}^n \left[\frac{2i-1}{2n} - F(x_i) \right]^2,$$

Equazione 30

che valuta quindi la bontà del fit del campione tramite la funzione F . Il valore di questo indicatore viene quindi confrontato con quello della sua statistica caratteristica. Se il valore è troppo alto il test non viene superato e viene rigettata l'ipotesi che il campione segua la statistica legata alla funzione F .

3. Anderson – Darling.

Anche il test di Anderson – Darling si utilizza per testare l'ipotesi che un campione provenga da una certa distribuzione di probabilità. Indichiamo con x_1, \dots, x_n le realizzazioni del campione ordinate per valori crescenti e con F la funzione di ripartizione della variabile aleatoria da cui si suppone provenga il campione. Il valore che va calcolato è dato da

$$A_n^2 = -n - s,$$

Equazione 31

dove s è dato da

$$s = \sum_{i=1}^n \frac{2i-1}{n} [\ln(F(x_i)) + \ln(1 - F(x_{n+1-i}))].$$

Equazione 32

Come al solito il valore ottenuto per A_n va confrontato con il valore critico della funzione di ripartizione della statistica-test A_n per determinare l'esito del test. Da rilevare che la statistica-test varia al variare della statistica della funzione F da cui si suppone provengano i valori x_1, \dots, x_n . I valori critici per cui il test è superato o no, sono quindi dipendenti dalla distribuzione della variabile che si sta testando.

1.1.1.1.7 Criteri informativi

1. Criterio informativo di Akaike (AIC).

Il valore che viene preso in considerazione per valutare il miglior modello è

$$A = 2k - 2\ln(L),$$

Equazione 33 AIC

in cui k è il numero di parametri ed L è il valore della funzione di verosimiglianza calcolata nel punto di massimo. Il modello che viene preferito è quello che restituisce il minore valore di A .

2. Criterio di Schwarz (SIC) o Bayesian Criterion (BIC).

Esso è analogo al precedente ma in questo caso per la scelta del modello si utilizza il valore

$$S = k \ln(n) - 2\ln(L),$$

Equazione 34 BIC

dove k ed L hanno lo stesso significato del caso precedente ed in più n è il numero di osservazioni. Anche in questo caso è preferito il modello che restituisce il valore minore di S .

1.1.1.1.8 Test parametrici avanzati

In molte applicazioni il dataset analizzato è incompleto nel senso che le osservazioni sono registrate solo se eccedono una certa soglia pre determinata, inoltre spesso è più importante testare la bontà del fitting sulla coda destra della distribuzione, ad esempio quando, come nel nostro caso, si cerca di determinare il VaR.

Si considerano perciò test d'ipotesi che, tenendo conto di questo vincolo, permettono di testare se la distribuzione empirica proviene o meno da una determinata famiglia di distribuzioni.

Sappiamo che l'ipotesi nulla da testare (appartenenza alla famiglia) è rifiutata se il p-value è più basso di un livello critico α , che di solito varia tra il 5% e il 10%. Se indichiamo con D il valore della statistica osservata (calcolata sul campione a disposizione) e con d il valore critico per un dato livello α , allora il p-value è calcolato come: $\text{Prob}(D \geq d)$. Dato che la distribuzione della

statistica non è indipendente dal parametro, un metodo per calcolare p-value e valore critico è attraverso una simulazione Monte Carlo.

Si definisce il valore di α , si calcola il valore D e si applica il seguente algoritmo:

1. Si genera dalla distribuzione teorica fittata un numero elevato di campioni (almeno 1000) di ampiezza n , uguale al numero di dati osservati.
2. Si fitta la distribuzione teorica sui nuovi dati e si stimano i nuovi parametri per ogni campione;
3. Si stima il valore della statistica test D_i per ogni campione i ;
4. Si calcola il p-value come la proporzione delle volte in cui la statistica campionaria supera il valore osservato D del campione originale;
5. Si rifiuta l'ipotesi nulla se il p-value è più basso di α

3. Supremum Kolmogorov – Smirnov Test.

Questa statistica misura il valore assoluto della massima distanza tra la distribuzione empirica e quella fittata e assegna uguale peso a ciascuna osservazione. La formula proposta tiene conto del fatto che i dati sono troncati a sinistra.

Indichiamo con $F_n(x)$ la CDF empirica e con $\hat{F}_n^*(x)$ la CDF teorica troncata. La statistica campionaria è definita come:

$$KS^* = \sqrt{n} \sup_x |F_n(x) - \hat{F}_n^*(x)|.$$

E può essere calcolata come:

$$\begin{aligned} KS^{+*} &= \sqrt{n} \sup_j \{F_n(x_{(j)}) - \hat{F}_n^*(x_{(j)})\} = \frac{\sqrt{n}}{1 - z_H} \sup_j \left\{ z_H + \frac{j}{n}(1 - z_H) - z_j \right\}, \\ KS^{-*} &= \sqrt{n} \sup_j \{\hat{F}_n^*(x_{(j)}) - F_n(x_{(j)})\} = \frac{\sqrt{n}}{1 - z_H} \sup_j \left\{ z_j - \left(z_H + \frac{j-1}{n}(1 - z_H) \right) \right\}, \end{aligned}$$

Dove H è il valore soglia, z_H è il percentile della distribuzione teorica calcolato in H e n la dimensione del campione soprasoglia.

4. Upper Quadratic Anderson – Darling test.

A differenza della precedente statistica che assegnava uguale peso a tutti gli elementi del campione, questa statistica permette di testare l'accostamento sulla coda della distribuzione, poiché è ponderata per la probabilità che rimane sulla coda.

Sia $\{X_j\}_{1 \leq j \leq n}$ un vettore di n statistiche ordinate, la forma generale della statistica per campioni completi è:

$$AD_{\text{up}}^2 = n \int_H^{\infty} \frac{(F_n(x) - \hat{F}(x))^2}{(1 - \hat{F}(x))^2} d\hat{F}(x)$$

Mentre la forma per campioni troncati a sinistra si dimostra essere pari a :

$$AD_{\text{up}}^{2*} = -2n \log(1 - z_H) + 2 \sum_{j=1}^n \log(1 - z_j) + \frac{1 - z_H}{n} \sum_{j=1}^n (1 + 2(n - j)) \frac{1}{1 - z_j}.$$

Dove H è il valore soglia, z_H è il percentile della distribuzione teorica calcolato in H , z_j è il percentile della distribuzione teorica completa in j e n la dimensione del campione soprasoglia.

5.1.5 Extreme Value Theory

5.1.5.1 Logica Block Maxima

La logica Block Maxima si concentra sulla distribuzione del massimo di una serie di osservazioni indipendenti.

Sia X_1, X_2, \dots, X_n una successione di variabili casuali non degeneri con funzioni di ripartizione date da $F_{X_i} = P\{X_i \leq x\}$. Esse rappresentano le osservazioni campionarie che saranno oggetto dell'analisi, cioè, in altre parole, le varie perdite da rischi operativi osservate in un dato periodo di riferimento.

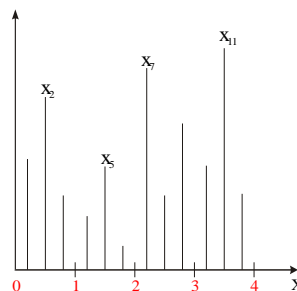
La distribuzione del massimo $M_n = \max(X_1, \dots, X_n)$ sarà data da:

$$F_{M_n}(m) = P\{M_n \leq m\} = P\{X_1 \leq m, \dots, X_n \leq m\}$$

se si assume che le X_i siano indipendenti ed identicamente distribuite (IID) si avrà:

$$F_{M_n}(m) = [F_X(m)]^n$$

La distribuzione esatta dei massimi dipende quindi dalle distribuzioni marginali delle variabili



casuali X_1, X_2, \dots, X_n (osservazioni campionarie della variabile casuale “perdite”). La ricerca di una distribuzione asintotica (per n tendente all'infinito) è motivata dalla necessità di ottenere una distribuzione degli estremi che non dipenda dalle distribuzioni marginali F_{X_i} , così da non dover fare alcuna assunzione sulla forma distributiva originaria di tutte le osservazioni.

Si dice che la variabile casuale X (ovvero la sua funzione di ripartizione F) appartiene al massimo dominio d'attrazione (MDA) della distribuzione dei valori estremi H , se a questa converge la distribuzione del massimo M_n (così come sopra definita) opportunamente riscalata:

$$c_n^{-1}(M_n - d_n) \xrightarrow{d} H \quad \text{con } c_n > 0 \text{ e } d_n \in \mathcal{R}.$$

dove c_n e d_n sono due successioni numeriche opportune.

Il teorema di *Fisher – Tippet* dimostra che la distribuzione dei massimi converge asintoticamente alla seguente:

$$H_\xi(x) = \begin{cases} \exp\left[-(1 + \xi x)^{-1/\xi}\right] & \text{se } \xi \neq 0, \\ \exp[-\exp(-x)] & \text{se } \xi = 0, \end{cases}$$

con $1 + \xi x > 0$

detta GEV (Generalized Extreme Value), distribuzione generalizzata dei valori estremi.

Fondamentale in essa è il parametro ξ , anche detto *tail index*, poiché indica lo spessore della coda della distribuzione; più grande è ξ , più spessa sarà la coda. La distribuzione H è detta generalizzata poiché racchiude in sé i tre tipi di distribuzione limite individuati dalla letteratura:

- se ξ è nullo, la H corrisponde ad una Gumbel; al MDA della Gumbel appartengono le distribuzioni la cui coda decresce con legge esponenziale (cioè a coda leggera), quali per esempio la Normale, la Lognormale, la Gamma;
- se ξ è negativo, la H corrisponde ad una Weibull; al MDA della Weibull appartengono le distribuzioni con supporto limitato, come ad esempio la Uniforme, la Beta;
- se ξ è positivo, la H corrisponde ad una Frechet, al MDA della quale appartengono distribuzioni a coda spessa (o pesante) quali Pareto, Burr, Log-gamma, Cauchy.

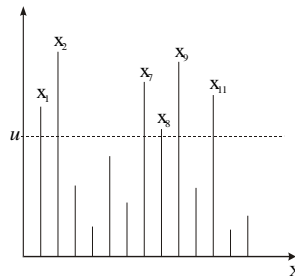
5.1.5.2 Logica Peaks Over Threshold (POT)

Un gruppo più moderno di modelli è rappresentato invece dalla classe dei modelli POT, i quali sono generalmente considerati più utili per le applicazioni pratiche, dato l'uso più efficiente che essi fanno dei dati (spesso limitati) a disposizione sui valori estremi.

Sia X una variabile casuale con una funzione di ripartizione F . Fissata una soglia u , la distribuzione condizionale:

$$F_u(y) = P(X - u \leq y \mid X > u), \quad y \geq 0$$

è l'*excess distribution function* della variabile casuale X (o della funzione di distribuzione F) sopra la soglia u .



Il teorema di *Pickands – Balkema – de Haan* dimostra che, per un'ampia classe di distribuzioni F , la distribuzione asintotica limite (per u che tende all'infinito) della excess distribution function $F_u(y)$ è rappresentata dalla seguente funzione:

$$G_{\xi}(y) = \begin{cases} 1 - [1 + \xi(y - \nu)/\beta]^{-1/\xi} & \text{se } \xi \neq 0 \\ 1 - e^{-(y-\nu)/\beta} & \text{se } \xi = 0 \end{cases}$$

dove

$$\begin{array}{ll} y \geq 0 & \xi \geq 0 \\ 0 \leq y \leq -1/\xi & \xi < 0 \end{array}$$

detta Generalized Pareto Distribution (GPD).

Come nel caso della GEV, al variare del segno di ξ cambiano le caratteristiche della distribuzione limite; in particolare il caso di maggior interesse dal punto di vista del risk management è quello in cui $\xi > 0$, corrispondente ad una GPD a coda spessa.

Gli altri due parametri che compaiono nell'espressione della GPD sono un parametro di *scale* (β) ed un parametro di *location* (ν), la cui variazione comporta, rispettivamente, un cambiamento di scala ed una traslazione della GPD.

In entrambi i casi l'importante conclusione a cui si perviene è la conoscenza della forma distributiva (asintotica) della coda della distribuzione campionaria, senza la necessità di fare alcuna supposizione sulla distribuzione delle perdite osservate. Ciò a cui si è interessati è la stima dei parametri caratterizzanti la suddetta forma distributiva asintotica, che nella logica Block Maxima è una Generalized Extreme Value distribution (GEV), mentre nella logica Peaks Over Threshold è una Generalized Pareto Distribution (GPD).

5.2 Approfondimenti sulle distribuzioni utilizzate

5.2.1 Distribuzioni di frequenza

5.2.1.1 Poisson

Nell'ipotesi che la frequenza di accadimento di un evento si distribuisca come una Poisson si hanno:

densità di probabilità:

$$f_X(x) = f_X(x, \lambda) = \begin{cases} \frac{e^{-\lambda} \cdot \lambda^x}{x!} & \text{per } x = 0, 1, 2, \dots \\ 0 & \text{altrimenti} \end{cases} = \frac{e^{-\lambda} \cdot \lambda^x}{x!} I_{\{0,1,\dots\}}(x) \quad \text{per } \lambda > 0$$

funzione di ripartizione:

$$F_X(x) = \sum_{(j=x_j \leq x)} f_X(x_j) = \sum_{x_j=0}^x \frac{e^{-\lambda} \cdot \lambda^{x_j}}{x_j!}$$

Dal momento che per tale distribuzione valgono le seguenti relazioni

Media = varianza = λ ⁴⁵

si suppone valida la seguente relazione fra l'unico parametro di tale distribuzione e la stima raccolta sulla frequenza media di accadimento tramite RSA:
stima soggettiva frequency media = λ

5.2.1.2 Binomiale Negativa

La frequenza di accadimento dell'evento può distribuirsi anche secondo una binomiale negativa, che, come la Poisson, descrive l'accadimento di eventi rari; è descritta dalla funzione di densità:

$$p_r(k) = \binom{r+k-1}{k} p^r (1-p)^k$$

Con k e r interi non negativi e p che assume valori in (0,1). Descrive la probabilità di eseguire r prove per ottenere k successi.

La media e la varianza sono date da:

$$\mu = \frac{rq}{p}$$

⁴⁵ Media e varianza della distribuzione, rispettivamente momento primo e secondo, coincidono e sono pari all'unico parametro della distribuzione stessa.

$$\sigma^2 = \frac{rq}{p^2} = \frac{\mu}{p}$$

Con $q=(1-p)$.

Gli indici di simmetria e curtosi sono definiti, rispettivamente, da:

$$\beta_1 = \frac{(1+q)^2}{rq}$$

$$\beta_2 = 3 + \frac{(1+4q+q^2)}{rq}$$

5.2.2 Distribuzioni di severity

5.2.2.1 Weibull

Nell'ipotesi che l'impatto del singolo evento ("sev(t)") si distribuisca come una Weibull:

$$(\text{densità di probabilità}) \text{ sev}(t) = \frac{\text{shape}}{\text{scale}} \cdot \left(\frac{t}{\text{scale}}\right)^{\text{shape}-1} \cdot \exp\left\{-\left(\frac{t}{\text{scale}}\right)^{\text{shape}}\right\}$$

$$(\text{funzione di ripartizione}) \text{ SEV}(t) = 1 - \exp\left\{-\left(\frac{t}{\text{scale}}\right)^{\text{shape}}\right\}$$

In particolare il momento primo (media) della distribuzione può essere così espresso:

$$(\text{impatto medio}) \text{ sev} = \text{scale} \cdot \Gamma\left(1 + \frac{1}{\text{shape}}\right)$$

dove si definisce "funzione gamma" la seguente espressione:

$$\Gamma(t) = \int_0^{\infty} x^{t-1} \cdot e^{-x} \cdot dx \quad , \quad t > 0$$

A partire dalla funzione di ripartizione, sopra riportata, si ottiene la funzione inversa (espressione dei quantili in funzione del livello di probabilità prescelto) con i seguenti passaggi:

$$\text{SEV}(t) = 1 - \exp\left\{-\left(\frac{t}{\text{scale}}\right)^{\text{shape}}\right\} \quad (\text{funzione di ripartizione})$$

$$\ln(1 - \text{SEV}(t)) = -\left(\frac{t}{\text{scale}}\right)^{\text{shape}}$$

$$\text{SEV}^{-1}(p) = \text{scale} \cdot (-\ln(1 - p))^{\frac{1}{\text{shape}}} \quad (\text{funzione inversa})$$

Stima dei parametri della Weibull: applicazione al caso delle valutazioni qualitative:

Ora, a partire dalle espressioni dell'impatto medio e della funzione inversa (qui riportate per comodità):

$$sev = scale \cdot \Gamma\left(1 + \frac{1}{shape}\right)$$

$$SEV^{-1}(p) = scale \cdot (-\ln(1-p))^{\frac{1}{shape}}$$

Nell'ipotesi che le stime soggettive raccolte per impatto medio e impatto peggiore assumano rispettivamente il significato di valore atteso dell'impatto e di quantile al 99,9% della distribuzione di impatto:

- stima soggettiva sull'impatto medio = valore atteso dell'impatto del singolo evento
- stima soggettiva sul impatto peggiore = quantile al 99,9% della distribuzione di probabilità dell'impatto,

è possibile impostare un sistema di equazioni che consente di ottenere l'espressione dei parametri incogniti scale e shape della Weibull, in funzione delle stime note raccolte "sev" e "wcase":

$$scale = \frac{sev}{\Gamma\left(1 + \frac{1}{shape}\right)}$$

$$wcase = scale \cdot (-\ln(1-.999))^{\frac{1}{shape}} = scale \cdot (3\ln 10)^{\frac{1}{shape}}$$

In particolare si osserva che la seguente grandezza:

$$\frac{wcase}{sev} = \frac{(3\ln 10)^{\frac{1}{shape}}}{\Gamma\left(1 + \frac{1}{shape}\right)}$$

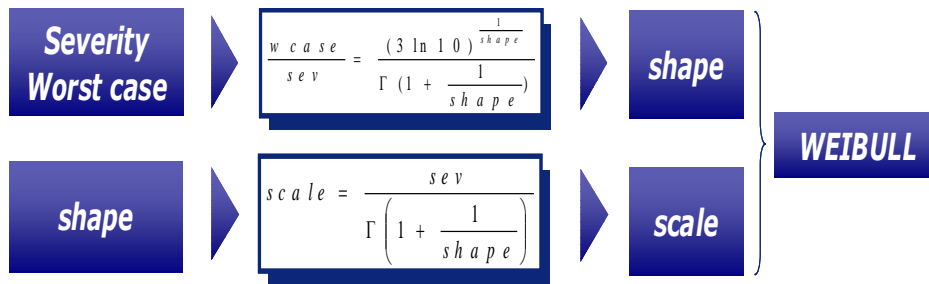
non dipende dal parametro scale.

Ciò equivale a dire che, in corrispondenza di spostamenti nel piano <wcase,sev> lungo rette uscenti dall'origine, il parametro shape rimane costante poiché costante è il rapporto fra impatto medio e impatto peggiore.

Esempio

Se il punto A(wcase,sev) corrisponde ad una Weibull di parametri (scale,shape), il punto B(k*wcase,k*sev) appartenente ad una retta di coefficiente angolare pari a "k" e passante per il punto A, corrisponderà ad una Weibull di parametri (k*scale,shape).

Alla luce di tali osservazioni i passi da compiere per giungere alla determinazione della Weibull partendo dalle stime raccolte possono essere così schematizzati:



5.2.2.2 Lognormale

Nell'ipotesi che l'impatto del singolo evento ("sev(t)") si distribuisca come una Lognormale:

$$f(t) = sev(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\ln t - \mu}{\sigma}\right)^2\right\} \quad (\text{densità di probabilità})$$

$$F(t) = SEV(t) = G\left(\frac{\ln(t) - \mu}{\sigma}\right) \quad (\text{funzione di ripartizione})$$

in cui G è la funzione di ripartizione della normale standard.

In particolare il momento primo (media) della distribuzione può essere così espresso:

$$E[t] = sev = \exp\left\{\mu + \frac{\sigma^2}{2}\right\} \quad (\text{impatto medio})$$

A partire dalla funzione di ripartizione, sopra riportata, si ottiene la funzione inversa (espressione dei quantili in funzione del livello di probabilità prescelto) con i seguenti passaggi:

$$G^{-1}(p) = \frac{\ln(SEV^{-1}(p)) - \mu}{\sigma}$$

$$SEV^{-1}(p) = \exp\{\mu + \sigma G^{-1}(p)\} \quad (\text{inversa})$$

$$E[t^k] = \exp\left(k\mu + \frac{1}{2}k^2\sigma^2\right) \quad (\text{momenti})$$

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} = (2 + \exp(\sigma^2))\sqrt{\exp(\sigma^2) - 1} \quad (\text{skewness})$$

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 = \exp(\sigma^2)^4 + 2\exp(\sigma^2)^3 + 3\exp(\sigma^2)^2 - 3 \quad (\text{kurtosis})$$

Stima dei parametri della Lognormale: applicazione al caso delle autovalutazioni di scenario.

A partire dalle espressioni dell'impatto medio e della funzione inversa (qui riportate per comodità):

$$sev = \exp\left\{\mu + \frac{\sigma^2}{2}\right\}$$

$$SEV^{-1}(p) = \exp\{\mu + \sigma G^{-1}(p)\}$$

nell'ipotesi che le stime soggettive raccolte per impatto medio e impatto peggiore assumano rispettivamente il significato di valore atteso dell'impatto e di quantile al 99,9% della distribuzione di impatto:

- stima soggettiva sul impatto medio = valore atteso dell'impatto del singolo evento
- stima soggettiva sul impatto peggiore = quantile al 99,9% della distribuzione di probabilità dell'impatto,

è possibile impostare un sistema di equazioni che consente di ottenere l'espressione dei parametri incogniti μ e σ della Lognormale, in funzione delle stime note raccolte "sev" e "wcase":

$$\mu = \ln sev - \frac{\sigma^2}{2}$$

$$\sigma = G^{-1}(.999) \pm \sqrt{G^{-1}(.999)^2 - 2 \ln \frac{wcase}{sev}} = 3.090 \pm \sqrt{3.090^2 - 2 \ln \frac{wcase}{sev}}.$$

Delle due soluzioni possibili per σ , viene utilizzata quella minore di 1. Infatti essa:

- assume una forma più coerente con la grandezza rappresentata (impatto di perdita);
- risulta essere più stabile dal punto di vista computazione.

Inoltre si osserva che la seguente grandezza:

$$\ln\left(\frac{wcase}{sev}\right) = \frac{\sigma^2}{2} - 3.090 \cdot \sigma$$

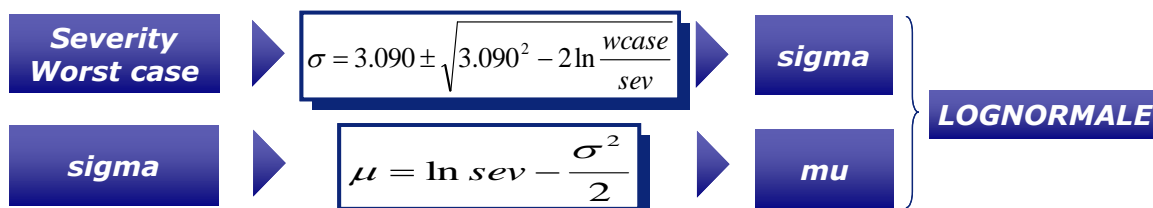
non dipende dal parametro μ .

Ciò equivale a dire che, in corrispondenza di spostamenti nel piano $\langle wcase, sev \rangle$ lungo rette uscenti dall'origine, il parametro σ rimane costante poiché costante è il rapporto fra impatto medio e impatto peggiore.

Esempio

Se il punto $A(wcase, sev)$ corrisponde ad una Lognormale di parametri (μ, σ) , il punto $B(k \cdot wcase, k \cdot sev)$ appartenente ad una retta di coefficiente angolare pari a "k" e passante per il punto A, corrisponderà ad una Lognormale di parametri $(\mu + \ln(k), \sigma)$.

Alla luce di tali osservazioni i passi da compiere per giungere alla determinazione della weibull partendo dalle stime raccolte possono essere così schematizzati:



Stima dei parametri della Lognormale: applicazione al caso delle autovalutazioni di scenario – nuova metodologia.

Ipotizziamo che le stime soggettive raccolte per impatto tipico e impatto peggiore assumano rispettivamente il significato di moda, o valore più frequente di impatto e di quantile associato ad un percentile variabile dal 90 al 99,9% della distribuzione di impatto.

Si può anche in questo caso impostare un sistema di equazioni che consente di ottenere l'espressione dei parametri incogniti μ e σ della Lognormale, in funzione delle stime note raccolte impatto tipico e caso peggiore:

$$\begin{cases} \sigma = \frac{1}{2} \left\{ -G^{-1}(P_{wc}) + \sqrt{[G^{-1}(P_{wc})]^2 + 4 \ln \frac{x_{wc}}{x_{\max}}} \right\} \\ \mu = \sigma^2 + \ln x_{\max} \end{cases}$$

5.2.2.3 Gamma

La distribuzione Gamma è una distribuzione di probabilità definita sui numeri reali non negativi, $[0, \infty)$.

La sua funzione di densità di probabilità è:

$$f(x) = \frac{(x/\theta)^\alpha e^{-x/\theta}}{x\Gamma(\alpha)}$$

dove $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$, $\alpha > 0$ è la funzione Gamma, e con $\alpha > 0$, $\theta > 0$.

$$F(x) = \Gamma(\alpha; x/\theta) \quad (\text{funzione di ripartizione})$$

$$E[X^k] = \frac{\theta^k \Gamma(\alpha+k)}{\Gamma(\alpha)}, k > -\alpha \quad (\text{funzione generatrice dei momenti})$$

$$E[X] = \alpha\theta \quad (\text{media})$$

$$\text{Moda} = \theta, \quad \alpha > 1 \text{ altrimenti vale } 0$$

$$E[X] = \alpha\theta^2 \quad (\text{deviazione standard})$$

Definendo i momenti campionari come:

$$m = \frac{1}{n} \sum_{i=1}^n x_i, \quad t = \frac{1}{n} \sum_{i=1}^n x_i^2$$

Si ottengono i valori degli stimatori (metodo dei momenti):

$$\hat{\alpha} = \frac{m^2}{t - m^2}, \quad \hat{\theta} = \frac{t - m^2}{m}$$

Altrimenti può essere usato il metodo classico della massima verosimiglianza.

5.3 Convoluzione di distribuzioni

La *convoluzione* è una moltiplicazione particolare fra le due distribuzioni in modo tale da tener presente le probabilità relative dei valori che vengono moltiplicati. Poiché la frequency rappresenta il numero di volte in cui accadono le perdite e la severity l'importo della singola perdita, le perdite totali sono date dalla moltiplicazione delle due, tenendo presente quale sia la probabilità di avere un certo numero di perdite e quale sia la probabilità che la singola perdita assuma un certo valore.

Per comprendere meglio cosa sia la convoluzione, prima di guardare alla descrizione matematica, si pensi ad un processo di perdita in cui l'entità delle singole perdite sia sempre uguale (supponiamo X) e si cerchi la distribuzione delle perdite totali in un certo lasso di tempo. Tale distribuzione dipenderà unicamente da quante perdite sono avvenute: date N perdite, l'importo perso è $N \cdot X$. L'importo totale dipende dalla probabilità che è associata al fatto che avvenga un certo numero N di perdite. Se però anche l'importo delle singole perdite è definito da una pdf, per calcolare la probabilità di avere un certo importo totale si deve tener conto di quale sia la probabilità di avere delle perdite X_i ($i=1, \dots, N$) la cui somma sia proprio $N \cdot X$.

In termini matematici, sia f la distribuzione di frequency e s quella di severity, e sia $L_{f=N}$ la somma di N perdite avvenute nell'anno (s_1, \dots, s_N). N rappresenta il numero di accadimenti nell'anno (è perciò un'estrazione dalla distribuzione f di frequency) e gli s_i sono elementi della distribuzione di severity. La distribuzione di probabilità che L sia minore di un certo valore x prefissato è data dalla combinazione di tutti i possibili casi in cui $L_{f=N}$ è minore di x , al variare delle estrazioni da f :

$$P\{L \leq x\} = \sum_{n=0}^{\infty} P\{L_{f=N} \leq x, f = N\} = \sum_{n=0}^{\infty} P\{f = n\} \cdot s^{\otimes n}(x)$$

dove $s^{\otimes n}$ rappresenta la distribuzione di probabilità della somma di n estrazioni di severity:

$$s^{\otimes n}(x) = P\{s_1 + \dots + s_n \leq x\}.$$

5.3.1 Convoluzione via simulazione Monte Carlo

La convoluzione avviene per simulazione Monte Carlo con tante estrazioni sull'impatto quanti sono gli accadimenti previsti dalla frequenza in una precedente estrazione.

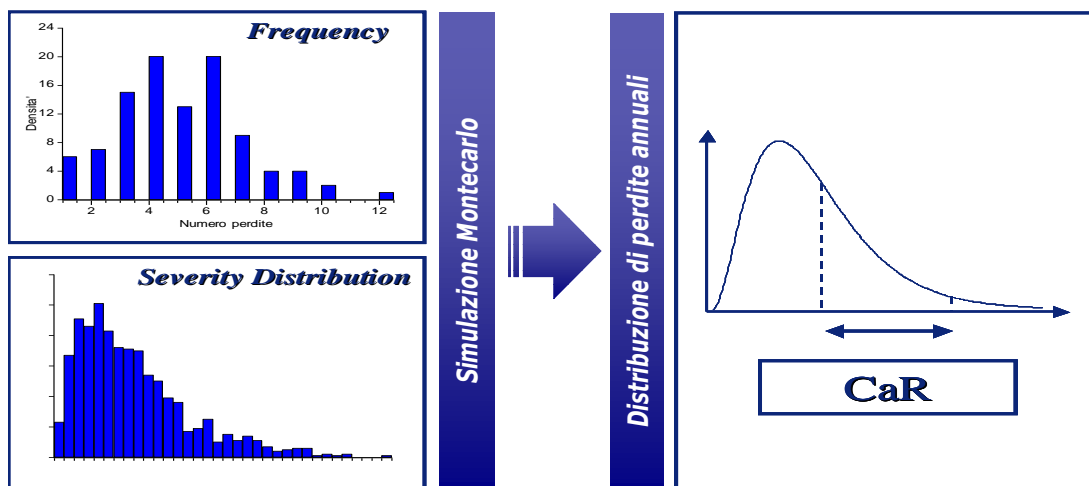
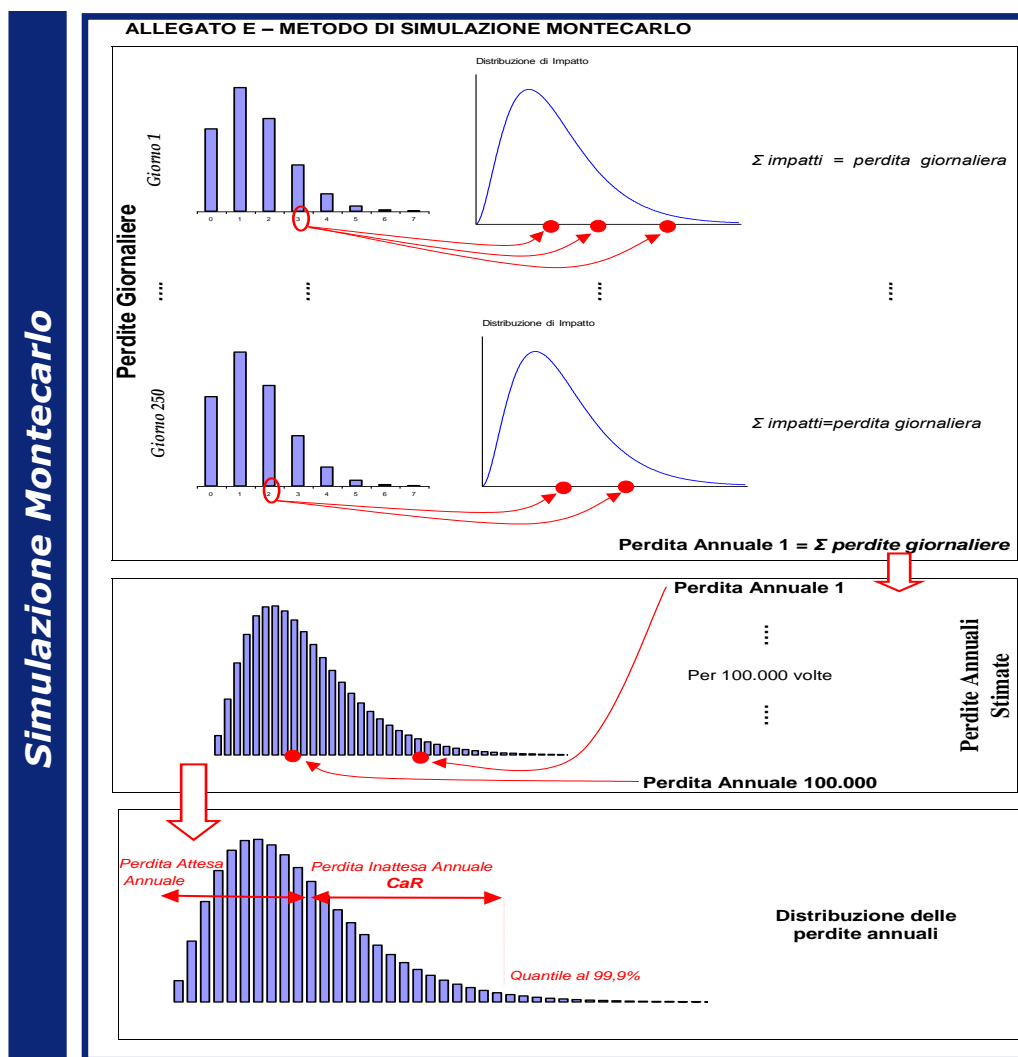


Figura 32 Schema di convoluzione

Nel caso specifico l'estrazione dalle distribuzioni di frequenza e impatto, prevista dalla simulazione MC, è stata attualmente implementata a partire da un generatore di numeri casuali uniforme [0, 1], con successiva applicazione del metodo della ripartizione inversa al fine di determinare le effettive estrazioni dalle distribuzioni d'interesse.



5.4 Sulle funzioni copula e la dipendenza di coda

Le funzioni copula sono utilizzate per descrivere la struttura di dipendenza tra diverse variabili casuali. Il principale vantaggio di utilizzare una funzione copula è quello di poter modellare la struttura di dipendenza tra variabili casuali senza dover fare assunzioni sulle distribuzioni marginali. Nel caso di una copula gaussiana, scegliendo distribuzioni marginali normali, si ricade nel caso specifico di approccio *varianza-covarianza*. Per catturare in maniera adeguata la dipendenza tra variabili a quantili alti, nota come dipendenza di coda o *tail dependence*, l'approccio prevalente in letteratura si basa sull'utilizzo della copula *t-Student* con n gradi di libertà.

L'uso della copula *t-Student* per l'integrazione dei diversi *event type* basa la propria fondatezza sulla capacità di cogliere non solo la correlazione lineare ma anche la correlazione di coda, in quanto il modello stima la misura di *VaR* tenendo in considerazione non solo i fenomeni lineari ma anche i fenomeni correlativi a quantili elevati.

Il manifestarsi di eventi estremi, infatti, può modificare la natura delle correlazioni a quantili elevati e condurre ad una diversa calibrazione della matrice di correlazione (input della funzione copula) in modo che la funzione copula tenga adeguatamente conto delle correlazioni di coda verificatesi.

In generale si dimostra che la copula *t-Student*, diversamente dalla copula normale o gaussiana, consente di modellare una dipendenza di coda, in modo inversamente proporzionale al numero di gradi di libertà⁴⁶.

5.4.1 Definizione di funzione copula

Una copula d -dimensionale è una funzione di densità cumulata C con marginali uniformemente distribuite nell'intervallo $[0,1]$, ovvero $U(0,1)$, tale che:

- 1) $F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$;
- 2) C ha distribuzioni marginali C_i che soddisfano la relazione $C_i(u) = C(1, \dots, 1, u, 1, \dots, 1)$ per ogni $u \in [0,1]$.

⁴⁶ P. Embrechts, F. Lindskog e A. McNeil, *Modelling Dependence with Copulas and Applications to Risk Management*, 2001

Dalla definizione data, si evince che se F_1, \dots, F_d sono funzioni di densità cumulata univariate, allora $C(F_1(x_1), \dots, F_d(x_d))$ è una funzione di densità cumulata con marginali F_1, \dots, F_d , in quanto la generica componente $u_i = F_i(x_i)$ è una variabile casuale $U(0,1)$. In questo modo, le funzioni copula possono essere utilizzate come utili strumenti per la costruzione e la simulazione di distribuzioni multivariate.

Il teorema più importante sulle funzioni di copula è noto come teorema di Sklar (1959).

Se F è una f.d.c. d-dimensionale con marginali continue F_1, \dots, F_d , allora ha una rappresentazione unica attraverso la funzione di copula:

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$$

Il teorema di Sklar garantisce che, in presenza di distribuzioni marginali di perdita continue, l'informazione contenuta nella distribuzione congiunta è esattamente uguale all'informazione contenuta nella copula che descrive la struttura di dipendenza.

5.4.2 Dipendenza di coda (tail dependance)

Il concetto di dipendenza di coda⁴⁷ è rilevante per lo studio della dipendenza tra valori estremi e le funzioni di copula sono adeguate a catturarlo.

Dato un vettore di variabili casuali continue (X_i, X_j) con f.d.c. marginali F_i e F_j , la dipendenza di coda esprime la probabilità del verificarsi di un valore estremo elevato di X_i condizionato al verificarsi di un valore estremo di X_j .

Il coefficiente di dipendenza di coda superiore λ_{ij} di (X_i, X_j) è definito come:

$$\lambda_{ij}^{upper} = \lim_{u \rightarrow 1} P\{X_j > F_j^{-1}(u) | X_i > F_i^{-1}(u)\}$$

Le variabili casuali (X_i, X_j) sono asintoticamente dipendenti nella coda superiore se $\lambda_{ij} \neq 0$. È possibile dare definizioni analoghe per la coda inferiore.

⁴⁷ P. Embrechts, F. Lindskog e A. McNeil, *Modelling Dependence with Copulas and Applications to Risk Management*, 2001

5.4.3 Proprietà fondamentali delle copule Gaussiane e t-Student

Fissate le marginali, il problema dell'aggregazione dei rischi si traduce nella misurazione della struttura di dipendenza incorporata nella copula C , che unisce le marginali F_1, \dots, F_d con la distribuzione congiunta F , secondo la seguente formula:

$$C(u_1, \dots, u_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d))$$

Poiché la distribuzione congiunta F può, in linea di principio, essere determinata sulla base dei dati di perdita congiunti, il problema dell'esistenza di C non sussiste; l'unico problema consiste nell'individuare una copula opportuna.

La copula con distribuzione normale multivariata è chiamata copula gaussiana. Infatti, un vettore casuale è normale multivariato se e solo se:

- 1) le distribuzioni marginali univariate F_1, \dots, F_d sono normali;
- 2) la struttura di dipendenze tra tali marginali è descritta da un'unica funzione di copula C , tale che:

$$C_{\rho}^{Ga}(u_1, \dots, u_d) = \Phi_{\rho}(\phi^{-1}(u_1), \dots, \phi^{-1}(u_d)),$$

dove Φ_{ρ} è la funzione di densità cumulata normale standard multivariata con matrice di correlazione lineare $\{\rho\}$ e ϕ^{-1} è l'inversa della funzione di densità cumulata normale standard univariata.

È possibile dimostrare⁴⁸ che la copula gaussiana bivariata non è dotata di dipendenza di coda se il coefficiente di correlazione lineare ρ_{12} è inferiore ad 1.

La copula t-Student è la copula della distribuzione t-Student multivariata.

La copula t-Student con n gradi di libertà può essere rappresentata analiticamente nel modo seguente:

$$C_{n,\rho}^t(u_1, \dots, u_d) = t_{n,\rho}^d(t_n^{-1}(u_1), \dots, t_n^{-1}(u_d))$$

dove

- $\rho_{ij} = \sigma_{ij} / \sqrt{\sigma_{ii}\sigma_{jj}}$, dove σ_{ij} è la covarianza tra le variabili i e j ;

⁴⁸ L'esempio 3.4 in Embrechts, Lindskog e McNeil, 2001.

- $t_{n,\rho}^d$ è la f.d.c multivariata t-Student, con vettore medio nullo, matrice di covarianza $\{\rho\}$ e gradi di libertà pari ad n;
- t_n^{-1} è l'inversa della f.d.c. marginale t-Student univariata con n gradi di libertà.

Si osserva che per definizione di copula t-Student, la matrice di covarianza Σ non è definita per gradi di libertà inferiori o uguali a 2.

Si dimostra⁴⁹ che, diversamente dalla copula gaussiana, la copula t-Student bivariata manifesta una dipendenza di coda superiore. Tale dipendenza è direttamente proporzionale a ρ_{12} ed inversamente proporzionale rispetto al numero di gradi di libertà n.

La relazione che lega la cosiddetta tail dependence (dipendenza di coda) alla correlazione lineare ed al numero di gradi di libertà è stata dimostrata nel 2001 da Embrechts, Lindskog e McNeil:

$$\lambda_{ij}^{implicita} = 2t_{n+1} \left(-\sqrt{(n+1) \cdot \left(\frac{1-\rho_{ij}}{1+\rho_{ij}} \right)} \right)$$

che viene chiamata per semplicità formula di Embrechts, dove:

- $\lambda_{ij}^{implicita}$ è il coefficiente di dipendenza di coda modellata dalla funzione copula t-Student;
- n è il numero di gradi di libertà;
- t_n è la funzione di distribuzione t-Student;
- ρ_{ij} è il coefficiente di correlazione lineare tra la variabile i e la variabile j.

In particolare, il coefficiente di dipendenza di coda $\lambda_{ij}^{implicita}$ della formula di Embrechts è noto come dipendenza di coda implicita nella funzione copula t-Student. Data questa relazione, fissando i gradi di libertà, conoscendo la dipendenza di coda è possibile determinare la correlazione lineare e viceversa.

⁴⁹ P. Embrechts, F. Lindskog e A. McNeil, *Modelling Dependence with Copulas and Applications to Risk Management*, 2001

5.5 Metodi di stima dei parametri della GPD e proprietà asintotiche

5.5.1 Metodo della massima verosimiglianza (MLE)

Dato un campione $X = \{x_1, x_2, \dots, x_n\}$ di realizzazioni indipendenti e identicamente distribuite (IID), la verosimiglianza di X è il prodotto delle probabilità dei suoi punti:

$$L(X) = \prod_{i=1}^n P(x_i).$$

Poiché gli x_i sono indipendenti,

$$\prod_{i=1}^n P(x_i) = P\left(\bigcap_{i=1}^n x_i\right),$$

ovvero la verosimiglianza quantifica la probabilità che tutte le realizzazioni avvengano “contemporaneamente” per costituire il campione X .

Poiché gli x_i sono identicamente distribuiti, la densità di probabilità f è la stessa per tutti gli elementi della produttoria e dipende sempre dallo stesso parametro (o vettore di parametri) θ .

$$L_\theta(X) = \prod_{i=1}^n f(x_i | \theta).$$

Il metodo di massima verosimiglianza cerca il valore del parametro θ che rende massima la probabilità che il campione X si comporti secondo f .

Nel caso la funzione di probabilità sia una GPD, la funzione di verosimiglianza per le perdite X_1, \dots, X_K è:

$$l_\theta(X) = \ln L_\theta(X) = \ln \left(\prod_k \text{gpd}(X_k) \right) = \ln \left(\prod_k \left(1 - \frac{\xi}{\sigma} (X_k - L) \right)^{-1 - \frac{1}{\xi}} \right)$$

Lo stimatore MLE è definito come valore di ξ che rende massima la funzione di verosimiglianza. Si verifica facilmente che esso soddisfa l'equazione:

$$\sum \ln \left(1 + \frac{\hat{\xi}}{\sigma} (X_k - L) \right) = (1 + \hat{\xi}) \sum_{k=1}^K \left[\frac{(X_k - L)}{\sigma + \hat{\xi}(X_k - L)} \right]$$

la quale può essere risolta numericamente, utilizzando i dati delle perdite storiche X .

5.5.2 Metodo dei momenti pesati in probabilità (PWM)

Il metodo dei Momenti Pesati in Probabilità (Probability Weighted-Moments o PWM) è basato sull'idea che gli stimatori dei parametri incogniti di una distribuzione possono essere derivati dall'espressione dei momenti della popolazione.

Data una variabile casuale X con funzione di distribuzione F i momenti pesati in probabilità sono definiti come

$$M_{p,r,s} = E\{X^p [F(X)]^r [1-F(X)]^s\} \quad p, r, s \in \mathbb{R}$$

Il metodo PWM si basa sull'eguaglianza dei momenti pesati in probabilità della distribuzione con i momenti pesati in probabilità del campione.

Supponiamo di avere un campione X_1, \dots, X_N di variabili random IID distribuite con una GPD. In questo caso è conveniente considerare i momenti pesati in probabilità con $p=1$, $r=0$ e $s=0,1,2$, che, da un semplice calcolo risultano essere

$$M_{0,1,s} = \frac{\sigma}{(s+1)(s+1-\xi)}, \quad \xi < 1$$

Sostituendo $M_{0,1,s}$ con le corrispondenti controparti empiriche troviamo:

$$\hat{M}_{0,1,s} = \frac{1}{N} \sum_{j=1}^N \left(\prod_{l=1}^s \frac{N-j-l+1}{N-l} \right) X_j$$

Eguagliando $M_{0,1,s}$ e $\hat{M}_{0,1,s}$ per $s=1,2$ e risolvendo le due equazioni per σ e τ otteniamo gli stimatori PWM

$$\begin{aligned} \hat{\xi}_{PWM} &= 2 - \frac{\hat{M}_{1,0,0}}{\hat{M}_{1,0,0} - 2\hat{M}_{1,0,1}} \\ \hat{\sigma}_{PWM} &= \frac{2\hat{M}_{1,0,0}\hat{M}_{1,0,1}}{\hat{M}_{1,0,0} - 2\hat{M}_{1,0,1}} \end{aligned}$$

5.5.3 Proprietà asintotiche degli stimatori MLE, PWM ed MPLE

In generale, sia il metodo ML, sia il metodo PWM forniscono degli stimatori per entrambi i parametri ξ e σ della GPD. Detti $\hat{\xi}$ e $\hat{\sigma}$ questi stimatori, si dimostra che:

- i. Nel limite $N \rightarrow \infty$ gli stimatori MLE $(\hat{\xi}_{ML}, \hat{\sigma}_{ML})$ sono asintoticamente normali per $\xi > -1/2$:

$$\sqrt{N}(\hat{\xi}_{ML} - \xi, \hat{\sigma}_{ML} - \sigma) \rightarrow N(0, V_{ML})$$

dove V_{ML} è la matrice di covarianza della distribuzione

$$V_{ML} = \begin{pmatrix} Var(\hat{\xi}_{ML} - \xi) & Cov(\hat{\xi}_{ML} - \xi, \hat{\sigma}_{ML} - \sigma) \\ Cov(\hat{\sigma}_{ML} - \sigma, \hat{\xi}_{ML} - \xi) & Var(\hat{\sigma}_{ML} - \sigma) \end{pmatrix} = (1 + \xi) \begin{pmatrix} 1 + \xi & -\sigma \\ -\sigma & 2\sigma^2 \end{pmatrix}$$

Mentre per $-1 > \xi > -1/2$ gli stimatori MLE convergono ad una distribuzione non-normale ad un tasso di consistenza pari a $N^{-\xi}$.

ii. Nel limite $N \rightarrow \infty$ gli stimatori PWM ($\hat{\xi}_{PWM}, \hat{\sigma}_{PWM}$) per $\xi < 1/2$ soddisfano:

$$\sqrt{N}(\hat{\xi}_{PWM} - \xi, \hat{\sigma}_{PWM} - \sigma) \rightarrow N(0, V_{PWM})$$

Dove la matrice di covarianza della distribuzione è data da:

$$V_{PWM} = \frac{1}{(1-2\xi)(3-2\xi)} \begin{pmatrix} (1-\xi)(2-\xi)^2(1-\xi+2\xi^2) & -\sigma(2-\xi)(2-6\xi+7\xi^2-2\xi^3) \\ -\sigma(2-\xi)(2-6\xi+7\xi^2-2\xi^3) & \sigma^2(7-18\xi+11\xi^2-2\xi^3) \end{pmatrix}$$

Per quanto riguarda il metodo MPLE, Coles e Dixon osservano che lo stimatore ($\hat{\xi}_{MPL}, \hat{\sigma}_{MPL}$) modifica la funzione di verosimiglianza di una quantità asintoticamente trascurabile per $\xi < 1$, pertanto tutti i risultati asintotici standard che valgono per il metodo MLE valgono anche per il metodo MPLE. In particolare, vale l'asintoticità normale dello stimatore MPLE per $\xi \geq 1/2$, con la medesima formula per la matrice di covarianza che nel caso MLE.

5.5.4 Proprietà asintotiche MDPDE

La matrice di varianza asintotica MDPDE è ottenibile come:

$$V_{MDPDE} = J_{\alpha}^{-1}(\xi_0, \beta_0) K_{\alpha}(\xi_0, \beta_0) J_{\alpha}^{-1}(\xi_0, \beta_0).$$

Le matrici K e J sono definite come:

$$J_{\alpha}(\xi, \beta) = \begin{bmatrix} h_{\xi\xi}(\xi, \beta, \alpha) & h_{\xi\beta}(\xi, \beta, \alpha) \\ h_{\xi\beta}(\xi, \beta, \alpha) & h_{\beta\beta}(\xi, \beta, \alpha) \end{bmatrix}$$

$$K_{\alpha}(\xi, \beta) = \begin{bmatrix} h_{\xi\xi}(\xi, \beta, 2\alpha) & h_{\xi\beta}(\xi, \beta, 2\alpha) \\ h_{\xi\beta}(\xi, \beta, 2\alpha) & h_{\beta\beta}(\xi, \beta, 2\alpha) \end{bmatrix} - \begin{bmatrix} h_{\xi}^2(\xi, \beta, \alpha) & h_{\xi}(\xi, \beta, \alpha) h_{\beta}(\xi, \beta, \alpha) \\ h_{\xi}(\xi, \beta, \alpha) h_{\beta}(\xi, \beta, \alpha) & h_{\beta}^2(\xi, \beta, \alpha) \end{bmatrix}$$

Si riportano i passaggi per determinare analiticamente ed in forma chiusa i valori delle componenti delle matrici.

Sia X una v.a. che segue una distribuzione GPD(ξ, β), per le proprietà della GPD valgono le seguenti uguaglianze:

$$E \left(1 - \xi \frac{X}{\beta} \right)^{-r} = \frac{1}{1 - r\xi} , \quad r\xi < 1$$

$$E \left\{ -\log \left(1 - \xi \frac{X}{\beta} \right) \right\}^s = \xi^s \Gamma(s+1) , \quad s \text{ intero}$$

$$E \left(1 - \xi \frac{X}{\beta} \right)^{-r} \left\{ -\log \left(1 - \xi \frac{X}{\beta} \right) \right\}^s = \frac{\xi^s \Gamma(s+1)}{(1 - r\xi)^{s+1}} , \quad r\xi < 1/2$$

La terza relazione può essere usata per calcolare le matrici K e J da utilizzare per la determinazione della matrice di varianza e covarianza asintotica dello stimatore.

La verifica dell'esistenza della terza relazione segue dalla disuguaglianza di Cauchy-Schwartz

$$E \left(1 - \xi \frac{X}{\beta} \right)^{-r} \left\{ -\log \left(1 - \xi \frac{X}{\beta} \right) \right\}^s \leq \left\{ E \left[\left(1 - \xi \frac{X}{\beta} \right)^{-2r} \right] \right\}^{1/2} * \left\{ E \left[\left\{ -\log \left(1 - \xi \frac{X}{\beta} \right) \right\}^{2s} \right] \right\}^{1/2}$$

Sia γ un reale positivo e si ponga $r = -(\xi^{-1} - 1)\gamma$ si ottengono le seguenti relazioni:

$$E \left\{ \left(1 - \xi \frac{X}{\beta} \right)^{(\xi^{-1}-1)\gamma} \right\} = \frac{1}{1 + \gamma - \gamma\xi} , \quad r\xi < 1$$

$$E \left\{ \left(1 - \xi \frac{X}{\beta} \right)^{(\xi^{-1}-1)\gamma-1} \right\} = \frac{1}{1 + \gamma - \gamma\xi - \xi} , \quad \xi(r+1) < 1$$

$$E \left\{ \left(1 - \xi \frac{X}{\beta} \right)^{(\xi^{-1}-1)\gamma-2} \right\} = \frac{1}{1 + \gamma - \gamma\xi - 2\xi} , \quad \xi(r+2) < 1$$

Se vale la $\xi(r+2) < 1$ tutti e tre i momenti esistono finiti; allo stesso modo valgono le uguaglianze:

$$E \left\{ \left(1 - \xi \frac{X}{\beta} \right)^{(\xi^{-1}-1)\gamma} \log \left(1 - \xi \frac{X}{\beta} \right) \right\} = \frac{\xi}{(1 + \gamma - \gamma\xi)^2} , \quad \xi r < 1/2$$

$$E \left\{ \left(1 - \xi \frac{X}{\beta} \right)^{(\xi^{-1}-1)\gamma-1} \log \left(1 - \xi \frac{X}{\beta} \right) \right\} = \frac{-\xi}{(1 + \gamma - \gamma\xi - \xi)^2} , \quad \xi(r+1) < 1/2$$

$$E \left\{ \left(1 - \xi \frac{X}{\beta} \right)^{(\xi^{-1}-1)\gamma} \log^2 \left(1 - \xi \frac{X}{\beta} \right) \right\} = \frac{2\xi^2}{(1 + \gamma - \gamma\xi)^3} , \quad \xi r < 1/2$$

La condizione $\xi(r+2) < 1$ equivale a $\xi < (1+\gamma)/(2+\gamma)$;

e $\xi(r+1) < 1/2$ equivale a $\xi < (0.5+\gamma)/(1+\gamma)$.

Tali limiti sono soddisfatti contemporaneamente se:

$$\xi < \min \left\{ \frac{1+\gamma}{2+\gamma}, \frac{0.5+\gamma}{1+\gamma} \right\} = \frac{1+\gamma}{2+\gamma} , \quad \gamma > 0$$

Consideriamo ora gli integrali che includono gli elementi della score function e della matrice di informazione di Fisher:

$$h_{\xi}(\xi, \beta, \gamma) = \int_{D(\xi, \beta)} S_{\xi}(x; \xi, \beta) f^{\gamma+1}(x; \xi, \beta) dx$$

$$h_{\beta}(\xi, \beta, \gamma) = \int_{D(\xi, \beta)} S_{\beta}(x; \xi, \beta) f^{\gamma+1}(x; \xi, \beta) dx$$

$$h_{\xi\xi}(\xi, \beta, \gamma) = \int_{D(\xi, \beta)} S_{\xi}^2(x; \xi, \beta) f^{\gamma+1}(x; \xi, \beta) dx$$

$$h_{\beta\beta}(\xi, \beta, \gamma) = \int_{D(\xi, \beta)} S_{\beta}^2(x; \xi, \beta) f^{\gamma+1}(x; \xi, \beta) dx$$

$$h_{\xi\beta}(\xi, \beta, \gamma) = \int_{D(\xi, \beta)} S_{\xi}(x; \xi, \beta) S_{\beta}(x; \xi, \beta) f^{\gamma+1}(x; \xi, \beta) dx$$

$$g_{\xi}(\xi, \beta, \gamma) = \int_{D(\xi, \beta)} i_{\xi}(x; \xi, \beta) f^{\gamma+1}(x; \xi, \beta) dx$$

$$g_{\beta}(\xi, \beta, \gamma) = \int_{D(\xi, \beta)} i_{\beta}(x; \xi, \beta) f^{\gamma+1}(x; \xi, \beta) dx$$

$$g_{\xi\beta}(\xi, \beta, \gamma) = \int_{D(\xi, \beta)} i_{\xi\beta}(x; \xi, \beta) f^{\gamma+1}(x; \xi, \beta) dx$$

Dove ⁵⁰

$$S_{\xi}(x; \xi, \beta) = \frac{\partial}{\partial \xi} \log f(x; \xi, \beta)$$

$$S_{\beta}(x; \xi, \beta) = \frac{\partial}{\partial \beta} \log f(x; \xi, \beta)$$

$$i_{\xi}(x; \xi, \beta) = -\frac{\partial^2}{\partial \xi^2} \log f(x; \xi, \beta)$$

$$i_{\beta}(x; \xi, \beta) = -\frac{\partial^2}{\partial \beta^2} \log f(x; \xi, \beta)$$

⁵⁰ Per dettagli si veda Smith 1987 “Estimating tails of probability distributions “

$$i_{\xi\beta}(x; \xi, \beta) = -\frac{\partial^2}{\partial \xi \partial \beta} \log f(x; \xi, \beta)$$

La soluzione degli integrali, espressi come combinazioni dei valori attesi descritti, porta a definire le quantità:

$$\begin{aligned} h_{\xi}(\xi, \beta, \gamma) &= -\frac{1}{\xi^2 \beta^{\gamma}} \left[-\frac{\xi}{(1 + \gamma - \gamma \xi)^2} \right] + \frac{1}{\xi} \frac{1}{\beta^{\gamma}} \left(\frac{1}{\xi} - 1 \right) \left[\frac{1}{(1 + \gamma - \gamma \xi)} - \frac{1}{(1 + \gamma - \gamma \xi - \xi)} \right] \\ h_{\beta}(\xi, \beta, \gamma) &= -\frac{1}{\xi} \frac{1}{\beta^{1+\gamma}} \left[\frac{1}{(1 + \gamma - \gamma \xi)} - \frac{1 - \xi}{(1 + \gamma - \gamma \xi - \xi)} \right] \\ h_{\xi\xi}(\xi, \beta, \gamma) &= \frac{1}{\xi^4 \beta^{\gamma}} \left[\frac{2\xi^2}{(1 + \gamma - \gamma \xi)^3} \right] \\ &\quad + \frac{1}{\xi^2 \beta^{\gamma}} \left(\frac{1}{\xi} - 1 \right)^2 \left[\frac{1}{(1 + \gamma - \gamma \xi)} - \frac{2}{(1 + \gamma - \gamma \xi - \xi)} + \frac{1}{(1 + \gamma - \gamma \xi - 2\xi)} \right] \\ &\quad + \frac{2}{\xi^3 \beta^{\gamma}} \left(\frac{1}{\xi} - 1 \right) \left[\frac{\xi}{(1 + \gamma - \gamma \xi)^2} - \frac{\xi}{(1 + \gamma - \gamma \xi - \xi)^2} \right] \\ h_{\beta\beta}(\xi, \beta, \gamma) &= \frac{1}{\xi^2 \beta^{2+\gamma}} \left[\frac{1}{(1 + \gamma - \gamma \xi)} \right] - \frac{2}{\xi} \frac{1}{\beta^{2+\gamma}} \left(\frac{1}{\xi} - 1 \right) \left[\frac{1}{(1 + \gamma - \gamma \xi - \xi)} \right] \\ &\quad + \frac{1}{\beta^{2+\gamma}} \left(\frac{1}{\xi} - 1 \right)^2 \left[\frac{1}{(1 + \gamma - \gamma \xi - 2\xi)} \right] \\ h_{\xi\beta}(\xi, \beta, \gamma) &= \frac{1}{\xi^3 \beta^{1+\gamma}} \left[-\frac{\xi}{(1 + \gamma - \gamma \xi)^2} \right] + \frac{1}{\xi^2 \beta^{1+\gamma}} \left(\frac{1}{\xi} - 1 \right) \\ &\quad * \left[\frac{\xi}{(1 + \xi - \gamma \xi - \xi)^2} - \frac{1}{(1 + \gamma - \gamma \xi)} + \frac{1}{(1 + \gamma - \gamma \xi - \xi)} \right] \\ &\quad + \frac{1}{\xi} \frac{1}{\beta^{1+\gamma}} \left(\frac{1}{\xi} - 1 \right)^2 \left[\frac{1}{(1 + \gamma - \gamma \xi - \xi)} - \frac{1}{(1 + \gamma - \gamma \xi - 2\xi)} \right] \end{aligned}$$

Per determinare J_{α} è necessario che sia $\gamma = \alpha$ mentre per determinare K_{α} è necessario che sia $\gamma = 2\alpha$.

5.6 Indice di continuità/discontinuità: un esempio

Mostriamo a titolo esemplificativo la costruzione dell'indicatore di continuità/discontinuità per il comparto del Credito per il 2013.

I dati forniti dalla funzione competente sono trimestrali.

CREDITO	Dati di input							
	2012				2013			
INDICATORI	1Q 2012	2Q2012	3Q2012	4Q2012	1Q 2013	2Q2013	3Q2013	4Q2013
Td Sofferenza (teste)	1.31%	1.37%	1.43%	1.67%	1.75%	1.38%	1.76%	2.23%
Sofferenze/ Crediti (teste)	14.96%	15.10%	15.44%	16.07%	16.33%	16.47%	17.06%	17.71%
Incagli/ Crediti (teste)	6.26%	6.43%	7.00%	7.72%	8.12%	8.58%	9.28%	9.40%
Td Sofferenza (importi)	2.53%	2.54%	2.59%	3.25%	3.48%	2.71%	3.92%	4.59%
Sofferenza/ Crediti (importi)	13.31%	13.86%	14.37%	15.49%	16.41%	17.51%	18.44%	20.54%
Incagli/ Crediti (importi)	6.01%	5.98%	6.81%	7.29%	8.23%	8.70%	9.08%	9.63%
Nuovi flussi su bonis (teste)	14.51%	14.03%	11.24%	10.96%	11.53%	7.32%	11.32%	13.07%
Nuovi flussi su bonis (importi)	5.11%	4.54%	4.68%	4.79%	4.61%	2.59%	2.62%	2.74%
Nuovi flussi su totale crediti (teste)	11.44%	10.99%	8.78%	8.52%	8.74%	5.49%	8.37%	9.55%
Nuovi flussi su totale crediti (importi)	4.00%	3.68%	3.77%	3.79%	3.55%	1.97%	1.94%	2.00%

La variazione vien calcolata per differenza tra un trimestre e lo stesso trimestre dell'anno precedente:

$$z = y(t) - y(t-4)$$

Viene poi calcolata la deviazione standard delle variazioni per la serie storica disponibile.

CREDITO	Variazioni				Deviazione standard serie storica variazioni
	2013-2012				
INDICATORI	1Q	2Q	3Q	4Q	
Td Sofferenza (teste)	0.44%	0.01%	0.33%	0.56%	0.29%
Sofferenze/ Crediti (teste)	1.37%	1.37%	1.62%	1.64%	0.69%
Incagli/ Crediti (teste)	1.86%	2.15%	2.28%	1.68%	0.93%
Td Sofferenza (importi)	0.95%	0.17%	1.32%	1.34%	0.57%
Sofferenza/ Crediti (importi)	3.10%	3.65%	4.07%	5.05%	1.14%
Incagli/ Crediti (Importi)	2.22%	2.72%	2.27%	2.34%	0.87%
Nuovi flussi su bonis (teste)	-2.98%	-6.71%	0.08%	2.11%	3.40%
Nuovi flussi su bonis (importi)	-0.50%	-1.96%	-2.06%	-2.04%	1.93%
Nuovi flussi su totale crediti (teste)	-2.70%	-5.50%	-0.41%	1.03%	2.65%
Nuovi flussi su totale crediti (importi)	-0.45%	-1.72%	-1.83%	-1.79%	1.60%

Per individuare se le variazioni sono significative si confrontano le differenze con la deviazione standard e si attribuiscono i valori 1/0 alle differenze a seconda che superino o meno, in valore assoluto, la deviazione standard delle variazioni.

CREDITO	Assegnazione valori dicotomici			
INDICATORI	1Q 2013	2Q2013	3Q2013	4Q2013
Td Sofferenza (teste)	1	0	1	1
Sofferenze/ Crediti (Teste)	1	1	1	1
Incagli/ Crediti (teste)	1	1	1	1
Td Sofferenza (importi)	1	0	1	1
Sofferenza/ Crediti (importi)	1	1	1	1
Incagli/ Crediti (Importi)	1	1	1	1
Nuovi flussi su bonis (numero)	0	1	0	0
Nuovi flussi su bonis (importi)	0	1	1	1
Nuovi flussi su totale crediti (numero)	1	1	0	0
Nuovi flussi su totale crediti (importi)	0	1	1	1

Per ottenere l'indicatore di continuità/discontinuità su base annuale si sommano le variabili dicotomiche ottenute e si divide per il numero di osservazioni dell'anno, ovvero quattro.

Per ciascun indice si ottiene così un indicatore di continuità/discontinuità, ad esempio per il primo indicatore relativo al tasso di decadimento delle Sofferenze (calcolato sulle teste) l'indicatore assume valore pari a 75%.

CREDITO	Indicatore di continuità
INDICATORI	2013
Td Sofferenza (teste)	75%
Sofferenze/ Crediti (teste)	100%
Incagli/ Crediti (teste)	100%
Td Sofferenza (importi)	75%
Sofferenza/ Crediti (importi)	100%
Incagli/ Crediti (Importi)	100%
Nuovi flussi su bonis (teste)	25%
Nuovi flussi su bonis (importi)	75%
Nuovi flussi su totale crediti (teste)	50%
Nuovi flussi su totale crediti (importi)	75%
Indicatore di comparto	78%

Dalla media di tutti gli indicatori presenti all'interno del comparto si ottiene l'indicatore di comparto. In questo caso l'indicatore del comparto Credito è pari a 78%.

6 Bibliografia

- [1] Banca d'Italia . (2006). *Circ. 263 - titolo II Cap. 5 Sez. IV*.
- [2] Beirlant J., Goetgebeur Y., Segers J., Teugels J.. (2004). *Statistics of Extremes: Theory and Applications*. Chichester: Wiley.
- [3] Chavez-Demoulin, Davison and Mcneil. (2004). *Estimating Value-at-Risk: A point process approach*. Department of Mathematics ETH Zurich.
- [4] Chernobai A., Rachev S., Fabozzi F.. (2005). *Composite Goodness-of-Fit Tests for Left-Truncated Loss Samples*. Department of Statistics and Applied Probability University of California.
- [5] Di Clemente A., Romano C.. (2003). *A Copula-Extreme Value Theory Approach for Modelling Operational Risk*.
- [6] Embrechts P., Kluppelberg C., Mikosch T.. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer.
- [7] Embrechts P., Lindskog F., McNeil A.. (2001). *Modelling Dependence with Copulas and Applications to Risk Management*. Zurich: Department of Mathematics ETHZ.
- [8] Juarez S.. (2003). *Robust and Efficient Estimation for the Generalized Pareto Distribution*. Statistical Science Department, Southern Methodist University. Available at <http://www.smu.edu/statistics/faculty/SergioDiss1.pdf>.
- [9] Juarez S., Schucany W.. (2004). *Robust and Efficient Estimation for the Generalized Pareto Distribution*. *Extremes* 7:237–251
- [10] Juarez S., Schucany W. (2006) *A note on the asymptotic distribution of the minimum density power divergence estimator*.
- [11] Klugman S.A., Panjer H.H., Willmot G.E.. (2004). *Loss Models, From Data to Decisions*. John Wiley & Sons, Inc.
- [12] Luceno A. (2005) “Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators”, *Computational Statistics & Data Analysis* (51) – pp 904-917.
- [13] Manistrea, B. J. (2008). *A Practical Concept of Tail Correlation*.
- [14] Mc Neil, Rudiger, Embrechts. (2005). *Quantitative Risk Management, Concepts, Techniques and Tools*. Princeton University Press.
- [15] Moscadelli, M. (2004). *The modelling of operational risk*. Paper 517 Bank of Italy.
- [16] O'Brien, J. (1980). *A limit theorem for sample maxima and heavy branches in Galton Watson trees*. *Applied Probability* 17.
- [17] Smith, R.L. (1987). *Estimating tails of probability distributions*. *The Annals of Statistics* 15.

- [18] Smith, R.L. S. Coles and M. Dixon. *Likelihood-based inference for extreme value models*. Extremes, 2:5-23, 1999.