



**Dipartimento di
impresa e management**

Cattedra di Statistica

**“I metodi statistici per il *credit scoring*: applicazione del
modello logistico per la valutazione del rischio di credito delle
imprese”**

RELATORE

Prof.ssa Livia De Giovanni

CANDIDATO

Filippo Preziosi

Matricola 221891

Anno Accademico 2019/2020

INDICE

INTRODUZIONE	5
CAPITOLO 1: IL CREDIT SCORING	7
1.1 DEFINIZIONE DI CREDIT SCORING.....	7
1.2 FASI DEL CREDIT SCORING	8
1.2.1 SPECIFICITÀ, SENSIBILITÀ E CURVA ROC	10
1.3 MODELLI DI CREDIT SCORING.....	13
1.3.1 MODELLO Z-SCORE DI ALTMAN.....	13
1.3.2 MODELLO DI ANALISI DISCRIMINANTE	14
1.4 MOTIVAZIONI ALLA BASE DELL'UTILIZZO DEI METODI DI CREDIT SCORING.....	16
1.5 VANTAGGI APPORTATI DALL'UTILIZZO DEL CREDIT SCORING.....	17
1.6 EFFETTI DEL CREDIT SCORING SULLA SELEZIONE DEI MUTUATARI E SULLA PERFORMANCE DEL CREDITO	18
CAPITOLO 2: RISCHIO DI CREDITO E RELATIVA REGOLAMENTAZIONE	21
2.1 DEFINIZIONE DI RISCHIO DI CREDITO	21
2.2 GESTIONE DEL RISCHIO DI CREDITO: ANALISI DELLE COMPONENTI	22
2.2.1 EXPECTED LOSS.....	23
2.2.1.1. PROBABILY OF DEFAULT	24
2.2.1.2 EXPOSURE AT DEFAULT	25
2.2.1.3 LOSS GIVEN DEFAULT.....	26
2.2.2. UNEXPECTED LOSS	27
2.3 BASILEA II.....	28
2.3.1 METODI DI PONDERAZIONE DEL RISCHIO DI CREDITO	29
2.3.1.1 STANDARDISED APPROACH.....	30
2.3.1.2 APPROCCI INTERNAL RATING BASED	32
2.3.1.2.1 INTERNAL RATING BASED FOUNDATION.....	32
2.3.1.2.2 INTERNAL RATING BASED ADVANCED.....	33
2.4 LIMITI DI BASILEA II	33
2.5 BASILEA III	35
CAPITOLO 3: IL MODELLO DI REGRESSIONE LOGISTICA	37
3.1 CONCETTI INTRODUTTIVI PROPEDEUDICI ALLA COMPrensIONE DEL MODELLO	37
3.3 SELEZIONE DEL CAMPIONE	40

3.4 SELEZIONE PRELIMINARE DELLE VARIABILI ESPLICATIVE.....	41
3.5 STIMA DEL MODELLO.....	42
3.5.1 STIMA DEI PARAMETRI.....	44
3.5.2 VALUTAZIONE FIT DEL MODELLO.....	46
3.5.3 CONTRIBUTO DEI SINGOLI PARAMETRI	47
3.6 VERIFICA DELLA CAPACITÀ PREDITTIVA DEL MODELLO	48
CAPITOLO 4: APPLICAZIONE DEL MODELLO DI REGRESSIONE LOGISTICA PER VALUTARE IL	
RISCHIO DI CREDITO DI UN CAMPIONE DI IMPRESE	49
4.1 SELEZIONE DEL CAMPIONE	49
4.2 SELEZIONE DELLE VARIABILI ESPLICATIVE	52
4.3 STIMA DEL MODELLO 1	56
4.3.1 VERIFICA DELLA BONTÀ DI ADATTAMENTO DEL MODELLO	59
4.3.2 VERIFICA DELLA CAPACITÀ PREDITTIVA DEL MODELLO.....	60
4.3.3 VALUTAZIONE DELLA SIGNIFICATIVITÀ DEI SINGOLI REGRESSORI.....	61
4.4 STIMA DEL MODELLO 2	62
CONCLUSIONE	65
BIBLIOGRAFIA	67

INTRODUZIONE

Dopo la crisi dei mutui subprime statunitensi e la conseguente crisi del debito sovrano europeo, sono emerse non poche preoccupazioni in relazione all'utilizzo dei *Credit Default Swap* e più in generale di ogni strumento finanziario in cui non fosse valutato in maniera adeguata il merito creditizio della controparte. L'ondata di instabilità finanziaria, nonostante il tempo trascorso, non ha ancora esaurito la propria portata: la rarefazione del *funding*, i riflessi del deterioramento dell'economia reale sulla qualità del credito e la compressione dei margini operativi, sono solo alcuni degli effetti che ancora si protraggono sul sistema bancario.

Conseguentemente, la gestione del rischio di credito è divenuta ormai un fattore fortemente rilevante che attira un'attenzione significativa da parte degli operatori di mercato. Sviluppare un adeguato modello di credit scoring è dunque uno dei punti focali per le istituzioni finanziarie, al fine di gestire efficacemente le esposizioni creditizie rischiose e ottimizzare i profitti. Esse hanno così sviluppato una serie di diverse tecniche statistiche di apprendimento automatico per costruire modelli di rating del credito.

L'obiettivo dell'elaborato è quello di presentare le caratteristiche generali delle tecniche statistiche utilizzate per il credit scoring, concentrandosi approfonditamente su uno dei modelli più utilizzati per la valutazione creditizia della controparte (prima attraverso una trattazione metodologica e poi tramite un'applicazione pratica dello stesso).

Il lavoro si compone di quattro parti.

Nella prima parte viene fornita una panoramica generale sulle tecniche di credit scoring. Ciò viene fatto attraverso una descrizione più analitica nella prima parte e più concettuale nella seconda, al fine di mostrare sia le motivazioni che i vantaggi alla base dell'utilizzo del credit scoring, ma soprattutto l'effetto dello stesso sulla performance del credito.

Nella seconda parte viene analizzato il concetto di rischio di credito. Dopo un primo focus sulla relativa definizione e sulla gestione dello stesso attraverso l'analisi delle componenti (perdita attesa e inattesa), l'impostazione del capitolo presenta un'evidente correlazione con l'evoluzione della regolamentazione bancaria nel corso degli ultimi anni (da Basilea I a Basilea III).

Nel terzo e nel quarto capitolo si approfondisce la trattazione sul credit scoring attraverso la presentazione di un modello specifico e l'applicazione pratica dello stesso. Nel terzo, di carattere analitico, si presenta infatti il modello di regressione logistica, attraverso la trattazione delle varie

fasi di costruzione del modello (selezione del campione, delle variabili esplicative, stima del modello e verifica della capacità predittiva). Nel quarto, di carattere applicativo, viene invece selezionato un campione di imprese italiane per l'applicazione del modello teorico analizzato precedentemente. La scelta della tipologia di imprese (startup innovative) non è affatto casuale: in Italia la stretta relazione tra banche e imprese riflette infatti alcune caratteristiche strutturali della nostra economia (su tutte, la frammentazione del tessuto produttivo in numerose PMI che non sono in grado di accedere direttamente al mercato dei capitali)¹. L'analisi empirica svolta tramite l'utilizzo del software Gretl mira ad analizzare praticamente il rapporto tra gli indicatori di bilancio scelti come variabili esplicative e la variabile dipendente Y (probabilità di default); inoltre, viene valutata la capacità predittiva del modello e la significatività dei singoli regressori.

¹ Signorini L. F., *Banche e imprese nella crisi*, Banca d'Italia, XLIV Giornata del credito, Ottobre 2012

CAPITOLO 1: IL CREDIT SCORING

1.1 DEFINIZIONE DI CREDIT SCORING

Il credit scoring è l'insieme delle tecniche statistiche utilizzate per la valutazione della solvibilità della controparte di un'operazione finanziaria.

Si tratta di un sistema di calcolo che presenta le caratteristiche di coerenza, velocità e accuratezza e che riesce ad analizzare una serie di informazioni sul potenziale cliente per fornire come risultato un punteggio (*score*) sul rischio di insolvibilità dello stesso. Tale score sarà poi discriminante nella scelta binaria sulla concessione del finanziamento.

Sebbene questo metodo di sottoscrizione sia ben consolidato nei mercati del credito al consumo, per circa l'ultimo decennio è stato ampiamente applicato solo ai crediti commerciali. Le motivazioni del ritardo nella risposta sono diverse: la significativa eterogeneità tra i mutuatari, la variazione importante negli approcci di sottoscrizione sia all'interno delle banche che tra le stesse, e la diversità dimensionale dei vari crediti.

Secondo la Banca d'Italia, le informazioni più rilevanti utilizzate per il *credit scoring* sono di quattro tipi: sul cliente, sul finanziamento da erogare, sul bene da finanziare e sul grado di indebitamento del richiedente credito (informazione registrata nella centrale rischi). Tra gli altri fattori presi in considerazione nell'utilizzo delle tecniche di *credit scoring*, sono presenti la reputazione finanziaria e la storia creditizia del richiedente: i dati registrati nelle centrali rischi *sic* e *crif* sono controllati al fine di verificare eventuali comportamenti che ostacolino la concessione del credito (ad esempio presunta morosità su pagamenti di prestiti rateali). Si inverte dunque la tendenza a concedere un finanziamento in base a valutazioni soggettive basate sul legame cliente-banca, in favore di tecniche quantitative basate sull'elaborazione automatica di informazioni standardizzate, in grado di definire quantitativamente il rischio connesso ad un'operazione di finanziamento.

1.1.1 TIPOLOGIE DI SCORING

Le tecniche di scoring possono essere utilizzate sia nell'ambito della gestione del credito al consumo che per valutare la probabilità di insolvenza delle piccole e medie imprese (oggetto di

analisi nel quarto capitolo): possiamo sostanzialmente affermare che l'impiego è correlato alla necessità più generale di prevedere il rischio che la controparte di un'operazione finanziaria risulti insolvente.

Si distinguono due casi di credit scoring. Se esso è finalizzato alla concessione di un prestito ad un nuovo cliente, si parla di scoring di accettazione (*application scoring*). Se invece lo scoring riguarda un cliente già relazionatosi con l'ente erogatore, e dunque si tratta di vera e propria "gestione" dello stesso, si parla di scoring comportamentale (*behavioural scoring*).

Come precedentemente affermato, l'obiettivo dell'elaborato è quello di applicare il modello logistico al fine di valutare la solvibilità di un campione di imprese, e, conseguentemente, i modelli presentati si muovono nel contesto dello scoring di accettazione. In relazione a ciò, la variabile che stabilisce l'evento da prevedere è necessariamente binaria, al fine di dividere il campione in due insiemi: creditori solvibili e non solvibili ("buoni" e "cattivi"), classificati come tali a seconda che il punteggio di credito del cliente sia rispettivamente maggiore o minore di un valore soglia precedentemente fissato.

1.2 FASI DEL CREDIT SCORING

Prescindendo dalla categorizzazione delle tecniche di scoring riportate, esse si distinguono in base alla modalità di costruzione della griglia di scoring. L'ipotesi principale alla base della suddetta costruzione è l'omogeneità della popolazione oggetto di studio in un arco temporale supposto sufficientemente ampio, in modo che le informazioni riguardanti la storia creditizia del cliente possano costituire la base di dati del sistema di scoring. Specificatamente, le fasi del processo di credit scoring possono essere così delineate²:

1. Definizione della popolazione obiettivo.

La prima fase del credit scoring consiste nel delineare una popolazione definita che, come già accennato, sia temporalmente omogenea rispetto a variabili economiche prefissate (dimensione, fatturato, settore, prodotto offerto).

2. Definizione della variabile di classificazione.

La variabile di classificazione viene anche detta *Flag* ed è la variabile che, nell'intervallo prefissato, definisce il manifestarsi o meno del fenomeno oggetto di studio. La scelta della variabile è puramente discrezionale, ma, nei casi di credit scoring di accettazione, è

² Secondo Stanghellini, introduzione ai metodi statistici per il credit scoring, 2009

solitamente connessa a fenomeni di inadempimento delle condizioni contrattuali prefissate in tema di rimborso del debito.

3. Formazione di una base di dati.

Nella terza fase del processo di credit scoring si definiscono le unità che formeranno il campione oggetto di analisi e sul quale sarà implementata la metodologia (fase 4), e le informazioni determinanti su di esse (che, nel credito alle imprese di modesta dimensione, derivano dal bilancio aziendale). Ovviamente, il campione deve avere caratteristiche rappresentative della popolazione obiettivo definita o deve essere estratto in maniera casuale dalla suddetta popolazione (es. scelta casuale tra unità che hanno presentato richiesta di finanziamento). Infine, è opportuno specificare che l'intervallo temporale in cui le unità sono osservate e studiate, e al termine del quale saranno classificate in uno dei due insiemi della *Flag*, sia mantenuto costante per tutte le osservazioni.

4. Determinazione e implementazione della metodologia.

Per quanto concerne la determinazione della metodologia sulla quale costruire il sistema di credit scoring, essa viene spesso scelta in relazione agli obiettivi che si vogliono perseguire attraverso l'analisi; riguardo all'implementazione, è prassi nota dividere il campione in un campione di sviluppo (*training sample*) ed un campione, meno numeroso, di convalida (*test sample*). La scelta di proseguire in questo modo deriva sostanzialmente dalla volontà di non utilizzare le stesse unità più volte (sia per costruire la metodologia che per convalidarla, appunto).

Al fine di migliorare la precisione della metodologia selezionata, occorre poi operare alcune scelte riguardanti le informazioni che saranno inserite nel modello e la forma funzionale delle stesse. Per tale motivo sono utilizzate tecniche statistiche inferenziali che riescono a evidenziare gli andamenti tipici del fenomeno rispetto a quelli casuali dovuti all'utilizzo di un determinato campione. Infine, stimato il modello, la tecnica di scoring scelta può tradursi in un sistema di parametri da apporre alle varie configurazioni delle variabili descrittive del cliente.

5. Scelta del valore soglia e stima degli errori di classificazione.

La scelta del valore soglia è tipica soprattutto nei modelli di credit scoring di accettazione, che, come detto, è la tipologia di scoring studiata nell'elaborato. In questa fase viene scelto il valore con il quale confrontare i risultati dell'applicazione del modello e dunque decidere l'ammissione al finanziamento della controparte. Non vi sono regole prefissate nella scelta

di tale valore, che viene più che altro fissato in base agli obiettivi aziendali e in relazione al concetto di costo opportunità³. Il valore soglia viene così scelto come il valore che minimizza il costo atteso ed è direttamente proporzionale al costo che l'azienda sopporta in caso di classificazione sbagliata.

Gli errori di classificazione possono dunque riguardare sia la possibilità di catalogare come solvibili unità che non lo sono, sia di classificare come non solvibili clienti solvibili. Per stimare la probabilità di commettere uno di questi errori si possono sia utilizzare altri metodi statistici, sia osservare le frequenze relative di errore nel campione di convalida (anche se in questo modo tendono ad essere sottostimati).

6. Controllo della precisione dello scoring nel tempo.

Tale ultima e conclusiva fase rappresenta la fase di *monitoring* della metodologia e permette di verificarne l'accuratezza nel tempo dopo l'implementazione. Vengono perciò valutate sia la capacità discriminate – separare clienti solvibili da quelli che non lo sono – che la calibrazione – linearità tra la probabilità di insolvenza stimata e quella osservata -.

1.2.1 SPECIFICITÀ, SENSIBILITÀ E CURVA ROC

Per approfondire meglio la tematica dell'errata classificazione, appare opportuno usufruire di due termini utilizzati soprattutto nel linguaggio biomedico: specificità e sensibilità. Con il primo termine si indica la probabilità di presenza di un falso positivo, ovvero del fatto che, nel nostro caso, un creditore non solvibile sia classificato come solvibile. Con il termine sensibilità si indica invece il concetto, opposto, di falso negativo, ovvero la probabilità di classificazione di un creditore potenzialmente solvibile come non solvibile.

Per esprimere tali concetti in termini statistici, si prenda una scelta s casuale del valore soglia, sia $P(Y = 1)$ la probabilità a priori di classificazione di creditori solvibili e $P(Y = 0)$ quella di classificazione dei creditori non solvibili. Data $S = S(X_1, X_2, \dots, X_p)$ la variabile casuale ottenuta come funzione delle v.c. (X_1, X_2, \dots, X_p) e la funzione di ripartizione $F(s) = P(S \leq s)$, il concetto di specificità può essere dunque espresso come $F(s | Y = 0)$ e quello di sensibilità come $1 - F(s | Y = 1)$.

³ Inteso in tal caso come il mancato profitto derivante dalla perdita di un cliente potenzialmente solvibile.

Considerando un'ipotetica analisi di credit scoring a risposta quantale (cioè finalizzata a fornire una risposta binaria: solvibile/ non solvibile), la valutazione della validità della suddetta analisi può essere dedotta dalla tabella di contingenza 2 x 2 seguente:

Situazione del creditore

		Solvibile	Non solvibile
Esito dell'analisi	Ammesso al finanziamento	S+ F+	S- F+
	Non ammesso al finanziamento	S+ F-	S- F-

Nella quale i simboli S+, S-, F+, F-, indicano rispettivamente i creditori solvibili, quelli non solvibili, quelli che hanno avuto accesso al finanziamento e quelli che non lo hanno avuto. In tal modo appare facile dedurre che nella casella S+ F+ sono contenuti i veri positivi (solvibili ammessi al finanziamento), nella casella S- F- i veri negativi (creditori non solvibili non ammessi al finanziamento), nella casella S- F+ i falsi positivi e in quella S+ F- i falsi negativi. La situazione finale è riassunta seguente nella tabella:

Situazione del creditore

		Solvibile	Non solvibile
Esito dell'analisi	Ammesso al finanziamento	Veri positivi	Falsi positivi
	Non ammesso al finanziamento	Falsi negativi	Veri negativi

Dalla tabella si intuisce che le caselle della diagonale principale (veri positivi e veri negativi) rappresentano i casi di analisi corretta, mentre le caselle della diagonale secondaria (falsi positivi e falsi negativi) rappresentano invece i casi in cui lo stesso esame da luogo a risultati sbagliati.

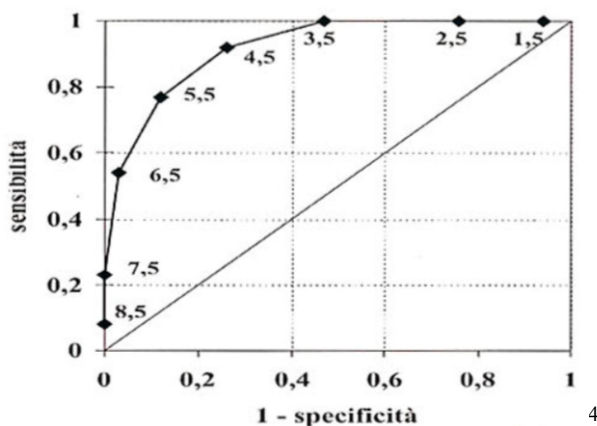
Una tabella così composta può dunque riassumere perfettamente i concetti di sensibilità e di specificità enunciati, definendoli come segue:

$$\text{Sensibilità: } VP/(VP + FN)$$

$$\text{Specificità: } VN/(FP + VN)$$

Principio fondamentale per una griglia di score è quello di presentare, per determinati valori della specificità, alti valori di sensibilità. La curva ROC – *Received Operating Characteristics* – si basa su tale principio ed è un diagramma che riporta, per valori diversi di una soglia s , in ordinata la

sensibilità e in ascissa il tasso dei falsi positivi (ovvero il complemento a 1 della specificità). Un esempio di curva ROC viene presentato in figura.



La scelta di riportare la curva ROC sta nella sua importanza in relazione alla costruzione di una griglia di scoring: per una data soglia s , essa fornisce la percentuale $F(s | Y = 1) * 100$ di clienti solvibili che devono essere esclusi dal finanziamento al fine di non includere la percentuale $F(s | Y = 0) * 100$ di clienti non solvibili.

Dal grafico è possibile notare alcune importanti informazioni ai fini della classificazione. Innanzitutto, la curva ROC è inscritta in un quadrato di area unitaria. La bisettrice rappresenta la curva con eguali valori dei due parametri – $F(s | Y = 1) = F(s | Y = 0)$ – e, essendo uguale per le due popolazioni, rappresenta una regola di classificazione casuale delle due unità. L'area sottesa alla curva ROC può essere interpretata come una probabilità: maggiore è l'area circoscritta tra essa e la bisettrice, tanto migliore è il classificatore.

Un indice sintetico utile ad esprimere in formule quanto detto è l'indice di Gini. Definita la curva ROC come $y(x)$, possiamo scrivere:

$$I_{ROC} = \frac{\int_0^1 y(x)d(x) - 1/2}{1/2}$$

Il numeratore della formula indica l'area sottesa alla curva ROC sopra la bisettrice; $1/2$ è invece la misura dell'area al di sotto della bisettrice stessa. L'indice sarà dunque 0 in caso di classificazione casuale e 1 in caso di perfetta.

⁴ Fonte: La valutazione di un esame diagnostico mediante curva ROC, Marco B. L. Rocchi, Istituto di Biomatemica, Facoltà di Farmacia, Università di Urbino

1.3 MODELLI DI CREDIT SCORING

L'intento di questo paragrafo è di analizzare modelli di credit scoring alternativi al modello di regressione logistica studiato nell'elaborato (capitolo 3). Nei seguenti paragrafi verranno così presentati, in limitazione ad una trattazione sintetica ed a scopo puramente conoscitivo, il modello Z-score di Altman e quello di analisi discriminante lineare.

1.3.1 MODELLO Z-SCORE DI ALTMAN

Appare opportuno presentare il modello sviluppato da Edward Altman poiché esso rappresenta probabilmente uno dei primi modelli di credit scoring della storia e soprattutto quello che per molto tempo è risultato il modello più utilizzato per la valutazione del rischio di credito della controparte.

Lo scopo del lavoro di Altman era quello di creare un modello di analisi multivariata da contrapporre ai classici modelli univariati utilizzati fino a quel momento, suscettibili di interpretazioni errate e generanti confusione nella valutazione creditizia. L'analisi discriminante multipla (MDA) utilizzata da Altman muove dunque dalla volontà di trovare una serie di indicatori della situazione economica finanziaria idonei a predire la probabilità di insolvenza di un'impresa e assegnare loro i relativi pesi all'intero di una funzione discriminante.

Il primo passaggio è la definizione di espliciti gruppi di classificazione. Con il metodo MDA viene derivata una combinazione lineare delle caratteristiche che meglio discriminano le variabili tra i gruppi oggetto di analisi.

Il modello di Altman si presenta in tale versione:

$$Z = 0,12 X_1 + 0,14 X_2 + 0,33 X_3 + 0,06 X_4 + 0,999 X_5$$

Dove:

X_1 = capitale circolante / totale attivo. Tale rapporto è una misura della liquidità netta dell'impresa rispetto alla capitalizzazione totale, ed è l'indice di liquidità solitamente più indicativo della situazione finanziaria aziendale.

X_2 = Utile non distribuito / totale attivo. Secondo Altman tale indice considerava implicitamente l'età di un'impresa, poiché imprese giovani presentano solitamente bassi rapporti RE / TA (*retained earnings / total assets*). Inserendo tale rapporto Altman considera dunque la possibilità che una

giovane impresa sia discriminata rispetto ad una più anziana, sostenendo che l'evidenza empirica sottolinea una maggiore probabilità di fallimento nei primi anni di attività.

$X_3 = \text{Utili ante interessi e imposte (EBIT)} / \text{totale attivo}$. La scelta di tale indicatore è giustificata dalla stretta correlazione dello stesso con la reale produttività del patrimonio dell'impresa. Tale rapporto appare dunque particolarmente adeguato nelle analisi sulla probabilità di insolvenza di un'azienda.

$X_4 = \text{Valore di mercato dell'equity} / \text{Book value del debito totale}$. L'*equity* è misurato dal valore di mercato di tutte le azioni mentre il debito include sia quello a breve che a lungo termine. Tale indicatore mostra quanto le attività aziendali possono diminuire di valore prima che le passività siano maggiori delle attività e le aziende diventino insolventi.

$X_5 = \text{Fatturato} / \text{totale attivo}$. Il rapporto tra capitale e fatturato è un rapporto finanziario standard in grado di illustrare la capacità di generare vendite delle attività dell'azienda. Misura sostanzialmente la capacità del management di gestire la concorrenza.

Il risultato del modello di Altman è uno *score (Z)* di ogni richiedente credito, confrontato con un valore soglia (fissato a 0,181): i prenditori con un punteggio inferiore a tale *cut-off* sono considerati non solvibili o comunque ad elevato rischio insolvenza.

1.3.2 MODELLO DI ANALISI DISCRIMINANTE

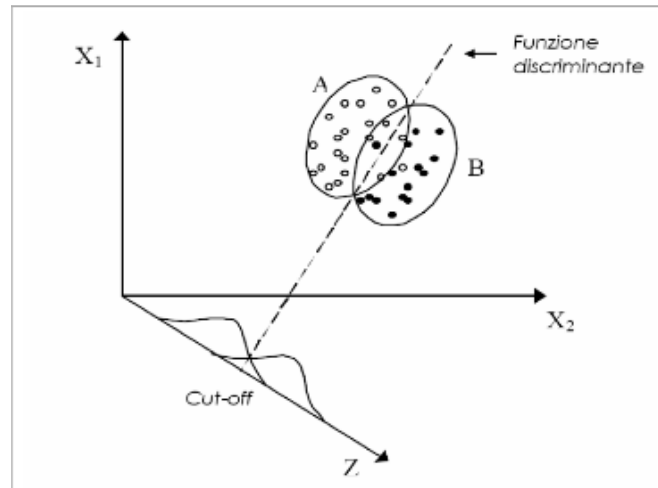
L'analisi discriminante nasce come strumento di classificazione nella prima formulazione di Fischer del 1936: essa costituiva un metodo per classificare e allocare due popolazioni nella popolazione di provenienza. Sebbene tale accezione si discosti in buona misura dall'obiettivo delle tecniche di scoring, la facilità discriminatoria dell'analisi l'ha resa uno strumento utile per la misurazione del rischio di credito e la classificazione delle unità nei due insiemi di creditori (solvibili e non).

L'obiettivo del modello era quello di costruire un'affidabile regola di classificazione con la quale distinguere le popolazioni e di classificare un'unità in una delle suddette categorie in linea con il potenziale comportamento che la stessa avrebbe avuto in futuro.

La classificazione avviene utilizzando una combinazione lineare delle variabili considerate per le due categorie, adottando la funzione che massimizza la distanza tra le medie delle due popolazioni a parità di variazioni.

La seguente figura rappresenta un'esplicazione grafica semplificata del modello di Fischer. Le imprese potenzialmente solvibili sono rappresentate nel gruppo A, mentre quelle potenzialmente

insolventi nel gruppo B, e sono rispettivamente rappresentate dalle variabili X_1 e X_2 . Lo score generato prende il nome di funzione discriminante, e viene riportato sull'asse Z, rappresentante il valore soglia che distingue le popolazioni.



5

La combinazione lineare delle variabili che genera lo score può essere riassunta nella seguente formula:

$$S_j = \sum_{i=1}^n \alpha_j * X_{ij}$$

Dove:

- S rappresenta lo score della j -esima impresa;
- n rappresenta il numero di variabili osservate.
- α_j rappresenta il coefficiente della variabile X ;
- X_{ij} rappresenta la variabile descrittiva i -esima dell'impresa;

La regola di classificazione lineare viene espressa in termini di distanze tra gli score, per cui l'impresa j -esima sarà assegnata alla categoria di aziende solvibili se vale la formula:

$$|S_j - S_1| < |S_j - S_2|, \text{ ovvero } S_j < \frac{1}{2} * (S_1 + S_2), \text{ per } S_1 < S_2$$

Altrimenti viene assegnato alla seconda categoria. Il termine $\frac{1}{2} * (S_1 + S_2)$ rappresenta quindi il *cut-off* per la classificazione.

⁵ Fonte: "Rischio e Valore nelle banche – misura, regolamentazione, gestione" – A. Resti, A. Sironi, Egea, 2008

1.4 MOTIVAZIONI ALLA BASE DELL'UTILIZZO DEI METODI DI CREDIT SCORING

Gli elementi che hanno portato ad una gestione automatizzata di informazioni riguardo il potenziale cliente e alla conseguente assegnazione di un punteggio sulla relativa affidabilità creditizia sono essenzialmente due: una sempre maggiore spersonalizzazione del rapporto tra le due controparti, dovuta alla crescita esponenziale sia nel volume delle richieste che nelle tipologie di prodotti finanziari offerti; la contemporanea e crescente competizione nel mercato finanziario, che ha reso necessaria la tempestività nella valutazione della richiesta del cliente, evitando ritardi che genererebbero fenomeni di selezione avversa della clientela. Ancora, il ricorso a tecniche di scoring automatizzate in operazioni di concessione del credito deriva necessariamente dalla più stringente e rigorosa legislazione finanziaria degli ultimi anni in merito alla gestione del rischio di credito, che verrà trattata in maniera specifica nel secondo capitolo.

È tuttavia impossibile non considerare altre motivazioni che hanno condotto le istituzioni erogatrici di prestiti ad adottare metodi di credit scoring. I progressi delle tecnologie dell'informazione e della comunicazione, un maggiore adattamento nella trasmissione e nell'elaborazione delle informazioni e la pressione competitiva, che spinge le banche a comprimere i costi, hanno aumentato gli incentivi e le opportunità per l'impiego di tecnologie di prestito che si affidano più intensamente alle informazioni hard (Petersen, 2004).

Tra le menzionate, il risparmio sui costi è sicuramente un incentivo chiave. Tale affermazione risulta particolarmente appropriata quando le banche acquistano *score* da fornitori esterni e li usano per rifiutare o accettare decisioni automatizzate e impostare termini di credito. Attraverso questo *modus operandi* la banca rinuncia sostanzialmente alla scelta degli input e dei pesi del modello, lasciando relativamente poca discrezionalità ai suoi funzionari addetti alla concessione di finanziamenti. Il metodo presentato, utilizzato soprattutto per crediti alle piccole imprese, risulta sicuramente l'applicazione più economica e rapida del credit scoring.

Altri incentivi all'utilizzo di tali tecniche sono la riduzione del problema dell'opacità e la definizione dei termini contrattuali in modo più accurato, al fine di ridurre future perdite su crediti e contemporaneamente aumentare le entrate attese sugli stessi.

Questi miglioramenti possono essere più probabili quando il punteggio di credito viene utilizzato insieme a un'altra tecnologia di prestito, ovvero analizzando i punteggi insieme alle informazioni raccolte utilizzando una o più tecnologie aggiuntive. Si ritiene che questa implementazione di credit scoring per le piccole imprese utilizzi maggiore discrezione, perché la banca lascia più margine di

operatività ai suoi funzionari addetti ai prestiti, per valutare diversi tipi di informazioni nell'approvazione o nel rifiuto di potenziali crediti e nella definizione contrattuale dei prestiti.

È tuttavia probabile che questa strategia aumenti i costi di sottoscrizione poiché la banca sostiene la spesa per il punteggio di credito oltre al costo dell'altra tecnologia di prestito. La distinzione presentata non è da sottovalutare: gli effetti dell'adozione del credit scoring diminuiscono infatti in relazione a come la banca decide di implementare la tecnologia. Elemento centrale è l'importanza che il credit scoring ha nella decisione di concedere il credito: se viene utilizzato automaticamente ed è il fattore principale che influenza tale decisione (regole) o se è solo un ingrediente tra gli altri (discrezione, appunto).

1.5 VANTAGGI APPORTATI DALL'UTILIZZO DEL CREDIT SCORING

Un appropriato utilizzo del credit scoring, soprattutto se applicato in operazioni che riguardano piccole e medie imprese, può apportare benefici a vari di *stakeholder* coinvolti nel processo finanziario.

Per i finanziatori, per esempio, il credit scoring riesce ad aumentare la redditività dei prestiti: grazie alla riduzione di tempi e costi necessari per l'approvazione del finanziamento, si rende possibile un ampliamento delle opportunità di prestito e dunque maggiori opportunità di ricavi, anche tramite specifiche tecniche di *pricing* basate sul rischio. A tal proposito, un'indagine della Federal Reserve di Atlanta⁶ ha dimostrato che le banche che attuano valutazioni creditizie tramite credit scoring ricevono un premio sui prestiti superiore del sette per cento rispetto alle banche che non lo utilizzano, in parte proprio a causa dell'applicazione del *risk based pricing*. Ancora, si stima⁷ che il costo totale della richiesta e della sottoscrizione di un prestito medio per le PMI elaborato senza punteggio sia compreso tra i 500 e i 1.000 dollari, mentre il costo stimato per l'elaborazione dello stesso, tramite credit scoring, sia compreso tra 48 e 95 dollari.

Anche i mutuatari ne beneficiano per svariate ragioni. In primo luogo, essi godono di una maggiore possibilità di accesso al credito: analizzando il già menzionato studio della Federal Reserve, si può osservare che in generale l'utilizzo di tecniche statistiche di scoring ha aumentato il numero di banche concedenti credito alle PMI, registrando un aumento del 37% del rapporto tra i prestiti alle PMI e il totale dei prestiti commerciali tra le banche che si avvalgono di credit scoring per i piccoli *business* (SBCS).

⁶ Berger, Allan M. and W. Scott Frame, "Small Business Credit Scoring and Credit Availability". Federal Reserve Bank of Atlanta (Maggio 2005)

⁷ Fonte: Fair Isaac and Company

In secondo luogo, l'utilizzo delle tecniche di scoring porta ad una maggiore trasparenza nel processo decisionale sull'approvazione del prestito e in generale ad una maggiore obiettività. Il credit scoring aiuta infatti i finanziatori a garantire l'applicazione di uguali criteri di sottoscrizione a tutti i mutuatari, prescindendo da informazioni sociodemografiche irrilevanti. L'utilizzo del credit scoring appare dunque molto utile anche in relazione all'“*Equal credit opportunity act*”, che vieta l'uso dei suddetti fattori (quali sesso, razza o estrazione sociale) nelle decisioni di credito, imponendo a sua volta che l'unica correlazione dei pesi di un modello sia con il merito creditizio. Il prestatore deve sostanzialmente dimostrare che ci sia una ragione commerciale per utilizzare un determinato fattore: un modello di credit scoring per la valutazione creditizia facilita indubbiamente l'utilizzo di fattori quantitativi nella fase di scelta binaria sulla concessione del finanziamento.

Ancora, i mutuatari beneficiano di un'accelerazione nelle tempistiche di approvazione, che possono ridursi da circa dodici ore a meno di una (anche se ciò dipende soprattutto dal fatto che la tecnologia venga usata come strumento per produrre decisioni automatiche o in maniera complementare ad altri metodi di valutazione, come già visto). Tale risparmio di tempo si tramuta in risparmio di costi e va a beneficio anche del cliente, che deve fornire solo le informazioni utilizzate nel sistema di punteggio, in modo che le applicazioni possano essere più brevi.

Infine, vi sono benefici anche a livello macro e microeconomico generale: il trasferimento di risorse verso investimenti produttivi in PMI, promuove inevitabilmente la crescita economica, con risvolti positivi in termini di occupazione, produzione e più equa distribuzione delle opportunità economiche.

1.6 EFFETTI DEL CREDIT SCORING SULLA SELEZIONE DEI MUTUATARI E SULLA PERFORMANCE DEL CREDITO

L'adozione del credit scoring può avere grandi effetti sulle modalità con cui le banche e le istituzioni finanziarie originano e monitorano i prestiti, e, allo stesso tempo, sull'offerta di credito e sui tassi di insolvenza dei mutuatari.

Innanzitutto, la possibilità di disporre di informazioni aggiuntive (eventualmente più adatte e precise) può indurre le banche ad una maggiore concessione di prestiti e ad una valutazione migliore del merito di credito dei mutuatari. L'adozione di tali tecniche può consentire alle banche, infatti, di abbassare i tassi, concedendo credito a mutuatari precedentemente marginali che verrebbero probabilmente respinti dal processo di sottoscrizione manuale.

Inoltre, se le informazioni hard ricavate dai modelli di credit scoring sostituiscono le informazioni soft tradizionalmente raccolte, le banche possono interrompere il credito ai mutuatari per i quali non sono disponibili le informazioni quantitative necessarie per i modelli di scoring.

Al netto di ciò, l'adozione del credit scoring non modifica la propensione delle banche a concedere prestiti alla generalità dei mutuatari, ma consente loro di selezionarne di più sicuri, riducendo la probabilità che un mutuatario sia inadempiente. Quest'ultimo effetto è sicuramente più forte per le banche che dichiarano di utilizzare il credit scoring principalmente come strumento per monitorare i mutuatari (e dunque che si muovono nel conteso del già menzionato *behavioural scoring*) e per i mutuatari più piccoli.

Altro nodo cruciale è invece l'effetto dell'adozione del credit scoring sul *banking lending* alle piccole e medie imprese, soprattutto in relazione alla sempre maggiore rilevanza della valutazione del merito di credito dei mutuatari dopo la crisi finanziaria e alla grande diffusione delle tecnologie di credit scoring tra le banche negli ultimi vent'anni. I risultati di uno studio svolto dalla Banca d'Italia⁸ mostrano che l'adozione del credit scoring per le piccole imprese non porta le banche ad allentare gli standard di credito: in primo luogo, la propensione a concedere prestiti alla generalità dei mutuatari non è influenzata dall'adozione di SBCS; in secondo luogo, le banche sono meno disposte a concedere prestiti ai mutuatari più rischiosi. L'effetto sull'andamento del credito è dunque positivo: il credit scoring riduce la probabilità di insolvenza di un credito.

Concludendo, possiamo affermare che, da un punto di vista teorico, il processo di indurimento delle informazioni soft nei prestiti, unito ai progressi della tecnologia della comunicazione, comporta principalmente vantaggi per le banche, facilitando il processo decisionale sulla concessione del credito e il monitoraggio del merito creditizio dei mutuatari. Da un punto di vista politico, invece, è bene annoverare che il quadro normativo di Basilea II ha consentito alle banche di basarsi su modelli interni di rating dei mutuatari per il calcolo dei loro coefficienti patrimoniali. I nuovi e più severi requisiti patrimoniali imposti da Basilea III e l'adozione di stress test per valutare la capacità del bilancio delle banche di sostenere periodi sfavorevoli del ciclo macroeconomico hanno ulteriormente portato le banche a migliorare le loro tecniche di credit scoring.

Dando per ormai assodata la sopramenzionata diminuzione dei costi dell'attività creditizia consentita dall'utilizzo delle tecniche di credit scoring, è possibile affermare, alla luce di quanto

⁸ “Le tecniche di credit scoring migliorano la selezione della clientela e la qualità del credito?”, Giorgio Albareto, Roberto Felici e Enrico Sette, ottobre 2016

detto, che le preoccupazioni sulle potenziali conseguenze indesiderate dell'adozione del credit scoring sono sopravvalutate, e che in realtà l'adozione porta ad un aumento della qualità del credito con un impatto complessivamente positivo sulla performance bancaria.

CAPITOLO 2: RISCHIO DI CREDITO E RELATIVA REGOLAMENTAZIONE

2.1 DEFINIZIONE DI RISCHIO DI CREDITO

Come accennato, il ricorso all'utilizzo di tecniche di credit scoring automatizzate deriva anche e soprattutto dalla stringente regolazione del rischio di credito introdotta dalle normative di Basilea. L'intento di questo capitolo è di analizzare in profondità il concetto di rischio di credito e i metodi per la relativa gestione.

Con la locuzione 'rischio di credito' si intende la possibilità che una variazione inattesa del merito creditizio di una controparte generi una corrispondente variazione inattesa del valore corrente della relativa esposizione creditizia.⁹

Ai fini di una migliore comprensione della suddetta definizione, occorre fare delle specifiche su alcune importanti tematiche.

Innanzitutto, appare evidente che il rischio di credito non riguardi solamente il verificarsi di un evento binario di fallimento (il creditore è o non è solvibile), ma incorpori allo stesso tempo il cosiddetto rischio di *downgrading*, ovvero il rischio connesso ad un deterioramento del merito creditizio della controparte. In tal caso, si parla di *mark-to-market paradigm*, in cui si tratta l'insolvenza solamente come evento estremo e si vagliano le possibilità di un aumento graduale della probabilità del verificarsi della stessa (trattata invece come evento quantale nel *default-made-paradigm*, in cui si analizza la variazione del merito creditizio in seguito ad avvenuta insolvenza).

Altro nodo cruciale è quello dell'attendibilità della variazione del merito creditizio. Nei prossimi paragrafi analizzeremo in dettaglio le componenti del rischio di credito, evidenziando la distinzione tra perdita attesa ed inattesa e come tale differenza sia determinante ai fini della quantificazione e gestione del rischio stesso.

Infine, appare opportuno considerare che, se si analizza il concetto di rischio di credito, è impossibile prescindere dall'analizzare altri rischi ad esso correlati:

- Il rischio che il tasso di recupero effettivamente registrato al termine della liquidazione delle attività di una controparte divenuta insolvente risulti inferiore a quanto originariamente stimato dalla banca;

⁹ A.Resti, A. Sironi, "Rischio e Valore nelle banche – misura, regolamentazione, gestione", Egea

- Il rischio di controparte, ovvero il rischio che la controparte di una transazione in derivati negoziati in un mercato *over-the-counter* (OTC) divenga insolvente prima della scadenza dello stesso e renda dunque necessario per la banca sostituire la posizione sul mercato a condizioni contrattuali differenti;
- Il rischio Emittente (per titoli obbligazionari), ossia il rischio connesso ad un eventuale rialzo degli spread richiesti dal mercato agli emittenti di una data classe di merito creditizio;
- Il rischio Paese, ovvero il rischio che una controparte non residente non sia in grado di adempiere alle proprie obbligazioni a causa di eventi di natura politica o legislativa, quali, per esempio, l'introduzione di vincoli valutari, che impediscono alla stessa controparte di rimborsare il proprio debito.

2.2 GESTIONE DEL RISCHIO DI CREDITO: ANALISI DELLE COMPONENTI

Il rischio di credito può esser dunque legato sia alle perdite subite dalla banca in relazione all'insolvenza del debitore (rischio di insolvenza) che al deterioramento del merito creditizio (rischio di migrazione). Entrambi i casi analizzati possono portare a una riduzione inattesa del valore attuale delle attività iscritte nel bilancio bancario, che a loro volta possono comportare l'insolvenza della banca stessa.

La gestione del rischio di credito ruota, come accennato, attorno al concetto di attendibilità della perdita. Le perdite parziali o totali che può subire la banca vengono infatti distinte in perdite attese (*expected losses - EL*) e perdite inattese (*unexpected losses - UL*), con importanti implicazioni sia a livello di rischiosità che a livello di relativa gestione. Prima di analizzare in dettaglio entrambe le categorie, occorre specificarne le principali differenze.

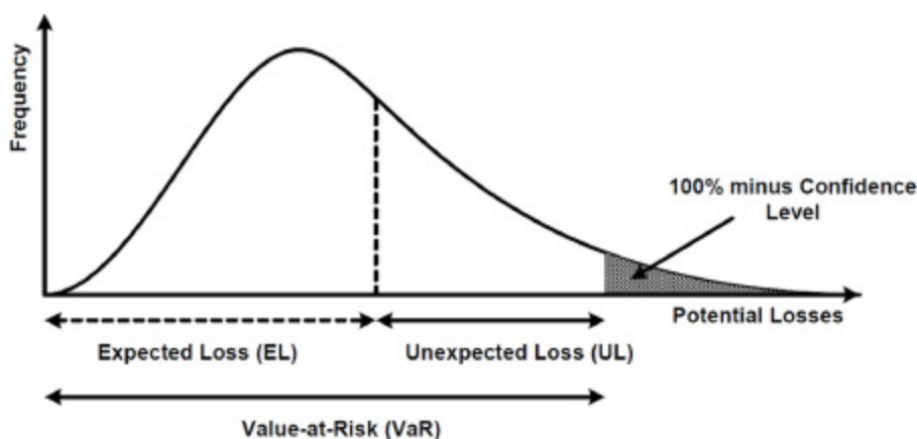
La perdita attesa non costituisce il vero rischio di un'esposizione creditizia: proprio perché è "attesa", viene stimata ex-ante e direttamente incorporata in termini di spread sulle condizioni di prezzo applicato dal mercato all'emittente di un'attività finanziaria, dando luogo ad una rettifica del valore dell'attivo o ad un accantonamento a fondo rischi passando attraverso il conto economico.

La componente del rischio di credito che rappresenta il vero rischio è invece la perdita inattesa, ovvero il rischio che la perdita effettiva risulti, a posteriori, maggiore di quella inizialmente stimata. Essa non si accantona, ma trova adeguata copertura nel patrimonio della banca; il modo migliore per calcolare l'incidenza sul patrimonio delle perdite inattese è il *Value-at-Risk*, seguito da *stress testing*.

Tali componenti si differenziano soprattutto per la possibilità di riduzione del rischio tramite diversificazione. La perdita attesa non può essere eliminata attraverso una politica di diversificazione del portafoglio, quanto al più stabilizzata mediante ampliamento dello stesso, ossia attraverso il conseguimento di un numero di impieghi della medesima natura tale da garantire, in base alla legge dei grandi numeri, che il livello di perdita media sia effettivamente quello conseguito dal portafoglio impieghi della banca. Ciò che conta, nello specifico, è la varianza delle perdite attese, perché è da lì che vengono calcolati gli accantonamenti da costituire: più si riesce a ridurre le varianze, maggiore è la probabilità che *l'expected loss* riesca ad essere predittiva.

La perdita inattesa, invece, può essere ridotta mediante un'adeguata politica di diversificazione del portafoglio - per aree geografiche, settori produttivi e classi dimensionali - che mira a diminuire il grado di correlazione dei vari impieghi.

La figura seguente riassume quanto finora detto.



10

2.2.1 EXPECTED LOSS

La perdita attesa è definita come il valore medio della distribuzione delle perdite che un'istituzione creditizia si attende di subire su un portafoglio prestiti.¹¹

Sostanzialmente, rappresenta un costo atteso dalla banca a fronte dell'esposizione creditizia nei confronti del debitore.

¹⁰ Fonte: Bank for international Settlements (2005): An Explanatory Note on Basel II IRB Risk Weight Functions

¹¹ Rossi M., (2003), *Perdita attesa, perdita inattesa e diversi*

L'utilizzo dell'*expected loss* viene introdotto dal principio IFRS 9, che impone come unico modello di *impairment* la rilevazione delle perdite attese in relazione al merito di credito dello strumento (metodo *expected credit loss* – ECL); esso è contrapposto al precedente IAS 39 che prevedeva sia diversi modelli di *impairment* in funzione della classificazione degli strumenti finanziari, ma soprattutto la rilevazione delle perdite sulla base di quelle verificatesi e non attese (metodo IBNR). Attraverso l'attuale metodo l'ammontare delle perdite attese viene aggiornato ad ogni data di riferimento del bilancio per riflettere la variazione del rischio. L'obiettivo è quello di rilevare le perdite attese per tutta la durata del rapporto e per tutti gli strumenti finanziari che hanno avuto un significativo incremento del rischio di credito dalla loro registrazione, siano essi valutati individualmente o collettivamente, considerando tutte le informazioni disponibili e incorporando elementi prospettici.

La stima della perdita attesa riferita alla singola posizione è ottenuta dal prodotto di tre variabili:

$$EL = PD * EAD * LGD$$

Dove:

PD = *probability of default*;

EAD = *exposure at default*;

LGD = *loss given default*.

2.2.1.1. PROBABILITY OF DEFAULT

La probabilità di *default* (PD) può essere definita come la probabilità di insolvenza della controparte nei confronti della quale si è esposti. La previsione della probabilità di *default* risulta fondamentale perché, per ogni mutuatario o gruppo di mutuatari, è l'input chiave per la stima del patrimonio di vigilanza e del capitale economico per le banche.

La stima della PD può seguire 5 principali approcci o modelli:

1. modelli analitico-soggettivi;
2. modelli statistici (modelli di analisi discriminante; modello probit/logit, analizzato nel terzo capitolo; modelli basati su reti neurali);
3. modelli fondati sui giudizi delle agenzie di *rating*;
4. modelli fondati sui dati di mercato dei capitali;

5. modelli di *option pricing*.

Prescindendo dal modello adottato, l'EBA¹² - *European Banking Authority* – stabilisce dei requisiti dei dati per lo sviluppo del modello. Gli enti sono infatti tenuti a garantire che l'insieme dei dati di riferimento contenga i valori dei fattori di rischio alle date appropriate, tenuto conto delle dinamiche e della frequenza di aggiornamento dei suddetti fattori durante l'intero periodo di presenza di un debitore nel portafoglio. In particolare, i fattori di rischio dovrebbero includere: caratteristiche del debitore, informazioni finanziarie e sulle tendenze dell'attività (es. crescita/contrazione del fatturato) e informazioni comportamentali, garantendo l'affidabilità delle informazioni nel tempo.

2.2.1.2 EXPOSURE AT DEFAULT

L'*exposure at default* (EAD) rappresenta l'esposizione in caso di insolvenza, ovvero la stima del valore effettivo del credito al verificarsi dello stato di insolvenza.¹³ Rappresenta sostanzialmente l'esposizione finanziaria del creditore corretta per l'eventuale presenza di un'opzione di aumento dell'esposizione in capo alla controparte.

Per calcolare quantitativamente l'EAD, occorre prima effettuare una distinzione sulla forma tecnica del prodotto finanziario collegato all'esposizione (in sostanza, sulla tipologia di finanziamento).

Se l'istituto conosce l'importo esatto del finanziamento concesso – es. per mutui e *leasing* – l'esposizione è detta “a valore certo”, e viene registrata nel *trading book* della banca. La stima è data da:

$$EAD = EC + EP$$

Dove:

- EC = esposizione corrente (*mark-to-market* della posizione)
- EP = esposizione potenziale (valore massimo che la posizione potrebbe assumere in relazione a variazioni del prezzo di mercato, considerato un determinato intervallo di confidenza).

Se invece l'istituto non conosce l'importo esatto del finanziamento concesso, non può determinare l'importo immediatamente, ma solo ad insolvenza verificata (es. scoperto su conto corrente). In tal caso l'esposizione viene detta “a valore incerto” e viene registrata nel *banking book* dell'istituto. La stima è data da:

¹² Guidelines on PD and LGD estimation, (2018).

¹³ “A. Resti” - “Misurare e gestire il rischio di credito: una guida per le banche” - Milano (2001)

$$EAD = UT + (ACC - UT) * \%TIRAGGIO$$

Dove:

- UT = utilizzato corrente, anche detto *drawn portion*;
- (ACC – UT) = porzione dell'accordato non ancora utilizzata, anche detta *undrawn portion*;
- %TIRAGGIO = percentuale della quota non utilizzata che si ritiene venga utilizzata dal debitore in corrispondenza del default. È anche detta *usage given default* (USG).

2.2.1.3 LOSS GIVEN DEFAULT

La *loss given default* (LGD) rappresenta la perdita di credito subita dalla banca in caso di insolvenza di un debitore della stessa.¹⁴

Una volta che si è verificato un evento di default, la perdita connessa può essere di tre tipi:

- La perdita del capitale;
- I costi di gestione dei crediti in sofferenza, come il mancato pagamento di interessi attivi;
- Spese di allenamento (riscossioni, legali, ecc.).

Ci sono sostanzialmente tre modi per misurare la LGD di uno strumento:

1. *Market* LGD. È il metodo utilizzato dalle agenzie di rating che consiste nell'osservare i prezzi di mercato delle obbligazioni o dei prestiti negoziabili in sofferenza subito dopo l'effettivo evento di inadempienza. La caratteristica principale di tali prezzi è che possono essere facilmente tradotti in una percentuale di recupero e che riflettono il recupero atteso dall'investitore, opportunamente attualizzato, includendo dunque i recuperi sia sul capitale scontato che sui mancati pagamenti di interessi.
2. *Workout* LGD. È il metodo che si basa sugli effettivi flussi di recupero realizzati nei periodi successivi al *default*, raggruppando i dati in base al tipo di procedura seguita, di debitore o di esposizione, per creare dei *clusters* con caratteristiche affini e analoghe di LGD e poi stimare quella attesa sui *default* futuri.
3. *Implied Market* LGD. Sono spesso utilizzati come controllo rispetto ai modelli di rating più convenzionali e consistono nel derivare LGD dai prezzi delle obbligazioni rischiose (non in *default*) utilizzando un modello teorico di *pricing* delle attività.

¹⁴ Federal Reserve Bank of New York, "What we do know about loss given default", 2004

In formule la *loss given default* può essere scritta come:

$$LGD = 1 - RR$$

Dove:

RR = tasso di recupero (*recovery rate*).

Tale tasso può dipendere da vari fattori, tra cui presenziano fattori interni alla banca, tipologie di garanzie applicate, stadio del ciclo economico, spese amministrative e operative. La formula può essere così riscritta come:

$$LGD = 1 - \frac{RA - CA}{EAD * \%GAR} * (1 + i)^{-t}$$

Dove:

- RA = recupero atteso;
- CA = costi amministrativi;
- EAD = *exposure at default*;
- %GAR = grado di copertura della garanzia;
- i = tasso di attualizzazione del *cash flow*;
- t = tempo stimato per il recupero del credito.

2.2.2. UNEXPECTED LOSS

Con perdita inattesa si intende “la misura del grado di variabilità della perdita attorno al proprio valore atteso”.¹⁵ Come già affermato, rappresentando il rischio che la perdita sia maggiore di quella stimata ex-ante, rappresenta il vero rischio di un’esposizione creditizia.

Avendo già analizzato nell’introduzione del paragrafo le caratteristiche dell’*unexpected loss* in relazione a diversificazione di portafoglio e copertura patrimoniale, e le relative differenze con l’*expected loss*, l’esposizione è limitata alla misurazione analitica della perdita.

In formule si ha che:

$$UL = \sqrt{LGD * EDF * \sigma_{EDF}^2 * \sigma_{LGD}}$$

¹⁵ Resti A. & Sironi A., (2008), *Rischio e valore nelle banche*, Milano, Egea, p.356.

Dove:

- $LGD = \text{loss given default}$;
- EDF = tasso di insolvenza atteso;
- $\sigma_{EDF}^2 = \text{varianza del tasso di insolvenza}$;
- $\sigma_{LGD} = \text{varianza del tasso di perdita}$.

È opportuno ricordare che nella suddetta scrittura della formula si considera come ipotesi di base l'indipendenza tra LGD e EDF: nel caso opposto di andamento dipendente andrebbe inserita anche la covarianza fra le due variabili.

Infine, concludiamo la trattazione sulla perdita inattesa specificando che la menzionata copertura patrimoniale viene disciplinata dalle normative introdotte dal comitato di Basilea sui requisiti minimi di capitale, trattati nel seguente paragrafo.

2.3 BASILEA II

Al fine di contenere e fronteggiare prospetticamente i rischi insiti nell'attività bancaria, che sono potenzialmente in grado di accrescere il valore delle passività e ridurre quello delle attività, le banche devono mantenere livelli di patrimonializzazione adeguati.

La normativa bancaria in tema di requisiti minimi di capitale trae origine nel primo documento redatto dal Comitato di Basilea per la Vigilanza Bancaria, denominato Accordo sul capitale e anche poi detto "Basilea I". Tale comitato è un'istituzione internazionale di sorveglianza delle operazioni bancarie, che promuove la stabilità del sistema attraverso regole che definiscono la metodologia che deve essere usata per valutare ogni tipo di rischio ed il capitale minimo che la banca deve detenere.

"Basilea I" introduceva un coefficiente di solvibilità – con valore maggiore o uguale all'8% dell'importo del credito ponderato per il *credit risk* dello stesso – come misura di adeguatezza patrimoniale bancaria. I limiti dell'accordo, in particolare nella non considerazione del grado di diversificazione del portafoglio prestiti, della vita residua delle esposizioni e degli strumenti di *credit mitigation*¹⁶, hanno portato il Comitato a varare una nuova versione dello stesso.

¹⁶ Si fa riferimento a strumenti di protezione del credito reale o personale utilizzati da molte banche.

Tale revisione culmina nel cosiddetto “Basilea II”, che entra in vigore nel 2007 e si fonda su tre pilastri.

1. Il primo pilastro si occupa dei requisiti patrimoniali minimi, vincolando in maniera più rigida le banche, soprattutto quelle attive internazionalmente, a far riferimento a quattro tipologie di rischio effettivo: rischio di credito, di mercato, di controparte e operativo. Occorre tuttavia specificare che i cambiamenti apportati da Basilea II non riguardano la mera quantificazione del coefficiente di solvibilità, quanto piuttosto le metodologie utilizzate per calcolare i rischi (soprattutto per il rischio di credito – analizzato nei paragrafi successivi – e per quello operativo).
2. Il focus del secondo pilastro si sposta invece sul miglioramento del processo di sorveglianza, soprattutto in relazione alla qualità del *risk management* delle istituzioni bancarie e dell’adeguatezza delle procedure di valutazione del capitale.

Tale processo di supervisione – il *Supervisory Review Process* – si costituisce di due fasi:

- i. Il processo interno di determinazione dell’adeguatezza patrimoniale (*Internal Capital Adequacy Assessment Process*, ICAAP).
 - ii. Il processo di revisione e valutazione prudenziale (*Supervisory Review and Evaluation Process*, SREP).
3. Il terzo pilastro mira a migliorare la disciplina del mercato aumentando la trasparenza del sistema, in modo che i partecipanti al mercato possano valutare più accuratamente la solidità patrimoniale, le esposizioni creditizie di una banca, i sistemi di gestione e controllo e i relativi modelli di *rating*. In particolare, si prevede la preparazione di tavole sinottiche che classifichino le informazioni da pubblicare in base ai rischi.

2.3.1 METODI DI PONDERAZIONE DEL RISCHIO DI CREDITO

Come già affermato, l’obiettivo del capitolo è quello di analizzare più approfonditamente il concetto di rischio di credito e i modelli utilizzati per la relativa gestione. Alla luce di ciò appare opportuno tralasciare gli altri interessanti, ma poco attinenti, aspetti di Basilea II e soffermarsi sui metodi di ponderazione del rischio, ossia sugli specifici sistemi di valutazione e calcolo che considerano affidabilità del cliente, caratteristiche dell’operatore e presenza di garanzie.

Rispetto a Basilea I, in cui si classificava ogni attività bancaria assegnando una determinata ponderazione di rischio a ciascuna categoria, il rischio di credito trattato in Basilea II considera la

sensibilità dei requisiti patrimoniali rispetto al merito di credito dei debitori. La formula di vigilanza tiene conto non solo della classe di debitori, ma anche del loro merito creditizio, rendendola il fulcro del legame tra la qualità degli attivi e il patrimonio di vigilanza e rinforzando quello tra rischio effettivo e coefficienti patrimoniali.

Per il calcolo dei coefficienti patrimoniali del rischio di credito – le già citate variabili PD, EAD, LGD, la *Maturity* dello strumento e il giudizio sul merito del creditore, il *rating* – si possono utilizzare differenti approcci. In particolare, il Comitato ha vagliato la possibilità dell'utilizzo di diversi sistemi di valutazione considerando le diverse esigenze dei vari istituti e relative differenze nei sistemi informativo-gestionali. Due sono i metodi di ponderazione del rischio previsti: un metodo Standard (*standardised approach* – SA) ed un metodo basato sui rating interni (*internal rating based* – IRB).

In entrambi i metodi i requisiti patrimoniali sono determinati in termini di attività ponderate per il rischio (RWA), stimate moltiplicando ciascuna ponderazione per il valore dell'attività corrispondente. Nell'ambito del metodo standardizzato, le ponderazioni di rischio associate a ciascuna esposizione sono fisse e costanti nel tempo, mentre nel metodo basato sui *rating* interni le ponderazioni dipendono dai modelli interni di rischio delle banche (convalidati dalle autorità di vigilanza). Proprio perchè i coefficienti patrimoniali sono legati al rischio patrimoniale con il metodo IRB, mentre sono fissati nel tempo con il SA, la letteratura ha sottolineato che il patrimonio di vigilanza è più sensibile al rischio per le banche che utilizzano l'approccio basato sui *rating* interni che per quelle utilizzando quello standardizzato.

Occorre comunque specificare che l'autonomia concessa alle banche nella determinazione della ponderazione del rischio viene mitigata dal Comitato, che ha stabilito dei requisiti di garanzia a cui attenersi in modo da garantire l'idoneità dei metodi di rilevazione del rischio del credito da parte della banca. Esse sono tenute a definire con completezza e chiarezza quali sono i criteri secondo cui verranno suddivise e distinte le classi di rating, considerando sia le esposizioni creditizie che la rischiosità complessiva del cliente.

2.3.1.1 STANDARDISED APPROACH

Il sistema standard prevede che i fattori di ponderazione del rischio siano tutti formulati da soggetti esterni all'istituto di credito.

Il *rating* viene valutato da agenzie indipendenti accreditate, dette ECAI¹⁷, mentre PD, LGD, EAD e *Maturity* sono fissati dall'Autorità di Vigilanza in base a determinate caratteristiche (categoria giuridico-economica di appartenenza dell'impresa, dimensioni aziendali, caratteristiche tecniche dell'operazione). Le banche che adottano questo sistema segmentano i loro crediti in categorie prudenziali e ad ognuna corrisponde una ponderazione del rischio fissa.

La figura illustra lo schema predisposto dalle due maggiori agenzie esterne, Standard & Poor's e Moody's: a seconda della tipologia di appartenenza, al richiedente prestito viene assegnata una determinata categoria di rischio.

Coefficienti di ponderazione prefissati correlati al rating assegnato alle imprese da società esterne:

VALUTATORE	CATEGORIE DI RATING = VOTO ASSEGNATO ALLE IMPRESE DA SOCIETÀ ESTERNE				
	SICUREZZA		VULNERABILITÀ	RISCHIOSITÀ	
STANDARD- POOR'S	da AAA a AA-	da A+ a A-	da BBB+ a BB-	< a BB-	SENZA RATING
MOODY'S	da Aaa a Aa3	Da A1 a A3	Da Baa1 a B3	< a B3A	SENZA RATING
PONDERAZIONE AL: (coefficienti di ponderazione prefissati)	20%	50%	100%	150%	100%
PONDERAZIONE UNICA PER LE PMI	75 %				

*le categorie di rischio vanno da AAA per le imprese meno rischiose a BB/B3 per quelle più rischiose.*¹⁸

Come è possibile leggere dalla tabella, i coefficienti di ponderazione sono diversi e dipendono dal *rating* che l'impresa richiedente finanziamento riceve dall'ECAI; possono ancora non avere *rating* (ultima colonna) o avere un'unica ponderazione (al 75%, per le imprese *retail*).

In formule, si può riassumere lo *Standardised Approach* in questo modo:

$$\text{ESPOSIZIONE} * \text{COEF. DA RATING ESTERNO} * 8\% = \text{REQUISITO PATRIMONIALE}$$

¹⁷ External Credit Assessment Institution

¹⁸ Fonte: Camera di commercio di Macerata.

2.3.1.2 APPROCCI INTERNAL RATING BASED

Le banche che hanno adottato i sistemi IRB effettuano il calcolo del *rating* del cliente al loro interno, valutando direttamente il rischio del credito e determinando il valore del credito stesso. Utilizzando tale metodo le banche calcolano i coefficienti patrimoniali in base alla propria stima del rischio associato a ciascuna esposizione del proprio portafoglio.

L'utilizzo di modelli basati su *rating* interni ha rappresentato un importante cambiamento nella valutazione regolamentare del rischio di credito: il metodo IRB consente di esercitare una maggiore discrezionalità nella valutazione del merito creditizio del mutuatario poiché le banche possono stimare il rischio di ogni esposizione in proprio (discrezionalità che si riduce in presenza di un *benchmark* esterno di riferimento).

La misurazione del merito creditizio viene effettuata sempre attraverso la stima delle variabili già citate in precedenza, ma, a differenza dell'approccio standardizzato, sono calcolate dalla banca (in tutto o in parte), e non prefissate dall'Autorità di Vigilanza. Gli elementi necessari per calcolare i coefficienti patrimoniali della banca sono il risultato di una combinazione tra input quantitativi, forniti dalle banche stesse, e formule indicate dal Comitato.

Tale approccio ha due versioni, una di base e una avanzata.

2.3.1.2.1 INTERNAL RATING BASED FOUNDATION

La banca che adotta l'approccio base (FIRB) del sistema basato sui *rating* interni utilizza sistemi di calcolo per determinare il giudizio sul merito creditizio del cliente ed individuare la classe di *rating* da assegnare, associando automaticamente una determinata PD. Per le altre componenti di rischio (LGD, EAD, M) sono invece utilizzati valori regolamentari. In tal modo il modello non si allontana molto da quello standard, dovendo effettuare in concreto solo il calcolo del *rating*.

Al fine di valutare coloro che richiedono un prestito, le banche devono predisporre un sistema di *rating* oggettivo, affidabile e che contenga criteri determinati, prevedendo sostanzialmente delle classi di rischio a cui associare una soglia minima e massima di probabilità di *default*.

Basilea II riconosce al metodo FIRB una categoria più ampia di strumenti di mitigazione del rischio rispetto all'approccio standardizzato, consentendo di ridurre la PD e migliorare la classe di *rating* del richiedente.

2.3.1.2.2 INTERNAL RATING BASED ADVANCED

La versione avanzata del metodo IRB è strutturata in un modello totalmente autonomo in cui la valutazione creditizia del cliente è totalmente effettuata dalla banca. In tal modo le variabili ormai note sono stimate attraverso sistemi di calcolo piuttosto complessi.

Il sistema AIRB può essere utilizzato solamente dagli istituti che ottengono una certificazione dall'Autorità di vigilanza del paese di appartenenza: all'interno dei criteri validati per la concessione, gli istituti possono detenere comunque una soglia di flessibilità.

Tale modello riconosce una categoria più ampia di strumenti in grado di mitigare il rischio, che consentono di ridurre oltre alla PD, come nel modello FIRB, anche la LGD.

In generale, il calcolo dei requisiti patrimoniali comporta la classificazione del portafoglio bancario in cinque categorie di esposizioni: imprese, clientela al dettaglio, banche, soggetti sovrani e patrimonio netto.

Per ognuno di essi, il Comitato propone una ponderazione del rischio basata su una specifica funzione continua in cui inserire i parametri (PD, LGD, EAD e M) e dando così luogo alla ponderazione del rischio per la singola esposizione. Un esempio di funzione continua può essere la seguente.

$$RWA = K \cdot 12,5 \cdot EAD$$

Dove:

1. Il requisito patrimoniale minimo (MCR) è l'8% dell'RWA;
2. Le attività ponderate per il rischio si ottengono dalla moltiplicazione dell'EAD per 12,5 e il fattore K (determinato in base al tipo di esposizione).

2.4 LIMITI DI BASILEA II

La crisi finanziaria ha sottolineato diversi punti deboli dell'accordo di Basilea II, che è stato criticato e considerato da molti, almeno parzialmente, responsabile della crisi stessa. Tali critiche non appaiono del tutto coerenti: è infatti possibile osservare che l'accordo è entrato in vigore

quando la crisi si era già ampiamente diffusa nei mercati finanziari; pur avendo evidenti limiti, riconosciuti dallo stesso comitato, l'accordo ha solamente esacerbato alcuni aspetti della crisi che ne hanno evidenziato le carenze.

I principali limiti di Basilea II possono riassumersi nei seguenti punti:

- Qualità e livello del capitale. La quantità di capitale in grado di prevenire le insolvenze per le banche era essenzialmente superiore a quanto fissato dal comitato. Inoltre, il patrimonio di vigilanza era formato da strumenti di bassa qualità: su tutti strumenti ibridi o innovativi che, anche se considerati come parte di capitale dagli istituti di vigilanza, erano percepiti dagli investitori come strumenti di debito (e che sono stati utilizzati proprio per rimborsare il capitale associatogli, e non per assorbire le perdite).
- Prociclicità, ovvero la tendenza dell'accordo ad accentuare fluttuazioni del ciclo economico¹⁹. Dovendo fronteggiare una fase di recessione, i requisiti di capitale sono necessariamente aumentati, costringendo le banche a *credit crunch* o a vendita di titoli. Ciò ha finito inevitabilmente per accentuare la fase negativa del ciclo.
- Aumento del *leveraging*. Pur detenendo livelli di capitale coerenti con l'impianto normativo, nel periodo di crisi e in quello immediatamente successivo le banche hanno aumentato a dismisura la leva finanziaria (non vincolata da nessun limite in Basilea II). Il processo di *deleveraging* successivo, portato avanti dalle banche al fine di aumentare il coefficiente patrimoniale, ha aiutato la solvibilità finanziaria delle singole istituzioni, ma ha portato instabilità generale nei mercati.
- Liquidità. Il rischio di liquidità²⁰ era trattato da Basilea II nel secondo pilastro non attraverso un requisito esplicito e uniforme, ma piuttosto attraverso un invito alle banche a dotarsi di strumenti per misurarli costantemente e gestirli. Tale mancanza di regole atte a prevenire specifiche crisi di liquidità si è mostrata quando, nel periodo post crisi, la caduta improvvisa della liquidità dei mercati ha provocato situazioni di elevata tensione per gli intermediari.
- Banche sistemiche. Le banche sistemiche sono banche ritenute fondamentali per l'intero sistema finanziario in virtù delle loro elevate interconnessioni con altre istituzioni. Basilea

¹⁹ Resti A. & Sironi A., (2008), Rischio e valore nelle banche, Milano, Egea,

²⁰ Rischio derivante dal fatto che, per far fronte a un improvviso e rapido aumento nel ritiro di o all'esercizio di un diritto di credito fuori bilancio, un intermediario debba liquidare attività in tempi brevi e a prezzi inferiori al loro valore di mercato.

Il non prevedeva specifici requisiti per tali banche, non considerando in tale maniera che le suddette interconnessioni potevano favorire la diffusione degli shock negativi attraverso il sistema finanziario.

- Rischi di mercato nel *trading book*. Le attività finanziarie inserite nel portafoglio di negoziazione – *trading book* – devono essere riportate al *fair value*, ed hanno pertanto risentito a pieno del crollo dei mercati. I requisiti patrimoniali previsti da Basilea II, che sottostimavano i rischi derivanti da strumenti finanziari complessi, sono risultati insufficienti ad assorbire le perdite.

2.5 BASILEA III

Concludiamo il capitolo sul rischio di credito e sulla relativa regolamentazione con una trattazione sintetica del nuovo accordo sul capitale pubblicato a fine 2010.

Con l'intento di rivedere ed integrare le regole di Basilea II, e nel tentativo di correggerne i limiti, il Comitato di Basilea ha avviato un processo di riforma sfociato in un ulteriore accordo, denominato Basilea III, destinato ad avere i suoi effetti per il periodo compreso tra il 2013 ed il 2020. Le finalità di Basilea III sono quelle di immunizzare l'attività bancaria da eventuali crisi, ridurre il rischio di trasmissione dei relativi effetti all'economia reale e rafforzare la solidità bancaria.

Il rafforzamento della qualità del patrimonio di vigilanza detenuto dalla banca è un fattore cardine di Basilea III, che affida maggiore importanza al *common equity*²¹. Inoltre, l'enfasi posta sulla quantità di capitale si è tradotta anche in interventi di natura micro e macro-prudenziale, con l'imposizione di due *buffer* patrimoniali:

- *buffer* di conservazione del capitale, formato da azioni ordinarie e riserve da utili in misura pari al 2,5% dell'attivo ponderato per il rischio. Impone vincoli alle distribuzioni discrezionali di utile se il capitale della banca scende al di sotto di determinati livelli.
- *buffer* anticiclico, in caso di rischio sistemico: le Autorità di vigilanza possono imporre una dotazione addizionale di copertura patrimoniale dell'attivo ponderato per il rischio, compresa tra 0 e 2,5%.

²¹ Parte di capitale costituita da azioni ordinarie, utili non distribuiti e riserve di utili al netto delle perdite.

Per quanto attiene al rischio di eccessiva leva finanziaria, Basilea III introduce anche un nuovo *leverage ratio* che contiene al proprio interno anche le poste iscritte fuori bilancio. È un indicatore semplice, trasparente e non basato sul rischio, calibrato in modo da ridurre il rischio di processi di *deleveraging* che possono arrecare danno al sistema finanziario.

$$\text{Leverage ratio} = \frac{\textit{Tier 1}}{\textit{Totale attivo + fuori bilancio}} \geq 3\%$$

Al fine di prevedere il rischio di liquidità delle banche sono stati inoltre introdotti dei requisiti prudenziali minimi, a presidio della solvibilità del singolo intermediario e del sistema nel complesso. Nello specifico:

- Il *Liquidity coverage ratio*, per assicurare un livello adeguato di attività di elevata qualità convertibili agevolmente in cassa per far fronte a uno scenario di *stress* acuto di breve periodo.
- Il *Net Stable funding ratio*, che promuove un rapporto equilibrato fra fonti di finanziamento “stabili” e corrispondente fabbisogno di finanziamento a medio-lungo termine di una banca.

Per le banche giudicate di rilevanza sistemica, infine, sono stati introdotti dei requisiti patrimoniali aggiuntivi che riguardano l’applicazione di un coefficiente patrimoniale con *common equity Tier 1* (da 1 a 2,5% dell’attivo ponderato per il rischio).

CAPITOLO 3: IL MODELLO DI REGRESSIONE LOGISTICA

3.1 CONCETTI INTRODUTTIVI PROPEDEUDICI ALLA COMPrensIONE DEL MODELLO

Il modello di regressione logistica, anche noto come modello *Logit* o modello logistico, è di norma utilizzato per analizzare la relazione causale tra una variabile dipendente dicotomica ed una o più variabili indipendenti quantitative.

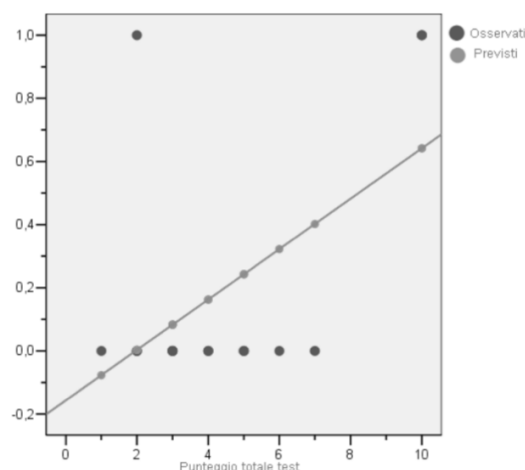
Se la variabile dipendente è appunto dicotomica – e assume valori 0 o 1 – la distribuzione teorica di riferimento dovrebbe essere la distribuzione binomiale (e non quella normale). In tali casi, anche se in realtà è possibile applicare il modello della regressione semplice, matematicamente parlando, un modello non lineare sarebbe maggiormente opportuno.

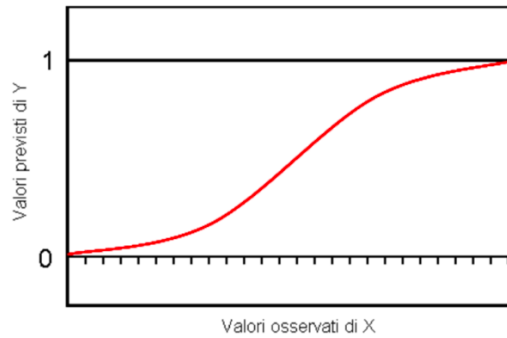
Il modello di regressione lineare semplice assume infatti che:

$$Y = a + bx + \varepsilon$$

Dove Y è una variabile di risposta continua e x una continua esplicativa, mentre ε esprime l'effetto di fattori che concorrono a formare il valore di Y .

Nella sua formulazione il modello lineare semplice implica che i valori assunti dalla Y possono essere compresi in un intervallo che va da $-\infty$ a $+\infty$. Tale implicazione risulta inadeguata considerando che, come già riportato, la variabile dipendente può assumere valori 0 o 1. In sostanza, essendo la variabile dipendente dicotomica influenzata dalla X , per valori molto alti di X il valore in Y dovrebbe avvicinarsi ad 1 senza superare tale limite (e in egual misura per lo 0 nel caso opposto). La relazione corretta tra X e Y viene riportata nella seconda figura, rappresentante la relazione nel modello logistico. I concetti spiegati appaiono inoltre più evidenti attraverso un confronto con la prima figura, che riporta invece il modello di regressione lineare semplice con variabile dipendente dicotomica.





22

La funzione rappresentata è chiamata funzione logistica, ed è una sigmoide che assume valori sempre compresi tra 0 e 1.

Prima di addentrarci più specificatamente nel modello di regressione logistica multipla, occorre fare ancora un'introduzione più dettagliata sul concetto di *odds*, che sarà fondamentale ai fini della stima del modello.

L'*odds* è un modo di esprimere una probabilità mediante un rapporto. Considerati due livelli, esso viene calcolato attraverso il rapporto fra frequenze osservate in un livello e quelle osservate nell'altro, al fine di esprimere il rapporto tra due categorie. A titolo esemplificativo si riporta un breve esempio.

Considerato un gruppo di persone formato da 25 uomini e 16 donne, la probabilità che un individuo scelto casualmente dal gruppo sia uomo è pari al 60%. Esprimendo la stessa informazione tramite l'*odds*, avremo una relazione tra uomini e donne pari a 1,56 (per ogni donna sono presenti 1,56 uomini). Sinteticamente si avrà che:

$$P(U) = 25 / 41 = 0,609$$

$$Odds(U) = 25 / 16 = 1,5625$$

In termini probabilistici, si esprime l'*odds* come:

$$Odds(X_1, X_2, \dots, X_k) = \frac{P(Y = 1 | X_1, X_2, \dots, X_k)}{1 - P(Y = 1 | X_1, X_2, \dots, X_k)}$$

Ancora, si può valutare l'associazione tra due variabili (ovvero esprimere la relazione tra due categorie in funzione di un'altra variabile) utilizzando l'*odds ratio* (anche detto rapporto tra gli *odds*). Esso è un indice ottenuto dal rapporto tra gli *odds* di una variabile (es. Y) per ciascun livello

²² Regressione Multipla e Regressione Logistica: concetti introduttivi ed esempi, Vincenzo Paolo Senese, 2014

di un'altra variabile (es. X). Solitamente si tende ad interpretare gli *odds ratio* in relazione alla distanza dell'indice dal valore unitario: se tale rapporto non è uguale a 1, si evidenzia associazione tra le variabili considerate. Riportiamo la formulazione matematica semplificata dell'*odds ratio*, che ci sarà utile nel quarto capitolo.

$$ODDS\ RATIO = \frac{\frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}}{\frac{1}{1 + e^{\beta_0 + \beta_1}}} \bigg/ \frac{\frac{e^{\beta_0}}{1 + e^{\beta_0}}}{\frac{1}{1 + e^{\beta_0}}}} = \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = e^{\beta_1}$$

Si conclude ricordando che, se si considera una distribuzione congiunta delle variabili X e Y, e si confronta la distribuzione delle frequenze, di quelle relative, degli *odds* e dei *Logit* (ottenuti come logaritmo naturale degli *odds*) si nota che tali statistiche forniscono essenzialmente la stessa informazione, sebbene con valori matematicamente differenti.

3.2 DESCRIZIONE GENERALE DEL MODELLO LOGIT

Il modello *Logit* è probabilmente il modello più utilizzato tra i modelli di scelta discreta, soprattutto per la facilità di interpretazione della probabilità di scelta (che assume forma chiusa).

Tuttavia, il modello non ha immediata interpretazione, poiché non si specifica una relazione tra le variabili osservate, ma tra la probabilità di un evento e le variabili esplicative.

Analizzando il modello logistico nell'ambito del credit scoring, la distribuzione su cui si focalizza l'analisi è infatti quella della variabile casuale di classificazione Y in relazione ai valori $X = (x_1, x_2, \dots, x_p)$ delle variabili esplicative. Il modello lega lo *score*, di cui si è già parlato nel primo capitolo, alle menzionate variabili esplicative, in una forma lineare nei coefficienti.

Il modello di regressione logistica può essere interpretato in diversi modi:

- Tramite l'interpretazione basata sugli *odds* (come accennato), utilizzati soprattutto nel linguaggio del *betting* e nel gergo medico-sanitario, biologico e ambientale;
- Attraverso l'analisi degli effetti marginali, utilizzata spesso dai sociologi;
- Utilizzando il concetto di variabile latente (interpretazione usata in ambito econometrico).

La costruzione di un modello di regressione logistica si basa essenzialmente sull'implementazione di quattro fasi:

- Selezione del campione;
- Selezione preliminare delle variabili esplicative;
- Stima del modello;
- Verifica della capacità predittiva del modello.

3.3 SELEZIONE DEL CAMPIONE

La costruzione di un modello logistico vede come prima fase quella della selezione di un numero r di imprese sufficientemente alto per rendere il campione statisticamente rilevante.

Selezionate le imprese, solitamente scelte in modo da avere come risultato finale un raggruppamento omogeneo in relazione a determinate variabili – si veda il paragrafo 1.2 –, esse vengono distinte in due gruppi in base al valore assunto da uno specifico indicatore economico-finanziario. In particolare, nel nostro ambito di studio del *risk management* e della stima della probabilità di *default* della controparte di un'operazione finanziaria, le imprese sono classificate come segue:

- Insolventi, se presentano valori della $Y=1$;
- Solventi, se presentano valori della $Y=0$.

Come già discusso nel capitolo introduttivo sulle tecniche di scoring, la classificazione delle imprese avviene sulla base del superamento del *cut-off* del valore assunto dall'indicatore economico-finanziario.

Si conclude menzionando che uno dei tipici problemi dei meccanismi di credit scoring è quello dello sbilanciamento della distribuzione della Y in favore delle unità buone, che comporta dunque una maggiore presenza di unità solventi rispetto a quelle insolventi. In alcuni modelli più complessi, è perciò possibile che si ricorra al cosiddetto campione bilanciato, in cui si segue uno schema di campionamento in cui si campionano le unità buone con una possibilità di inclusione inferiore rispetto a quelle delle cattive.

3.4 SELEZIONE PRELIMINARE DELLE VARIABILI ESPLICATIVE

Se il modello di regressione logistica viene applicato in ambito finanziario per la stima della probabilità di insolvenza della controparte richiedente un finanziamento, le variabili esplicative del modello sono solitamente degli indici economico-finanziari deducibili dal bilancio (come è riscontrabile nel quarto capitolo).

In generale, la scelta delle variabili esplicative di un modello *Logit* è una delle fasi più importanti per la costruzione dello stesso: la selezione del modello si riduce infatti alla scelta delle variabili più adatte a formare la matrice e all'eliminazione di quelle che non hanno influenza rilevante nella formazione dello score.

Oltre a rendere più semplice e rapido il processo operativo, una scelta accurata dei parametri comporta conseguenze sulla relativa affidabilità: al crescere delle variabili scelte diminuisce infatti l'accuratezza delle stime. Risulta pertanto necessario utilizzare un numero limitato di variabili indipendenti per eliminare i problemi di *overfitting* e multicollinearità.

Con il primo termine si intende il fenomeno per il quale si ha una varianza gonfiata degli *odds ratio* stimati dovuta ad un rapporto troppo elevato tra numero di osservazioni (soggetti trattati nel modello) e numero di covariate incluse nel modello di punteggio²³. Al fine di potere ricavare stime abbastanza affidabili, studi di simulazione oggettiva dimostrano infatti che il rapporto tra parametri e numero di eventi debba essere non inferiore a 1/10 o 1/20.

Il problema della multicollinearità si presenta invece quando, in modelli statistici espressi mediante un'equazione lineare, le variabili sono fra loro fortemente correlate, per cui risulta molto difficile, e talora impossibile, individuare separatamente l'influenza delle variabili e anche ottenere una stima sufficientemente attendibile dei loro singoli effetti.²⁴

Per procedere alla scelta delle variabili da inserire nel modello, si seguono delle procedure dette "incrementali", poiché vengono svolte per passi successivi confrontando modelli gerarchici tra loro.

Le tecniche di scelta delle variabili sono essenzialmente tre:

- Procedura *backward*. In tale procedura si parte da un modello saturo rimuovendo progressivamente alcuni parametri. Il *driver* che guida l'eliminazione è il *p-value* (si

²³Overfitting in propensity score model, Journal of Clinical Epidemiology, Hajage, David, Tubach, Florence; De Rycke, Yann, 2017, volume 88

²⁴ Treccani, vocabolario on line

eliminano i parametri con valore più elevato fino ad avere tutti parametri con valori uniformi, inferiori ad una soglia α prefissata).

- Procedura *forward*. Si può considerare quasi opposta alla procedura precedente, perché parte da un modello che presenta solo l'intercetta aggiungendo man mano il parametro con *p-value* minore.
- Procedura *stepwise*. Il metodo *stepwise* può essere considerato un ibrido dei modelli precedenti, perché alterna fasi di inclusione di parametri a fasi di esclusione (seguendo le logiche menzionate).

È bene ricordare che le procedure conducono a modelli simili tra loro, soprattutto se è presente una forte relazione tra risposta e variabile esplicativa. Inoltre, nell'ambito del credit scoring più generale, le tecniche menzionate sono spesso utilizzate apportando piccole variazioni.

Il grande vantaggio apportato dalle tecniche incrementali è quello di risultare facilmente implementabili nei software statistici e dunque portare velocemente a modelli finali. A sottolineare la sopracitata importanza della selezione di variabili nei processi di credit scoring, in cui è spesso presente una conoscenza aprioristica del fenomeno, si ricorda infine che è opinione diffusa quella di preferire l'utilizzo di un modello non considerato ottimale ma in cui sono contenute variabili statisticamente rilevanti sulla base di tali conoscenze a priori.

3.5 STIMA DEL MODELLO

Pur disponendo di una variabile dipendente dicotomica, e potendo come detto applicare un modello di regressione lineare semplice, è più appropriato utilizzare un modello non lineare per evitare i problemi già spiegati nel paragrafo introduttivo del capitolo. La non linearità dei parametri non consente infatti l'applicazione del metodo dei minimi quadrati²⁵; prescindendo da ciò, occorre effettuare alcune trasformazioni per rendere la relazione lineare nei termini dei parametri e derivare in questo modo la formulazione finale del modello.

La trasformazione utilizzata è la trasformazione logaritmica della variabile dipendente. Nel modello di regressione logistica studiato tale variabile definisce l'appartenenza di un soggetto ad uno dei due

²⁵ Anche detto OLS (*Ordinary Least Squares*), il metodo dei minimi quadrati è una tecnica di regressione che permette di trovare una funzione, rappresentata da una curva di regressione, che si avvicini il più possibile ad un insieme di dati. Il metodo dei minimi quadrati può essere utilizzato per la stima dei parametri, anche se ormai viene utilizzato più spesso il metodo della massima verosimiglianza per cui non sussiste il complicato processo di verifica degli assunti (presente nel primo). Si è dunque scelto di trattare il metodo della massima verosimiglianza nel paragrafo 3.5.1. È bene ricordare che, se le assunzioni sono verificate, i due metodi portano a risultati identici.

gruppi, in cui i valori assegnati ai livelli sono scelti in modo arbitrario. Al netto di ciò, ciò che interessa non è tanto il valore predetto, come nella regressione lineare, ma la probabilità di appartenenza ad uno dei due gruppi del soggetto considerato. L'utilizzo del già menzionato *odds*, al posto della probabilità, consente di risolvere il problema dell'assegnazione dei livelli.

Per esprimere in termini lineari la relazione variabile dipendente/indipendente si può applicare la trasformazione esponenziale alla funzione che esprime il valore atteso della variabile dipendente (ovvero la probabilità $Y = \mu_Y = P_{(X=1)}$).

Si ottiene che la funzione lineare di X che esprime la probabilità di $Y = 1$, ovvero

$$P(Y = 1) = \alpha + \beta X$$

Diventa:

$$P(Y = 1) = e^{\alpha + \beta X}$$

Anche tale trasformazione non risolve completamente il problema del restringimento dei valori dall'intervallo $-\infty$ a $+\infty$ a quello $[0;1]$ (ovvero l'intervallo di valori assumibili dalla probabilità), poiché restringe il *range* solamente tra 0 e $+\infty$.

La trasformazione adeguata è, appunto, la trasformazione logistica, che trasforma l'equazione in:

$$P(Y = 1) = \frac{e^{\alpha + \beta X}}{1 + e^{\alpha + \beta X}}$$

Possiamo in questo modo esprimere il rapporto tra la probabilità di *default* e quella di non *default*, ovvero l'*odds* di $Y = 1$, come:

$$Odds_{Y=1} = \frac{\frac{e^{\alpha + \beta X}}{1 + e^{\alpha + \beta X}}}{\frac{1}{1 + e^{\alpha + \beta X}}} = e^{\alpha + \beta X}$$

Infine, applicando la proprietà dei logaritmi per la quale $\ln(e^x) = x$, il logaritmo naturale del suddetto *odds* è funzione lineare della variabile $X - \ln(odds_{Y=1}) = \alpha + \beta X$.

La formulazione finale del modello è dunque:

$$P(Y = 1 | \mathbf{x}) = \frac{e^{(\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p)}}{1 + e^{(\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p)}}$$

$$P(Y = 0 | \mathbf{x}) = \frac{1}{1 + e^{(\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p)}}$$

Si ricorda che nel modello di regressione logistica non sono presenti le variabili casuali errori ε_i poiché il modello è direttamente espresso, come visto, attraverso una trasformazione del valore medio della variabile dipendente Y che diventa la variabile casuale di interesse.

McFadden dimostra a tal proposito che il modello logistico presuppone per le variabili casuali ε_i una distribuzione di Gumbel che è simmetrica ed abbastanza simile (ad eccezione delle code) alla distribuzione Normale.

3.5.1 STIMA DEI PARAMETRI

Originariamente, il metodo iterativo utilizzato per stimare i parametri del modello era il già menzionato metodo dei minimi quadrati ponderati. Sovente è però vero che lo stesso non può essere applicato al modello di regressione logistica, poiché non ne sono completamente verificati gli assunti.

La stima viene dunque effettuata attraverso il metodo della massima verosimiglianza (*ML - maximum likelihood*). Tale metodo stima i parametri del modello in modo da massimizzare la funzione (*log-likelihood function*) che quantifica la probabilità ottenere il valore atteso di Y dati i valori delle variabili indipendenti.

La stima dei parametri di un modello di regressione logistica attraverso il metodo *maximum likelihood* si basa dunque sulla suddetta funzione di verosimiglianza, L. Essa si presenta nella formulazione finale come:

$$L(\beta_0, \beta_1) = \prod_{i=1}^n f(y_i | x_i) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{(1-y)}$$

Nella formula riportata si specifica che $\pi(x_i)^{y_i}$ è la probabilità che la variabile dipendente Y assuma valore 1 per una data osservazione i , mentre $[1 - \pi(x_i)]^{(1-y)}$ è, ovviamente, la probabilità che sia 0. Partendo da esse e scrivendo prima la distribuzione di probabilità condizionata, considerato che la probabilità congiunta è il prodotto delle probabilità condizionate per ogni osservazione i^{26} , si ottiene la suddetta funzione di verosimiglianza.

La stima ottimale della verosimiglianza è raggiunta attraverso un processo iterativo fino al punto in cui la capacità di miglioramento converge: sostanzialmente, si parte inserendo dei valori arbitrari che vengono poi modificati per migliorare la funzione. La stima della massima verosimiglianza dei parametri è ottenuta determinando il valore di β che massimizza il logaritmo di $L(\beta)$, che viene detto log-verosimiglianza.

Le condizioni per la massimizzazione sono due:

- La derivata prima rispetto ai parametri deve essere posta uguale a zero per trovare i punti estremanti;
- la derivata seconda, sempre rispetto ai parametri, è posta minore di zero per determinare le concavità della funzione (garantendo in tal modo che quelli trovati siano solo punti di massimo).

Nel caso specifico del modello di regressione logistica studiato, la distribuzione di probabilità condizionata è:

$$\begin{aligned}
 P(Y_i = y_i | X_{1i} \dots X_{ki}) &= p_i^{y_i} (1 - p_i)^{1-y_i} \\
 &= \left(\frac{e^{(\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p)}}{1 + e^{(\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p)}} \right)^{y_i} \left(1 - \frac{e^{(\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p)}}{1 + e^{(\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p)}} \right)^{1-y_i}
 \end{aligned}$$

Considerato quanto affermato poc'anzi sulla probabilità congiunta, scriviamo dunque la funzione log-verosimiglianza da cui saranno stimati i parametri.

²⁶ Si assume che per ogni osservazione $(X_{1i}, X_{2i}, \dots, X_{ki}, Y_i)$ siano indipendenti e identicamente distribuite.

$$\ln(\beta) = \sum_{i=1}^n \left[y_i \ln\left(\frac{e^{(\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p)}}{1 + e^{(\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p)}}\right) (1 - Y_i) \ln\left(1 - \frac{e^{(\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p)}}{1 + e^{(\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p)}}\right) \right]$$

3.5.2 VALUTAZIONE FIT DEL MODELLO

Con la locuzione “valutazione *fit*” si intende la valutazione della bontà di adattamento del modello (*goodness-on-fit*). Le misure di bontà di adattamento sono solitamente usate per descrivere la discrepanza tra valori osservati e attesi nel modello considerato.

Per quanto riguarda il modello di regressione logistica, si utilizzano statistiche molto simili a quelle che esprimono l’adeguatezza del modello nella riproduzione dei dati nella regressione lineare (ossia l’errore standard della stima e l’ R^2 , che esprime la parte di varianza della variabile dipendente spiegata attraverso il modello).

Considerato che per la scelta dei parametri da inserire nel modello si utilizza il *log-likelihood*, utilizziamo per la valutazione il valore dello stesso moltiplicato per -2, abbreviato con -2LL. Distinguiamo due casi.

Se il modello contiene la sola intercetta, il valore di -2LL rappresenta ciò che nella regressione lineare semplice corrisponde alla devianza, indicata con D_0 . Essa è calcolata come:

$$D_0 = -2 \{ \{n_{Y=1} \ln[P(Y = 1)] + n_{Y=0} \ln[P(Y = 0)] \}$$

Dove:

- $n_{Y=1}$ è il numero di casi in cui $Y=1$,
- $n_{Y=0}$ quelli in cui è uguale a 0,
- N il numero totale dei casi;
- $P(Y = 1) = n_{Y=1}/N$ è la probabilità che Y sia uguale a 1.

Nel modello che invece contiene, oltre all’intercetta, anche la variabile indipendente, -2LL è la devianza d’errore (indicata con D_M), ovvero la parte di variabilità dei dati non spiegata dal modello. La differenza tra D_0 e D_M è dunque la parte di variabilità delle indipendenti spiegata dal modello. Essa è indicata con G_M e si ottiene come:

$$G_M = D_0 - D_M$$

Tale valore è solito essere chiamato anche Chi-quadrato e, indicato col simbolo χ^2 , rappresenta quantitativamente la riduzione di errore derivante dall'uso del modello considerato.

Si specifica che tale affermazione è vera solo se i modelli sono nidificati (*nested*), ovvero se tutti i termini contenuti nel modello A sono anche contenuti in un modello B, e non viceversa (i parametri di A sono un sottoinsieme dei parametri di B). Si può così utilizzare il valore G_M per verificare l'ipotesi nulla del modello:

$$H_0 \Rightarrow \beta_1 = \beta_2 = \beta_3 = \dots = \beta_k = 0$$

Si rifiuta tale ipotesi quando il valore ha una $p < 0,05$ (ovvero se G_M è statisticamente significativo): ciò equivale sostanzialmente a dire che la previsione che Y sia uguale a 1 è migliorabile considerando i predittori.

Mantenendo inoltre l'ipotesi che la statistica $-2LL$ possa essere considerata come devianza della regressione, il rapporto di verosimiglianza può essere utilizzato per ottenere una statistica detta pseudo R^2 o indice di Mc Fadden (simile all' R^2 utilizzato nella regressione lineare semplice). In formule:

$$R_L^2 = \frac{G_M}{D_0} = \frac{G_M}{G_M + D_M}$$

Tale statistica quantifica la riduzione dell'errore dovuta al modello (ovvero di quanto la considerazione di determinati predittori riduce la variazione dei dati).

3.5.3 CONTRIBUTO DEI SINGOLI PARAMETRI

Il contributo apportato da ogni variabile indipendente su quella dipendente si analizza valutando la significatività di ogni parametro.

Per effettuare tale valutazione, si prendono in considerazione i coefficienti di regressione: se si è interessati alle unità di misura, si analizzano coefficienti non standardizzati; se si vuole evidenziare la relazione tra variabili in termini di deviazione standard, si utilizzano coefficienti standardizzati (ovvero quei coefficienti calcolati su variabili che come unità di misura presentano la deviazione standard).

La statistica test maggiormente utilizzata per valutare il contributo di ogni predittore è probabilmente la statistica di Wald (W_k):

$$W_k^2 = \left(\frac{b_k}{S_{b_k}} \right)^2$$

Dove con il termine al numeratore si indica un generico parametro e con quello al denominatore l'errore standard della stima.

3.6 VERIFICA DELLA CAPACITÀ PREDITTIVA DEL MODELLO

Si conclude il capitolo trattando, infine, la verifica della capacità predittiva del modello, ovvero la verifica sulla capacità del modello di determinare correttamente l'appartenenza ad uno dei due gruppi considerati.

In tal caso l'attenzione si sposta sulla tabella delle classificazioni, che nel modello di regressione logistica è una 2x2 in cui si confrontano i valori previsti dal modello con quelli osservati (simile a quella riportata nel paragrafo 1.2.1).

L'indice più utilizzato è il seguente:

$$\text{Efficienza predittiva} = \frac{(\text{errori senza il modello}) - (\text{errori con il modello})}{\text{errori senza il modello}}$$

Gli errori con il modello rappresentano i casi in cui valore osservato \neq valore predetto, mentre quelli senza il modello possono essere calcolati in modi differenti in relazione all'uso che si fa del modello stesso.

CAPITOLO 4: APPLICAZIONE DEL MODELLO DI REGRESSIONE LOGISTICA PER VALUTARE IL RISCHIO DI CREDITO DI UN CAMPIONE DI IMPRESE

Il quarto e ultimo capitolo dell'elaborato ha come obiettivo quello di applicare praticamente il modello di regressione logistica studiato.

Si noter  attraverso la lettura che la struttura dello stesso risulta molto simile a quella del capitolo precedente. Tale scelta non   casuale, ma strumentale a sottolineare la stretta relazione tra teoria e applicazione pratica e la tangibilit  ed adeguatezza del modello per la valutazione della solvibilit  delle imprese. Tale adeguatezza viene in particolare analizzata attraverso la verifica della capacit  predittiva del modello, riportata nella parte finale del capitolo.

4.1 SELEZIONE DEL CAMPIONE

Come riportato nel primo capitolo, gli effetti pi  rilevanti del credit scoring vengono evidenziati soprattutto nella concessione del prestito alle piccole e medie imprese o nel credito al consumo.

Tale motivazione aveva inizialmente portato alla scelta di un campione di imprese dal settore pi  ampio delle PMI innovative. Tuttavia, considerate le difficolt  di reperimento di dati finanziari per la suddetta categoria (soprattutto per i valori del ROE), si   scelto di selezionare 42 imprese appartenenti al settore delle startup innovative, considerando la relativa affinit , per struttura e caratteristiche, con la categoria precedentemente menzionata.²⁷

La denominazione "Startup innovative"   introdotta nel 2012 per rappresentare un nucleo di imprese giovani, ad alto contenuto tecnologico, e con forti potenzialit  di crescita. Tali imprese contribuiscono in maniera notevole allo sviluppo innovativo e tecnologico italiano, rappresentando dunque uno dei punti focali della politica industriale nazionale.

L'iscrizione di un'impresa nell'apposita sezione del registro delle imprese, che avviene solamente in seguito ad adeguata verifica del possesso di determinati requisiti²⁸, comporta ad esse svariate agevolazioni, tra le quali si menzionano l'accesso gratuito e semplificato al fondo di garanzia per le piccole e medie imprese, incentivi fiscali all'investimento nel capitale, esonero da diritti camerali e imposte di bollo, deroghe alla disciplina societaria ordinaria.

²⁷ Dati ottenuti tramite l'utilizzo della piattaforma AIDA

²⁸ Essi sono: nuova costituzione o avvenuta in un periodo inferiore a 5 anni, residenza in Italia, fatturato inferiore a 5 milioni, non quotazione in un mercato regolamentato, non distribuzione di utili, oggetto sociale esclusivo e prevalente in commercializzazione di un prodotto o servizio ad alto valore tecnologico.

Come già affermato, una volta scelto il campione, le imprese che ne fanno parte sono divise in due gruppi in base ai valori assunti da uno specifico indicatore economico-finanziario. La distinzione effettuata nel nostro caso avviene in base ai valori assunti dall'indice di indipendenza finanziaria, o indice di patrimonializzazione.

L'indice di indipendenza finanziaria è calcolato attraverso il rapporto tra il patrimonio netto (ovvero il finanziamento tramite propri mezzi) e l'indebitamento finanziario netto (approvvigionamento finanziario garantito da fonti esterne). Tale indice è considerato come una delle migliori misure per valutare la struttura finanziaria di un'impresa e più in particolare la relativa stabilità, proprio perché indica la proporzione tra mezzi propri e mezzi di terzi. Inoltre, risulta essere uno degli indici di struttura più completi perché, oltre a specificare le fonti di finanziamento dell'impresa, è un buon parametro per analizzare il rischio di investimento di terzi finanziatori nell'impresa stessa.

In base al valore che assume l'indice si possono delineare diverse tipologie di struttura finanziaria. Se l'impresa presenta un indice con valori negativi o uguali a 0, non ha mezzi propri e si finanzia totalmente all'esterno; per valori positivi ma inferiori a 0,55, l'impresa presenta una struttura gravemente ($< 0,33$) o mediamente squilibrata; per valori superiori a tale soglia ma minori di 1 l'impresa risulta essere mediamente equilibrata.

Al netto di ciò, il *cut-off* che permette di definire propriamente un'impresa come sana è quello unitario: imprese con indice maggiore di 1 presentano minore incidenza dell'indebitamento finanziario netto rispetto *all'equity*, risultando pertanto raramente vulnerabili a livello strutturale. Alla luce di quanto detto, sono state classificate come imprese sane tutte le imprese che presentano valori del rapporto considerato superiori a 1; sono invece state classificate come "rischiose" le imprese con valori inferiori.

Sulla base di questa classificazione avremo:

- 28 imprese sane, con indice di indipendenza finanziaria > 1 ;
- 14 imprese rischiose, con indice < 1 .

I valori dell'indice delle imprese selezionate e la relativa classificazione sono riportati in tabella.

IMPRESE SANE	VALORE DELL'INDICE	IMPRESE RISCHIOSE	VALORE DELL'INDICE
Absolutelx SRL	1,34	AN-KY SRLS	0,74
Alipalo SRL	1,18	Athic SRL	0,87
Banco Energia Forniture SRL	1,4	Autismo domani SRL	0,59
Beprime SRL	1,14	Autismo domani SRL	0,55
Bio Circulatory Sistem delovepment	1,45	Bruscatech SRL	0,65
BT solutions SRLS	1,41	Energia Agricola Verde	0,96
Centro ricerca Ratinf SRL	1,19	Gruppostg SRL	0,55
Domotica Lucana SRL	1,09	We Laser SRL	0,55
Eco Pets	1,46	Wire Gaming SRL	0,9
Fibreacustics SRL	1,39	Società agricola Methabios	0,85
Gruppo IMC innova SRL	1,34	Negozi diretti SRL	0,97
Hendal SRL	1,02	Oroverde Società Agricola	0,88
Intensivecare SPA	1,1	S.I.DRO	0,8
Jotto SRL	1,2	SM energy SRL	0,58
LSH società benefit	1,37		
Mattermove SRL	1,44		
Needo SRL	1,19		

OMM Technologies SRL	1,05
Omnitek SRL	1,01
Organic & Blue Solutions SRL	1,07
Pearfid SRL	1,37
PolyHedrica SRL	1,5
Revive SRL	1,36
Salusmundi SRLS	1,46
Tecnopolimeri International SRL	1,32
Thrive X Company	1,31
Ticopter SRL	1,11

29

4.2 SELEZIONE DELLE VARIABILI ESPLICATIVE

La scelta delle variabili esplicative è, come detto, una delle fasi più importanti della costruzione di un modello di scoring. Nel nostro caso la scelta da considerare avviene in maniera aprioristica rispetto alla stima dei parametri, tentando di selezionare le variabili che assumono maggiore rilevanza in ottica di valutazione complessiva di un'impresa.

Tenuto conto dei problemi di *overfitting* e multicollinearità già ampiamente trattati, la scelta ricade dunque su un numero esiguo di variabili rappresentanti i vari profili della gestione aziendale.

²⁹ La tabella è stata autonomamente costruita in base ai dati di bilancio disponibili

La redditività aziendale è probabilmente il profilo più considerato nelle analisi economico-finanziarie sulle imprese. Gli indicatori di redditività sono particolarmente utilizzati poiché tramite essi si riesce a osservare la capacità di un'impresa di generare reddito o produrre risorse; inoltre, possono aiutare gli investitori a prevedere ritorni economici su eventuali investimenti o possono essere utilizzati per valutare in generale l'affidabilità di un'impresa. Per tali ragioni, sono scelti come variabili esplicative due indicatori di redditività.

- ROE: il ROE (*return on equity*) è l'indicatore per eccellenza della redditività aziendale, poiché mostra una misura immediata della stessa attraverso un calcolo relativamente semplice. Esso si ottiene infatti rapportando una variabile flusso (utile) ed una stock (patrimonio netto) nel modo in cui segue:

$$ROE = \frac{\text{Utile d'esercizio}}{\text{Patrimonio netto}}$$

Solitamente è difficile assegnare valori soglia del ROE che possano permettere di giudicare in termini assoluti la redditività aziendale. Piuttosto, per stabilire se un determinato valore è positivo o negativo, conviene confrontare il ROE con investimenti alternativi a basso rischio (es. BOT). Una misura soddisfacente del ROE è solitamente superiore di 3 o 4 punti percentuali rispetto ai suddetti investimenti. Inoltre, si ricorda che la differenza fra il ROE e il rendimento *risk-free* determina il premio al rischio (ovvero il premio concesso all'investitore per scegliere di conferire capitali verso un'impresa, rispetto ad acquistare bond sicuri).

Se il ROE è negativo (sostanzialmente quando il numeratore < 0, ovvero si ha una perdita), l'azienda si trova in una situazione di squilibrio economico tale da erodere i mezzi propri.

- ROA: anche il ROA (*return on assets*) è uno degli indici di redditività aziendale più utilizzato. L'indice è calcolato come:

$$ROA = \frac{\text{Utile netto}}{\text{Totale attivo}}$$

Il ROA è un indicatore di efficienza dell'uso delle risorse dell'impresa per produrre utile. Esso può anche essere scomposto nel prodotto tra redditività delle vendite ((reddito

operativo – imposte) / totale vendite) e tasso di rotazione dell'attivo (totale vendite/totale attivo).

Il ROA è un indicatore molto simile al ROI (*return on investments*). Appare opportuno specificare che la scelta di utilizzare il primo indice rispetto al secondo, per certi tratti inconsueta rispetto ai tradizionali modelli di regressione logistica, trova spiegazione nella stretta correlazione esistente tra il ROE e il ROI. Il ritorno sull'equity dipende infatti in maniera diretta dal ROI, dal *leverage*, e dall'incidenza della gestione non caratteristica. Per non inserire due variabili sostanzialmente dipendenti tra loro, e incorrere dunque nel problema di multicollinearità, si è optato per la variabile ROA (discorso analogo al mancato inserimento degli indici di solidità, per i quali il più adatto era, appunto, il *leverage*).

Altro aspetto fondamentale nella gestione aziendale è quello della liquidità. Con tale termine si intende la disponibilità di mezzi di pagamento in contanti a brevissimo termine o comunque di titoli di pagamento monetizzabili in maniera immediata. I vari indici di liquidità esprimono la capacità aziendale nel fronteggiare impegni finanziari in relazione ai propri mezzi liquidi.

La variabile esplicativa utilizzata per rappresentare tale profilo gestionale è l'Indice di Liquidità corrente (anche detto indice corrente o *current ratio*). Esso è calcolato come:

$$ILC = \frac{\text{Attività a breve termine}}{\text{Passività a breve termine}}$$

Con la locuzione al nominatore si intendono le attività realizzabili (materie prime, merci e scorte) sommate a quelle disponibili (denaro in cassa e in banca, valori di pronto realizzo, crediti a breve termine); con quella al denominatore il valore dei debiti da pagare immediatamente a vista o a breve termine. Tale indice evidenzia una situazione soddisfacente se assume valore maggiore o uguale a 2, poiché si considera sufficientemente liquida un'azienda il cui attivo disponibile e realizzabile sia almeno il doppio dei debiti a breve termine. Risulta comunque difficile calcolare un ammontare standard, perché spesso i valori assunti derivano dalla politica finanziaria della singola impresa.

L'ultimo profilo aziendale considerato è quello dell'efficienza. L'indice scelto da inserire come variabile è il Tasso di Rotazione del capitale investito, anche detto ROT, che esprime la capacità aziendale nel trasformare il capitale investito in ricavi di vendita. La formula appare infatti come:

$$ROT = \frac{\text{Ricavi di vendita}}{\text{Capitale investito}}$$

Nel capitale investito rientrano tutte le forme di finanziamento (il capitale proprio e i debiti finanziari) utilizzate per gli investimenti in capitale circolante netto e in attività fisse al netto dei fondi. Solitamente, un indice ROT inferiore a 0,5 indica un'efficienza produttiva non soddisfacente, mediamente soddisfacente se compreso fra 0,5 e 1, e adeguata se è maggiore di 1.

Dalla lettura del paragrafo emerge chiaramente la presenza di una correlazione tra situazione aziendale positiva e indici considerati: all'aumentare di tutte le variabili, infatti, diminuisce la probabilità di insolvenza dell'impresa.

Concludiamo la trattazione riportando in tabella alcune misure statistiche (media, mediana, varianza e deviazione standard) dei vari indici, attraverso una divisione tra imprese sane e rischiose. Segue un breve commento sui risultati ottenuti.

		MEDIA	MEDIANA	VARIANZA	DEVIAZIONE STANDARD
ROE	IMPRESE SANE	1,13	6,76	3447,39	59,79
	IMPRESE RISCHIOSE	-7,63	-0,84	6008,39	80,43
ROA	IMPRESE SANE	1,17	0,92	2,74	1,70
	IMPRESE RISCHIOSE	0,875	0,31	1,58	1,30
ROT	IMPRESE SANE	0,72	0,42	0,71	0,86
	IMPRESE RISCHIOSE	0,25	0	0,17	0,43
ILC	IMPRESE SANE	1,57	0,96	3,31	1,86
	IMPRESE RISCHIOSE	0,89	0,52	1,25	1,16

La prima cosa che si nota è sicuramente la grande differenza tra i valori medi assunti dal ROE: le imprese sane sono infatti caratterizzate da valori mediamente positivi, mentre quelle rischiose da valori negativi (anche maggiori in valore assoluto). Ciò conferma quanto detto sulla struttura finanziaria: in caso di instabilità si può arrivare persino all'erosione di mezzi propri (l'impresa distrugge valore). Discorso simile può essere fatto analizzando il valore medio del ROA (più alto e maggiore di 1 per le imprese sane), anche se per tale indice la differenza tra imprese sane e rischiose risulta meno marcata.

Altro indice che conferma la corretta classificazione delle imprese nelle due categorie è il Tasso di Rotazione del capitale investito. Ricalcando quanto affermato precedentemente, il valore medio appare inferiore a 0,5 (0,25) per le imprese rischiose, indicando dunque scarsa efficienza produttiva; tale valore presenta un netto miglioramento, alzandosi fino a 0,72, per le imprese sane.

Infine, l'indice di liquidità corrente mostra una discreta differenza in valore assoluto per le due categorie. Pur essendo vero che l'indice non supera il menzionato valore soglia 2 per le imprese sane, tale differenza risulta comunque significativa. Ricordiamo infatti che il campione analizzato è formato da imprese operanti nel settore dell'innovazione, dove si è soliti attuare una politica finanziaria di reinvestimento, piuttosto che di detenzione di grandi quantità di liquidità in pancia all'impresa. Possiamo sostanzialmente affermare, dunque, che il campione analizzato rispetta le caratteristiche e le implicazioni degli indici scelti: in modo particolare in relazione ai valori assunti dal ROE, in misura minore ma significativa per i valori del ROT e dell'ILC.

4.3 STIMA DEL MODELLO 1

La stima del modello è effettuata attraverso l'utilizzo del software Gretl, che analizza la relazione tra gli indici finanziari scelti come variabili esplicative indipendenti e la probabilità di default dell'impresa, variabile dipendente dicotomica.

La stima dei coefficienti del modello avviene, come spiegato nella sezione metodologica, attraverso il metodo della massima verosimiglianza, individuando per ogni variabile il valore che massimizza la funzione log-verosimiglianza. La convergenza del modello, in particolare, risulta essere raggiunta alla settima iterazione.

³⁰ Fonte: propria elaborazione; indici ricavati attraverso l'utilizzo del software Excel.

I risultati ottenuti sono riportati in tabella.

Modello: Logit, usando le osservazioni 1-42

Variabile dipendente: Y

Errori standard basati sull'Hessiana

	Coefficiente	Errore Standard	Z	P-value
Costante	0,5248	0,6266	0,8375	0,4023
ILC	-0,4331	0,3097	-1,411	0,1582
ROT	-2,6321	1,3801	-1,906	0,0566
ROA	0,4114	0,4251	0,9677	0,3332
ROE	0,0011	0,0066	0,1711	0,8642

Numero di casi previsti correttamente: 34 (81%)

Test del rapporto di verosimiglianza: Chi quadro (4) = 9,51151 [0,0945]

Il modello empirico ha dunque la seguente formulazione:

$$P(Y = 1 | X_1, X_2, X_3, X_4) = \frac{e^{(0,5248 - 0,4331X_1 - 2,6321X_2 + 0,4114X_3 + 0,0011X_4)}}{1 + e^{(0,5248 - 0,4331X_1 - 2,6321X_2 + 0,4114X_3 + 0,0011X_4)}}$$

Dove:

- $X_1 = \text{ILC}$;
- $X_2 = \text{ROT}$
- $X_3 = \text{ROA}$
- $X_4 = \text{ROE}$.

Si riportano alcune considerazioni sui valori assunti dai coefficienti delle variabili.

Innanzitutto, notiamo il valore negativo dei coefficienti dell'indice di liquidità corrente e del tasso di rotazione del capitale investito. Tali valori confermano quanto affermato nel precedente paragrafo, poiché il segno negativo del coefficiente simboleggia proprio la relazione negativa tra

probabilità di default (Y) e la variabile considerata (all'aumentare del valore assunto da ILC o ROT diminuisce, in pratica, la probabilità che un'impresa sia insolvente). Inoltre, dalla comparazione dei valori assoluti dei suddetti indici, si nota come l'indice ROT abbia maggiore impatto sulla riduzione della probabilità di insolvenza.

Discorso diverso e divergente dalla trattazione teorica effettuata si può fare invece per i due indicatori di redditività (ROA e ROE), poiché il valore dei coefficienti stimati risulta positivo (e non negativo, come sarebbe lecito attendersi per le ragioni sopra elencate). Una delle spiegazioni plausibili di tale risultato trova luogo nella natura atipica del campione selezionato: le startup innovative differiscono infatti molto rispetto alle imprese tradizionali – tipicamente oggetto di analisi tramite regressione logistica - e, come è fisiologico per imprese di nuova costituzione ad alto contenuto tecnologico, possono presentare una correlazione anomala tra tali indicatori e solidità strutturale aziendale. L'incidenza di una società in perdita tra le startup innovative risulta infatti sensibilmente più elevata rispetto a quella rilevabile nell'universo delle società di capitali.³¹ Infine, si nota comunque che l'incidenza di tali parametri è relativamente bassa (valori inferiori a 1 e prossimi allo 0, soprattutto per il ROE).

Oltre ad analizzare l'incidenza del parametro sulla propensione della Y ad assumere valore 1 tramite l'analisi dei valori assoluti dei coefficienti, è possibile effettuare la stessa operazione attraverso l'analisi degli *odds ratio*. Come già affermato, infatti, si può utilizzare l'*odds ratio* per valutare l'associazione tra due variabili. Nello specifico, se il valore è unitario non esiste relazione tra la possibilità che la variabile dipendente assuma valori pari a 1 e la variabile indipendente; se il valore è maggiore di uno c'è relazione positiva; se è positivo ma minore di 1 c'è relazione negativa. I risultati degli *odds ratio* sono i seguenti (si veda la parte finale del paragrafo 3.1 per la formula completa):

- $OR (ILC) = e^{-0,433} = 0,649$
- $OR (ROT) = e^{-2,632} = 0,0719$
- $OR (ROA) = e^{0,411} = 1,5$
- $OR (ROE) = e^{0,001} = 1,001$

³¹ Tremolada L., *I nuovi numeri sull'ecosistema startup. Bene ma non benissimo*, Il sole 24 ore, Gennaio 2018

I risultati confermano quanto affermato poc'anzi: la presenza di una relazione negativa tra le variabili ILC e ROT e la probabilità di insolvenza dell'impresa, la relazione positiva tra gli indici di redditività (ed in particolare il valore molto vicino all'1 del ROE, ad indicarne la poca incidenza).

Completiamo infine la trattazione sui valori dei parametri presentando gli intervalli di confidenza per ognuno di essi. Spesso è infatti insufficiente presentare la sola individuazione del parametro e risulta opportuno specificare un intervallo di valori plausibili (che viene definito appunto intervallo di confidenza o intervallo di fiducia) per lo stesso.

VARIABILE	COEFFICIENTE	INTERVALLO DI CONFIDENZA	
COSTANTE	0,5248	-0,7033	1,7529
ILC	-0,4331	-1,0347	0,1684
ROT	-2,632	-5,3385	0,0741
ROA	0,4113	-0,4219	1,2446
ROE	0,0011	-0,0119	0,0141

4.3.1 VERIFICA DELLA BONTÀ DI ADATTAMENTO DEL MODELLO

La verifica della bontà di adattamento del modello avviene osservando il valore assunto dall'indice G_M (anche detto Chi-quadrato)

Nello specifico, utilizziamo G_M per valutare la capacità del modello di migliorare la previsione della variabile Y considerando l'ipotesi H_1 (ovvero prendendo come valore di riferimento quello stimato dal modello) o l'ipotesi nulla H_0 (considerando il valore medio di Y).

Per quanto già spiegato nel paragrafo 3.5.2, nel nostro caso possiamo rifiutare l'ipotesi nulla ed accettare l'ipotesi alternativa H_1 : il valore di G_M risulta infatti pari a 0,0495. La previsione che Y sia uguale a 1 è dunque migliorabile considerando i predittori.

È bene notare, tuttavia, che il valore considerato è comunque molto vicino al valore soglia di 0,05: ciò potrebbe portare a considerare che la bontà di adattamento sia relativamente buona e che possa essere migliorata (come nel modello 2 riportato in seguito).

4.3.2 VERIFICA DELLA CAPACITÀ PREDITTIVA DEL MODELLO

Per verificare la capacità predittiva del modello ci serviamo della seguente tabella dove sono presentati gli output del modello di regressione logistica.

		Previsto	
		0	1
Effettivo	0	25	3
	1	5	9

Dalla tabella notiamo che dei 42 casi totali previsti dal modello, i casi stimati correttamente sono 34 (25 + 9, ricavabili dalla diagonale principale della tabella). I casi stimati in maniera errata sono invece 8 (3+5, ricavabili dalla diagonale secondaria).

Seguendo l'impostazione del paragrafo 1.2.1, possiamo interpretare dunque il numero 3 come il numero di falsi positivi ed il numero 5 come il numero di falsi negativi. Inoltre, ricaviamo anche la misura di specificità e sensibilità del modello:

- Sensibilità: $VP/(VP + FN) = 25 / (25+5) = 0,833$
- Specificità: $VN/(FP + VN) = 9 / (9 + 3) = 0,75$

Infine, calcoliamo l'accuratezza del modello attraverso un calcolo relativamente semplice (rapporto tra numero di imprese stimate correttamente e numero di imprese totali). Il numero, se analizzato in relazione al valore massimo 1, appare abbastanza elevato, proprio a simboleggiare una buona capacità predittiva del modello.

$$Model\ accuracy = \frac{34}{42} = 81\%$$

4.3.3 VALUTAZIONE DELLA SIGNIFICATIVITÀ DEI SINGOLI REGRESSORI

La valutazione del contributo dei singoli parametri e la rilevazione della relativa significatività risultano quanto mai importanti in un modello di scoring.

Nella sezione metodologica si è menzionato come la statistica test maggiormente utilizzata per valutare il contributo di ogni predittore fosse probabilmente la statistica di Wald. Tuttavia, esistono modi differenti per valutare la significatività del singolo regressore: su tutti, uno dei metodi più facili ed intuitivi è sicuramente il confronto tra il p-value assunto dalla variabile indipendente e un livello di significatività stabilito ex-ante. Sono dunque verificabili due situazioni:

- Il p-value della variabile è maggiore del valore soglia scelto: in questo caso la variabile esplicativa non risulta significativa, ed è dunque poco utile ai fini dell'analisi empirica (equivale a dire che i parametri della regressione valgono zero);
- Il p-value della variabile è minore del valore soglia scelto: in tal caso la variabile è statisticamente significativa.

Appare opportuno procedere all'analisi della significatività delle variabili indipendenti del modello studiato. Scegliendo come livello di significatività un valore α (a due code) pari a 0,10, le variabili omesse dal modello sono:

- ROE, p-value = 0,864;
- ROA, p-value = 0,311;
- ILC, p-value = 0,119.

Tutti i valori risultano essere maggiori di α , e pertanto le relative variabili sono poco significative ed omesse dal modello. Solamente il tasso di rotazione del capitale investito risulta essere statisticamente significativo in relazione alla probabilità di insolvenza di un'impresa.

Tali risultati ci portano ad analizzare un secondo modello, che è un modello ridotto formato a partire dal modello 1 in seguito all'omissione delle 3 variabili per le ragioni sopra elencate.

4.4 STIMA DEL MODELLO 2

Il modello 2 prende in considerazione l'unica variabile statisticamente significativa rimasta dopo la valutazione dei regressori sulla base del metodo di confronto tra p-value e livello di significatività fissato. Operando in tale maniera, non si sta facendo altro che applicare la procedura *backward* precedentemente analizzata: si parte da un modello saturo e si utilizza il *driver* p-value per l'eliminazione progressiva dei parametri.

Il modello, che presenta dunque come variabile esplicativa indipendente solamente l'indice ROT, viene stimato anche questa volta tramite Gretl. Il software restituisce come output:

	Coefficiente	Errore standard	z	Pendenza
Costante	-0,092	0,4316	-0,2140	
ROT	-1,415	0,8184	-1,730	-0,2914

Numero di casi previsti correttamente = 28 (66,7%)

Test del rapporto di verosimiglianza: Chi-quadro (1) = 4,82682 [0,0280]

La formulazione del modello risulta dunque essere:

$$P(Y = 1 | X_2) = \frac{e^{(-0,092 - 1,415X_2)}}{1 + e^{(-0,092 - 1,415X_2)}}$$

Dove:

- $X_2 = \text{ROT}$.

Il segno del coefficiente risulta essere ancora negativo, anche se in valore assoluto inferiore rispetto al modello precedente stimato. Ciò significa sostanzialmente che l'incidenza della variabile indipendente su quella dipendente è diminuita.

Per sottolineare ciò, possiamo anche in questo caso calcolarne l'*odds-ratio*:

$$OR(ROT) = e^{-1,415} = 0,243$$

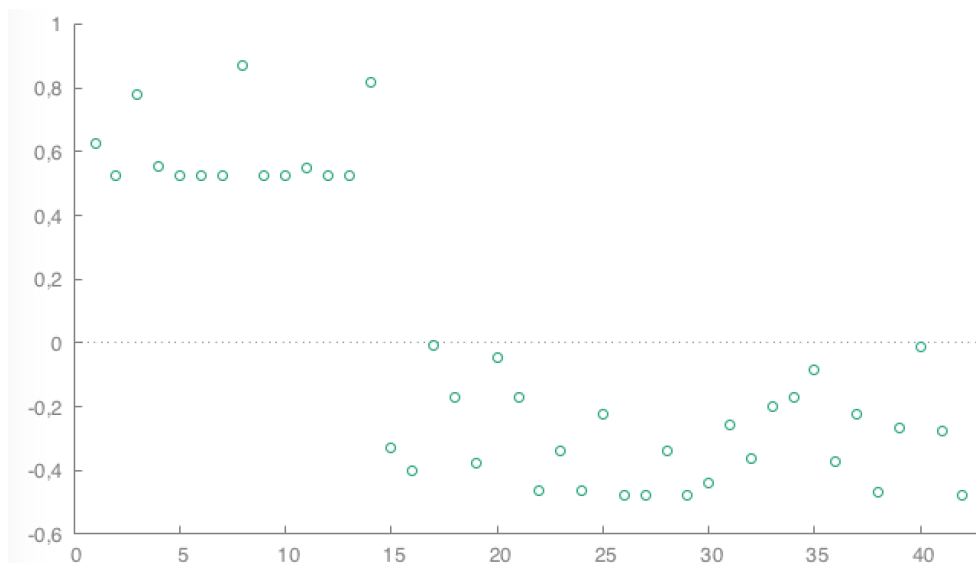
Il risultato ottenuto era facilmente prevedibile: esso è maggiore in valore assoluto rispetto al precedente (0,0719), proprio a simboleggiare la minore incidenza sulla probabilità di default, come

appena detto; allo stesso tempo l'*odds ratio* è sempre minore di uno, ed indica dunque una relazione negativa tra tasso di rotazione del capitale investito e probabilità di default della controparte.

Tale risultato conferma quanto enunciato a livello teorico riguardo l'indice ROT.

Ragionando in ottica puramente finanziaria, appare abbastanza semplice spiegare questa relazione osservando la formula analitica del ROT. L'indice viene infatti calcolato tramite il rapporto tra ricavi di vendita e capitale investito: un incremento dello stesso può essere determinato basicamente o da un aumento dei ricavi (a parità di investimenti), oppure da una diminuzione del capitale investito (a parità di ricavi), che simboleggia essenzialmente un aumento della redditività del capitale stesso. Vien da sé che entrambe le situazioni che portano il valore del tasso di rotazione del capitale investito ad alzarsi allontanano ovviamente l'impresa da una situazione di possibile default.

Seguendo l'impostazione tenuta per il modello 1, passiamo ora a verificare la bontà di adattamento del modello. In questo caso è possibile utilizzare sia il metodo grafico che quello analitico. Nello specifico, il grafico dei residui è:



Per il modello 2, la bontà di adattamento è dimostrata dalla presenza di tutti i residui nell'intervallo $[-0,6; 1]$.

Analiticamente, verifichiamo ancora l'adattamento del modello tramite il valore del Chi-quadrato (G_M). Il modello 2 risulta essere valido: G_M è statisticamente significativo, poiché il valore assunto dallo stesso è pari a 0,0280.

Considerando quanto affermato sul rifiuto dell'ipotesi nulla (quando il valore ha una $p < 0,05$) possiamo accettare il modello e specificare una validità maggiore rispetto al precedente (che non sarebbe stato accettato per valori di $\alpha < 0,03$, per esempio).

Infine, verifichiamo la capacità predittiva del modello servendoci della stessa tabella utilizzata precedentemente. In questo caso abbiamo:

		Previsto	
		0	1
Effettivo	0	28	0
	1	14	0

I casi stimati correttamente sono 28 (28 + 0, dalla diagonale principale) mentre quelli stimati in maniera errata sono invece 14 (0+14, dalla diagonale secondaria). Notiamo inoltre che si azzerava il numero di falsi positivi, ovvero di un'impresa non solvibile classificata come solvibile, mentre aumenta a 14 quello dei falsi negativi (classificazione di un'impresa potenzialmente solvibile come non solvibile).

Calcoliamo inoltre l'accuratezza del modello:

$$Model\ accuracy = \frac{28}{42} = 66,7\%$$

Notiamo che il modello risulta meno accurato del precedente, poiché l'accuratezza passa dall'81% al 66,7%. Tale valore risulta essere comunque abbastanza elevato.

Al netto di ciò, il modello risulta essere più valido, poiché è formato solamente da variabili statisticamente significative e presenta un valore Chi-quadrato inferiore.

CONCLUSIONE

Il rischio di credito risulta da sempre uno dei rischi principali dell'attività bancaria e finanziaria in generale. Considerato l'attuale periodo di crisi economica, in parte sugli strascichi della crisi finanziaria del 2008 e in parte dovuto all'emergenza sanitaria, tale affermazione risulta quanto mai adeguata.

Alcuni numeri che fanno riflettere sono quelli relativi alla possibilità di default delle piccole e medie imprese, categoria più volte menzionata nel corso dell'elaborato. Le aziende più a rischio sono infatti

quelle con un *rating* tra B e BBB (rappresentanti ben il 65% delle PMI italiane): stress test effettuati per capire l'aumento del rischio di default testimoniano che un'impresa con rating B passa dallo 0,98% al 3,29% (considerato che prima della crisi il rischio di default per rating CCC era del 2,38%).³²

In un periodo economico caratterizzato da forte incertezza e generale instabilità, in ambito creditizio appare lampante la necessità di accurata valutazione preventiva della controparte e di monitoraggio del proprio portafoglio clienti. Anche se a mio avviso i reali effetti della crisi economica si vedranno tangibilmente nei bilanci delle imprese solamente nei prossimi mesi, risulta quanto mai opportuno in un periodo come questo valutare adeguatamente l'affidabilità commerciale della controparte ed il relativo rischio di credito. Come affermato nell'introduzione, sviluppare un adeguato modello di credit scoring risulta ormai uno dei punti focali per le istituzioni finanziarie per gestire efficacemente le proprie esposizioni creditizie.

La tematica della costruzione di un solido sistema di scoring risulta dunque fortemente attuale ed è proprio da tale convinzione che nasce la scelta di dedicare il primo capitolo interamente alla relativa trattazione. Emerge chiaramente che l'adozione di tecniche di credit scoring porta ad un aumento della qualità del credito con un impatto complessivamente positivo sulla performance bancaria.

Il secondo capitolo mira invece a dare una definizione precisa di rischio di credito: analizzandone le componenti, i metodi di ponderazione e l'evoluzione delle normative (al fine di provvedere una gestione completa ed adeguata), si è cercato di evidenziarne la complessità e l'importanza nel sistema finanziario.

³² Fonte: ModeFinance

Infine, si è spiegato come taluni modelli statistici (soprattutto quelli a scelta dicotomica) possano essere utilizzati per determinare la probabilità di default della controparte. Dopo una trattazione teorica di uno di tali modelli – il modello di regressione logistica – si è passati all'applicazione pratica dello stesso ad un campione di imprese, per testarne la validità e la capacità predittiva.

In seguito ai risultati ottenuti, possiamo trarre diversi spunti interessanti sul modello Logit.

Per prima cosa, emerge in maniera abbastanza chiara come il modello di regressione logistica costituisca un ottimo strumento per la valutazione della probabilità di default, considerate le alte percentuali di accuratezza ottenute in entrambi i modelli stimati.

Risulta evidente, allo stesso tempo, come il modello presenti notevoli limiti che possono inficiare sulla relativa applicabilità. Esso presenta, infatti, una forte componente di discrezionalità, presente in diverse fasi del modello. Innanzitutto, nella scelta dell'indicatore economico-finanziario da porre alla base della suddivisione delle imprese in “sane” e “rischiose”: tale classificazione può risultare infatti completamente differente in base alla scelta di un'altra variabile discriminante. Inoltre, anche il valore del *cut-off* risulta determinante: in base al valore fissato si incide fortemente sul numero di imprese presenti in una o nell'altra categoria.

In secondo luogo, come è risultato evidente nel quarto capitolo, forte discrezionalità caratterizza anche la scelta delle imprese da includere nel campione: i risultati del modello (soprattutto nella stima dei coefficienti) possono dipendere molto dalla tipologia di imprese considerate. Risulta pertanto opportuno limitare la selezione ad un gruppo di imprese che presentino caratteristiche omogenee, soprattutto rispetto al settore di appartenenza.

Infine, si menziona che anche la scelta delle variabili esplicative, che come detto è discrezionale e aprioristica alla stima del modello, risulta una componente importante nella determinazione dei risultati finali e che gli stessi sono fortemente legati a tali scelte.

In conclusione, possiamo affermare che, al netto di tali limiti, il modello *Logit* è tra gli approcci dominanti in tale settore e risulta avere ancora elevata valenza pratica. Tuttavia, facendo un discorso più generale riguardo tutti metodi statistici per il credit scoring, è probabilmente più opportuno considerarli come punto di partenza per l'analisi del merito creditizio, unendo poi i risultati ottenuti ad altre informazioni raccolte utilizzando tecnologie aggiuntive.

BIBLIOGRAFIA

Albareto G., Felici R., Sette E., *Le tecniche di credit scoring migliorano la selezione della clientela e la qualità del credito?*, Banca d'Italia, Ottobre 2016

Allen N. Berger, W. Scott Frame, *Small Business Credit Scoring and Credit Availability*, Atlanta, Federal Reserve Bank of Atlanta, Maggio 2005

Altman I. Edward, *Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy*, Wiley for the American Finance Association, Settembre 1968

Bandyopadhyay A., Predicting probability of default of Indian corporates bond: logistic and Z-score model approach, *The Journal of Risk Finance*, Maggio 2006

Brealey R.A., Myers S.C., Allen F., *Principles of corporate finance*, McGraw-Hill Education, 11th Global Edition, 2014, p.73

Camera di Commercio Industria Artigianato e Agricoltura di Macerata, *I metodi di ponderazione del rischio*

Comitato di Basilea sulla supervisione bancaria, *An Explanatory Note on Basel II IRB Risk Weight Functions*, Bank for international Settlements, 2005

Cuicui L., Desheng W., Dexiang W., *A deep learning approach for credit scoring using credit default swap*, Stockholm, Stockholm Business School, Stockholm University, 2017

Duca P., *Analisi della curva ROC: significato dell'area ad essa sottesa e sua stima non parametrica*, Milano, Istituto di Statistica e Biometria, 1995

European Banking Authority, *Orientamenti sulla stima della probabilità di default (PD) e delle perdite in caso di default (LGD) e sul trattamento delle esposizioni in stato di default*, Aprile 2018

Fisher R., *The use of multiple measurements in taxonomic problems*, Annali dell'Eugenetica, Settembre 1936

Gallo R., *The impact of the IRB approach on the relationship between the cost of credit for public companies and financial market conditions*, Banca d'Italia, Luglio 2020

Hajage D., Tubach F., De Rycke Y., *Overfitting in propensity score model*, Journal of Clinical Epidemiology, Volume 88, 2017

Kenneth E. Train, *Metodi di scelta discreta con simulazione*, Università della California, Berkeley, Cambridge University Press, 2009

Marques Al, Garcia V., Sanchez JT, *A literature review on the application of evolutionary computing to credit scoring*, The Journal of the Operational Resear Society, Settembre 2013

Mester Loretta J., *What's the point of credit scoring?*, Philadelphia, Federal Reserve Bank of Philadelphia, 1997

Mishkin F., Eakins S., Beccalli E., *Istituzioni e mercati finanziari*, Pearson, 2019

Orgler E. Y., *A Credit Scoring Model for Commercial Loans*, Journal of Money, Credit and Banking, Ohio State University Press, Novembre 1970

Piccolo F., *Modello di regressione logistica*, Università degli studi di Napoli Federico II

Resti A., *Misurare e gestire il rischio di credito: una guida metodologica*, Milano, Alpha Test, 2001

Resti A., Sironi A., *Rischio e Valore nelle banche – misura, regolamentazione, gestione*, Egea, Agosto 2008

- Rocchi M., *La valutazione di un esame diagnostico mediante curve ROC: alcune osservazioni*, Istituto di Biomedica, facoltà di Farmacia, Università di Urbino, 2001
- Schuermann T., "What we do know about loss given default", New York, Federal Reserve Bank of New York, Febbraio 2004.
- Senese V. P., *Regressione Multipla e Regressione Logistica: concetti introduttivi ed esempi*, Ottobre 2014
- Sheldon M. Ross, *Probabilità e statistica per l'ingegneria e le scienze*, Trento, Apogeo, 2003,
- Signorini L. F., *Banche e imprese nella crisi*, Banca d'Italia, XLIV Giornata del credito, Ottobre 2012
- Stanghellini E., *Introduzione ai metodi statistici per il credit scoring*, Milano, Springer Italia, 2009
- Tremolada L., *I nuovi numeri sull'ecosistema startup. Bene ma non benissimo*, Il sole 24 ore, Gennaio 2018
- Wendel C., Harvey M., *SME Credit Scoring: Key Initiatives, Opportunities, and Issues*, The World Bank Group, Marzo 2006

SITOGRAFIA

<http://bankpedia.org/index.php/it/90-italian/c/19550-credit-scoring>

http://www.bankpedia.org/termine.php?c_id=20585#:~:text=L'indice%20di%20liquidit%C3%A0%20totale,e%20debiti%20a%20breve%20termine.&text=L'indice%20di%20liquidit%C3%A0%20totale%20%C3%A8%20uno%20degli%20indici%20pi%C3%B9,dalle%20situazioni%20patrimoniali%20d'impresa.

<https://www.borsaitaliana.it/notizie/sotto-la->

[lente/roe.htm#:~:text=Il%20ROE%20\(Return%20On%20Equity,%2FCapitale%20Proprio%20*%20100\).&text=In%20caso%20di%20perdita%2C%20il,da%20erodere%20i%20mezzi%20propri.](https://www.borsaitaliana.it/notizie/sotto-la-lente/roe.htm#:~:text=Il%20ROE%20(Return%20On%20Equity,%2FCapitale%20Proprio%20*%20100).&text=In%20caso%20di%20perdita%2C%20il,da%20erodere%20i%20mezzi%20propri.)

<https://www.cloudfinance.it/rot-rotazione-del-capitale-investito-netto.html>

https://learn.luiss.it/pluginfile.php/791619/mod_resource/content/1/2020_08%20Gestione%20dei%20rischi.pdf

<https://www.mise.gov.it/index.php/it/impresa/competitivita-e-nuove-imprese/start-up-innovative#agevolazioni>

https://www.treccani.it/enciclopedia/multicollinearita_%28Dizionario-delle-Scienze-Fisiche%29/#:~:text=multicollinearit%C3%A0%20%5BComp.,impossibile%2C%20individuare%20o%20anche%20semplic.