POLITECNICO DI TORINO

Collegio di Ingegneria Gestionale – Classe LM-31

Corso di Laurea Magistrale in Ingegneria Gestionale

Tesi di Laurea di II Livello

Sviluppo di modelli di credit scoring nel settore tessile

Relatore: Candidato :

Prof. Franco Varetto Amedeo Donà

Dicembre 2019

Sommario

1 IL SETTORE TESSILE IN ITALIA 7

1.1 CENNI SUL PROCESSO PRODUTTIVO TESSILE 10
1.2 IL BILANCIO PRECONSUNTIVO DEL 2018 NELL'INDUSTRIA TESSILE ITALIANA 12
1.3 IL COMMERCIO CON L'ESTERO AL TERZO TRIMESTRE 2018 15
1.4 L'EVOLUZIONE DEL SETTORE TESSILE DAL PERIODO DI CRISI 20
1.5 IL RUOLO DELLA TESSITURA NEGLI ANNI DI CRISI 24

2 IL RISCHIO DI CREDITO 29

2.1 PERDITA ATTESA E PERDITA INATTESA 32
2.2 LE PRINCIPALI TIPOLOGIE DI RISCHIO DI CREDITO 35
2.3 FATTORI CHE INFLUENZANO O DETERMINANO L'INSOLVENZA 37
2.4 I MODELLI DI CREIDT-SCORING 38
2.5 UN ESEMPIO DI VALUTAZIONE DEL RISCHIO DI CREDITO TRAMITE L'ANALISI DISCRIMINANTE: IL MODELLO DI ALTMAN (1968) 43
2.6 DALLO SCORE ALLA PROBABILITÀ DI DEFAULT 45
2.7 ALCUNE IPOTESI SOTTOSTANTI L'ANALISI DISCRIMINANTE 46
2.8 I MODELLI DI REGRESSIONE 47
2.9 MODELLI INDUTTIVI 50
2.10 UTILIZZO, LIMITI E PROBLEMATICHE RELATIVE ALL'IMPIEGO DEI MODELLI DI CREDIT-SCORING 55

3 LA REGRESSIONE LOGISTICA 57

3.1 INTRODUZIONE 57
3.2 PRINCIPIO DI FUNZIONAMENTO 58
3.3 TEST SULLA REGRESSIONE LOGISTICA 64
3.4 UN ESEMPIO DI REGRESSIONE LOGISTICA: IL MODELLO DI OHLSON (1980) 67

4 ANALISI ED ELABORAZIONE DEI DATI 70

4.1 SCELTA DEL SETTORE 70
4.2 IL CAMPIONE DI IMPRESE 73
4.3 DESCRIZIONE DELLE PROCEDURE DI ELABORAZIONI DATI 76
4.3.1 INSERIMENTO DATI 76
4.3.2 VALUTAZIONE DEGLI ERRORI NEI DATI 77
4.3.3 ELABORAZIONE DEI DATI 78
COSTRUZIONE DEI MODELLI DI SCORING 80

5 ILLUSTRAZIONE DEI MODELLI DI SCORING 85

5.1 I MODELLO SOCIETÀ 85
5.2 II MODELLO SOCIETÀ 87
5.3 III MODELLO SOCIETÀ 88
5.4 IV MODELLO SOCIETÀ 90
5.5 I MODELLO ANNO 92
5.6 II MODELLO ANNO 94
5.7 III MODELLO ANNO 95

6 ANALISI DELLE PERFORMANCE E DEGLI ERRORI DEI MODELLI 97

6.1 PERFORMANCE DEI MODELLI 97
6.2 VALUTAZIONE DEGLI ERRORI DEI MODELLI 103

6.3 ANALISI ERRORI I MODELLO SOCIETÀ **103**
6.4 ANALISI ERRORI I MODELLO ANNO **108**

CONCLUSIONI **112**

SITOGRAFIA **114**

BIBLIOGRAFIA **114**

Introduzione

In questo lavoro di tesi si analizza la rischiosità creditizia nel settore tessile italiano, in particolare nel sotto settore della tessitura. Il settore tessile ha subito, nel corso degli anni, un forte rallentamento dovuto in parte alla forte crisi globale del 2007, le cui ripercussioni a distanza di pochi mesi si sono manifestate soprattutto nella manifattura italiana e in parte alla mutazione progressiva del settore a partire dagli anni'90 del secolo scorso.

Più nel dettaglio, il primo capitolo affronta la condizione del settore tessile in Italia, fornendo una panoramica dello scorso 2018 e un confronto con gli anni successivi alla crisi del 2007, con un focus relativo alla fase di tessitura. Si accenna al processo produttivo, al ruolo del settore manifatturiero e alla sua evoluzione nel tempo.

Nel secondo capitolo si è descritto nel dettaglio il rischio di credito e i suoi fondamenti, il concetto di perdita attesa e perdita inattesa, e quali fattori siano più significativi nella valutazione dell'insolvenza. Si è inoltre descritto il funzionamento dei modelli di credit-scoring e fornito un esempio rappresentativo delle prime applicazioni di tali metodi (modello di Altman). Successivamente si sono descritti quali siano le principali procedure e metodi per la valutazione del rischio di credito, descrivendo l'analisi discriminante, i modelli di regressione e i modelli induttivi.

Nel terzo capitolo si è descritto il funzionamento della regressione logistica, tecnica utilizzata in questo lavoro di tesi, descrivendone anche i relativi test e fornendone un esempio.

Il quarto capitolo descrive invece tutti i passaggi che si sono seguiti per l'analisi ed elaborazione dei dati. In particolare si descrive la scelta del settore e il campione di imprese che ne fanno parte su cui si basa tutto il lavoro di tesi. Successivamente si è inoltre descritto in modo dettagliato quali siano stati i passaggi necessari a costruire i modelli di scoring, dall'inserimento dei dati scaricati dal database AIDA , alla sistemazione dei dati di bilancio ed infine alla regressione logistica tramite il software statistico *R* e la scelta delle variabili da utilizzare.

Successivamente si descrivono e illustrano nel dettaglio nel quinto capitolo i modelli di scoring che si sono costruititi passo dopo passo. In particolare si descrivono i quattro modelli relativi al flag società e i tre modelli relativi al flag anno. Si riporta per ogni

modello inoltre l'output statistico direttamente prodotto dal software utilizzato e un breve commento sulla significatività. Inoltre per ogni modello è stato costruito il relativo *logit.*

Si sono inoltre analizzati gli errori dei due migliori modelli società ed anno nell'ultimo capitolo. A tale scopo si sono costruiti degli istogrammi utilizzando variabili non presenti nei relativi modelli e con un basso grado di correlazione per valutare e mettere in luce al meglio la differenza tra le medie delle società sane correttamente classificate e sane erroneamente classificate, e tra i gruppi di società anomale correttamente identificate ed anomale errate.

1 Il settore tessile in Italia

Lo storico settore tessile e dell'abbigliamento rappresenta, in Italia, uno dei settori dell'industria manifatturiera più importanti. Si tratta di un settore che vanta un'antica tradizione, infatti la produzione organizzata dei primi tessuti risale addirittura al XII secolo quando la confraternita degli *Umiliati* che si occupava della lavorazione della lana, partendo dal milanese, costruì insieme ai conventi una prosperosa industria tessile che presto si diffuse in tutto il Nord Italia.

Il settore tessile in particolare si caratterizza da un'attività manifatturiera intensa al fine di lavorare e produrre fibre tessili utilizzate successivamente nella produzione di capi di abbigliamento o prodotti tecnici e tessili per comparti che variano dalla moda all'arredamento.

Il commercio delle lane del XIII secolo portò notevoli ricchezze a diverse città italiane tra cui Milano, Bologna, Firenze e Vicenza. L'industria tessile italiana ha trovato grande sviluppo in particolare nelle zone del Nord del Paese come il milanese, il biellese, il bergamasco e il vicentino dove sono ancora presenti testimonianze archeologiche della presenza delle prime forme di produzione manifatturiera.

Nel centro Italia il tessile si è sviluppato particolarmente nella zona industriale pratese che vanta il primato di maggior numero di aziende e occupati in Italia.

L'industria tessile italiana ha occupato una posizione di supremazia in Europa fin dal tardo Medioevo. Tuttavia, nel XV e nel XVI sec., in Francia e in Inghilterra vennero alla ribalta nuovi poli di produzione tessile e, un secolo più tardi, questi Paesi estesero la loro egemonia sui mercati tessili internazionali, sia nel Nord dell'Europa sia nel bacino del Mediterraneo; di conseguenza, a partire dalla metà del XVII sec. città italiane come Firenze, da lungo tempo specializzate nella produzione tessile, perderono importanza.

Ad oggi, le industrie tessili e dell'abbigliamento sono perlopiù presenti:

- In Piemonte nel distretto industriale biellese, dove la lavorazione della lana ha origini molto lontane. Il distretto rappresenta uno dei principali centri al mondo dell'industria tessile, specializzato nella produzione dei tessuti e filati in cashmere, pelo di cammello, alpaca. Nella zona si sono inoltre diffuse diverse aziende che producono macchinari destinate alla produzione dell'industria tessile;
- In Toscana, nel distretto industriale di Prato, rilevante a livello mondiale e uno dei più importanti a livello nazionale nella produzione di filati e tessuti in lana destinati sia all'abbigliamenti che all'arredamento. Si ricorda inoltre una vasta produzione di capi di abbigliamento;
- In Veneto, nel settore industriale veronese dove sono presenti molte aziende generalmente di piccole dimensione che si occupano, anche per conto terzi, della produzione di tessuti.

Con oltre 400.00 addetti, questo settore occupa circa il 12% del totale dei lavoratori appartenenti al settore manifatturiero; il suo fatturato rappresenta circa il 9% del fatturato dell'industria manifatturiera.

Per quanto riguarda la produzione di filati, il settore più importante è quello della lana. I lanifici infatti garantiscono all'Italia una posizione rilevante a livello mondiale: l'Italia è la quarta nazione al mondo per produzione di filati di lana. Rilevante è inoltre la produzione di tessuti di lana, dove l'Italia si aggiudica la seconda posizione a livello globale per produzione. È inoltre fortemente sviluppata la produzione di tessuti di lino e di cotone.

Per quanto riguarda le esportazioni, il settore tessile e dell'abbigliamento rappresenta circa il 10% del totale export italiano. A livello mondiale più del 6% delle esportazioni sono di provenienza italiana. I Paesi verso cui l'Italia esporta maggiormente sono USA, Cina, Francia e Germania.

La forza del settore tessile italiano risiede nella capacità di trovare il giusto mix tra progettazione stilistica e produzione e innovazione tecnologica.

Con il passare degli anni, il settore tessile italiano ha manifestato cambiamenti che ne hanno intaccato la natura, l'organizzazione, la redditività e le performance.

Con l'avvento della crisi del 2007 il settore ha accusato una forte flessione su diversi aspetti quali per esempio l'occupazione, lo sviluppo, il numero di produttori ed introiti. Nonostante sia tuttora considerato uno dei settori trainanti l'economia nazionale, si evidenziano all'interno del tessile mutamenti che non riconducono ancora alla buona condizione in cui versava nel periodo ante crisi. Il progressivo spostamento della produzione di fibre tessili nei paesi in via di sviluppo si è rivelata una scelta strategica dettata in particolar modo da:

- possibilità di impiego di manodopera a costo notevolmente ridotto rispetto ai salari medi italiani;
- sfruttamento di regimi fiscali più favorevoli;
- possibilità di trasferimento della produzione all'estero senza riduzione di qualità produttiva.

A questi fattori si possono inoltre elencare altre cause che hanno in parte contribuito alla crisi del settore quali per esempio il rafforzamento della posizione di alcuni paesi emergenti asiatici, sempre più competitivi e l'assestamento dei consumi europei verso valori sempre più bassi.
Tra i fattori elencati si può affermare che un rallentamento di domanda estera ha minato notevolmente la redditività del settore.

1.1 Cenni sul processo produttivo tessile

Il settore tessile si contraddistingue all'interno dei settori manifatturieri a causa della sua forte articolarità dovuta alla presenza di diverse fasi differenti tra di loro che devono necessariamente connettersi al fine di garantire un prodotto finito di qualità. Il processo produttivo si instrada selezionando le fibre da impiegare, siano esse fibre naturali, artificiali o sintetiche. Le fibre naturali sono quelle di natura animale o vegetale come seta, lana, cotone e lino. Tra le principali fibre artificiali si elencano il nylon, conosciuto a tutti. In alcuni casi, al fine di soddisfare particolari esigenza tecniche, si utilizzano anche fibre sintetiche come quelle costituite in carbonio.
Successivamente segue la fase di filatura che si contraddistingue in funzione del tipo di materia prima utilizzata al fine di produrre un filato conforme alla futura lavorazione della tessitura.

Figura 1.1 Le macro-fasi della produzione tessile

Una volta terminata la fase di filatura, il filato è pronto ad essere lavorato in tessitura. Le lavorazioni in tessitura sono fra loro molto differenti ma si possono distinguere principalmente in capi di tessuto e maglieria. In funzione del prodotto finito che si intende realizzare si impiegano diverse tecniche produttive ma soprattutto differenti macchinari. Una volta terminata la tessitura si ottiene un materiale detto "pezza", la quale subirà successive lavorazioni prima del confezionamento quali nobilitazione e finissaggio: si prevede la colorazione o la stampa della pezza in funzione della metodologia scelta per l'ideazione del prodotto finale. La tintura si può effettuare sulla rocca di filo oppure sulla pezza o ancora il prodotto già confezionato a differenza della stampa che invece si effettua sulla pezza. La fase di finissaggio svolge il compito di "fissare" le precedenti lavorazioni sul prodotto tramite l'utilizzo di processi chimici: tale fase si distingue ovviamente in funzione del tipo di fibre che si stanno trattando.

Terminata la fase di finissaggio, inizia la fase di confezionamento dove di fatto si porta alla vita il prodotto finito. Ogni diversa tipologia di prodotto finito subisce lavorazioni che lo contraddistinguono dagli altri, ciò rende questa fase molto particolare.
Tra le diverse tipologie di prodotto finito si distinguono infatti:

- capi di abbigliamento;
- tessuti tecnici;
- tessuti per arredi.

La filiera del tessile termina con la distribuzione e commercializzazione dei prodotti realizzati siano essi destinati al consumo o all'impiego industriale.
Come si evince da questa breve e sommaria descrizione della filiera tessile, le diverse fasi conducono ad un'elevata frammentazione ed eterogeneità. Poche sono le imprese integrate verticalmente che si occupano, cioè, di ogni fase del processo produttivo dalla preparazione delle fibre alla distribuzione e messa in commercio dei prodotti finiti. Tale situazione si è instaurata anche grazie all'evoluzione del settore a partire dagli anni '90 del XX secolo. Si può quindi concludere che il settore tessile si caratterizza dalla presenza di un numero molto elevato di imprese di piccole o piccolissime dimensioni rispecchiando le caratteristiche della realtà imprenditoriale italiana.

1.2 Il bilancio preconsuntivo del 2018 nell'industria tessile italiana

La tessitura "made in Italy" comprendente tessitura liniera, cotoniera, laniera, serica e a maglia secondo quanto stimato dal Centro Studi di Confindustria Moda per SMI conclude il 2018 in calo rispetto al risultato ottenuto nel 2017.

Si evidenziano in Tabella 1 le variazioni nel turnover dell'industria tessile italiana nel periodo relativo agli anni 2012-2018.

Il fatturato della tessitura al termine dell'anno 2012 ammontava a 7983 milioni di euro e ha manifestato nel corso degli anni una riduzione nel corso dell'anno successivo, per poi risalire negli anni 2014 e 2015 e infine si è manifestato un trend negativo che ha condotto i ricavi ad assestarsi intorno ai 7,8 miliardi di euro.

Tabella 1.1 L'industria della tessitura italiana (2012-2018). Fonte: Confindustria Moda su dati ISTAT e indagini interne. (Dati in milioni di Euro).

		2012	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
Fatturato		7983	7730	7985	7910	7839	7882	7859
Var. %		-	-3,2	3,3	-0,9	-0,9	0,5	-0,3
Valore della produzione		6237	6021	6173	6080	6011	6025	6051
Var. %		-	-3,5	2,5	-1,5	-1,1	0,2	0,4
Esportazioni		4374	4276	4400	4337	4294	4332	4333
Var. %		-	-2,2	2,9	-1,4	-1,0	0,6	0,3
Importazioni		1836	1912	2036	2042	2000	2003	1885
Var. %		-	4,1	6,5	0,3	-2,1	0,1	-5,9
Saldo commerciale		2538	2365	2364	2295	2294	2319	2447
Consumo apparente		3699	3656	3809	3785	3716	3706	3604
Var. %		-	-1,1	4,2	-0,6	-1,8	-0,3	-2,7
Indicatori Strutturali (%)								
Esportazioni/Fatturato		54,8	55,3	55,1	54,8	54,8	54,8	55,1

Importazioni/Consumo apparente		49,6	52,3	53,5	54,0	53,8	54,0	52,3

Il fatturato complessivo al termine dell'anno 2018 si assesterebbe, pertanto, intorno ai 7,8 miliardi di euro. La tessitura nel suo insieme rappresenta il 14,3% del fatturato totale generato dal complesso della filiera del settore Tessile-Moda
(cfr. Figura 1.2).

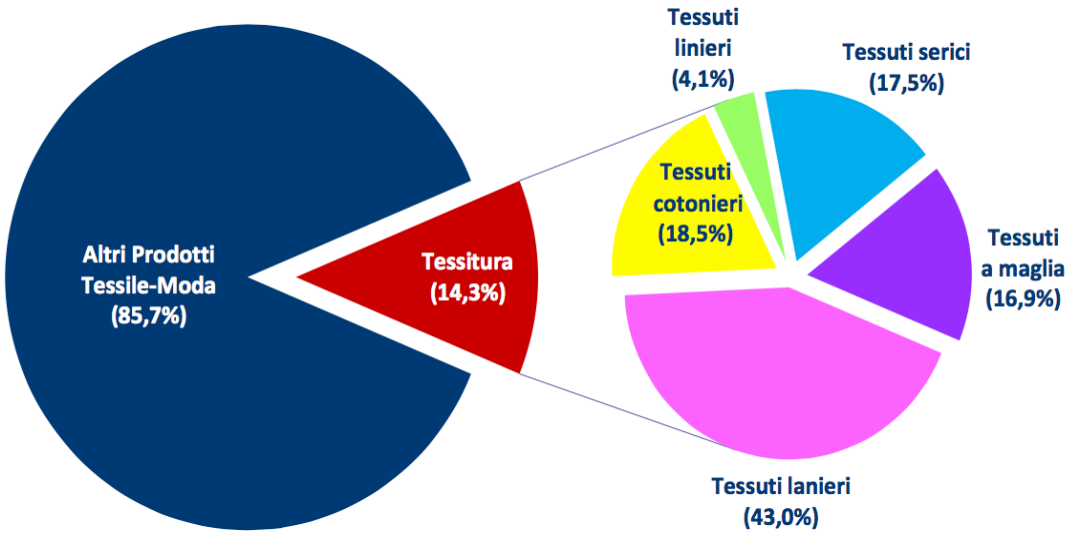


Figura 1.2 Il contributo della tessitura nell'industria Tessile-Moda italiana (2018). Dati in % sul fatturato. Fonte Confindustria Mode (Stime)

Confindustria Mode stima che la produzione nel settore della tessitura italiana al termine 2018 sarà suddivisa in tessuti lanieri che ne rappresentano la parte più consistente in termini di produzione (43,0%), seguono successivamente in ordine decrescente di produzione tessuti cotonieri (18,5%), tessuti serici (17,5%), tessuti a maglia (16,9%) ed infine la produzione di tessuti linieri rappresenta la porzione più piccola dell'intera fase di tessitura (4,1%).

Secondo i dati raccolti da *Confindustria Mode,* la tessitura serica, liniera e laniera si attende seguano un trend di sviluppo positivo occupando uno spazio sempre maggiore all'interno del settore tessitura. Per quanto concerne la tessitura cotoniera ci si attende una leggera discesa nella produzione. La tessitura a maglia si stima seguirà un trend negativo anche al termine del 2018.

Il valore della produzione si attende in aumento del 0,4%.

In riferimento agli scambi con l'estero, al termine del 2018 si nota una conclusione positiva per quanto relativo alle esportazioni extraeuropee con un aumento stimato delle vendite oltreconfine. Si evidenzia inoltre un calo significativo delle importazioni di materiale tessile necessario alle lavorazioni come l'import di semilavorati tessili.

Nell'arco dell'anno solare 2018 ci si attende infatti un aumento dell'export intorno allo 0,3% che condurrebbe il fatturato a quota 4,3 miliardi di euro, dall'altro lato le importazioni si prospettano in callo di circa il 6% dato che porta il fatturato a quota inferiore ai 2 miliardi di euro.

1.3 Il commercio con l'estero al terzo trimestre 2018

Concentrando l'analisi sugli scambi con l'estero si rileva nel commercio dei tessuti una leggera riduzione dello 0,4%, in termini di fatturato, rispetto allo stesso periodo di analisi dell'anno 2017 che aveva invece visto un timido incremento pari allo 0,3%.

L'export 2018 ammonta a 3028 milioni di euro con una "perdita" di 11 milioni di euro rispetto al 2017. In termini di volumi la variazione ammonta a -3,9%.

Contestualmente l'import dall'estero vede un ribasso notevole pari al -6,7% portandosi a 1301 milioni di euro mentre per quanto riguarda il volume la riduzione ammonta a 9,7%.

Focalizzando invece l'attenzione sulle macro aree geografiche l'export presenta diverse evoluzioni tra intra-UE ed extra-UE: le aree intra-UE rappresentano il 49,5% del volume di scambi e si notano in calo del 3,4% mentre le esportazioni verso le aree extra-UE si vedono in aumento e chiudono con una crescita del 2,9%.

Relativamente all'import, entrambe le aree si vedono in calo: gli approvvigionamenti extra-UE che rappresentano il 67,6% dell'import totale di tessuti in Italia si vedono in calo del 4,8% mentre quelli intra-comunitari che rappresentano il restante 32,4% vedono una flessione del -10,4%.

Al di là delle analisi effettuate sulle macro aree geografiche è utile analizzare anche le dinamiche che hanno interessato i singoli Paesi che ne fanno parte, principali partner commerciali delle aziende italiane di tessitura, che presentano andamenti necessari di singole analisi.

Nei primi dieci mesi del 2018 Germania e Romania, prima e seconda destinazione per volumi di traffici commerciali subiscono, rispettivamente, una riduzione degli scambi pari al -2,1% e del -5,3%; la Francia risulta invece sostanzialmente stabile e ha manifestato una leggera crescita pari allo 0,2%.

Le esportazioni aumentano ,nei primi dieci mesi del 2018 ,in Cina +3,0%, che manifesta comunque una netta frenata rispetto alla crescita a doppia cifra del 2017, mentre Hong Kong mostra un aumento del +6,1%; i 191 milioni di export venduti in Cina, se sommati ai 133 di Hong Kong vedono l'area Cina-Hong Kong risultare il primo sbocco della tessitura made in Italy, distanziando ormai ampiamente la Germania,

superandone i volumi con un ammontare di 324 milioni di euro contro i 285 milioni tedeschi (cfr. Tabella 1.2).

Si può notare, analizzando i dati relativi ai principali paesi che commercializzano tessuti *made in Italy*, la presenza di molti paesi a corto raggio particolarmente attenti e interessati da un'evoluzione favorevole dell'export nazionale.

In particolar modo Paesi come la Tunisia e la Turchia manifestano una dinamica positiva per quanto riguarda i volumi richiesti evidenziando un aumento delle importazioni italiane rispettivamente del 2,2% e del 3,1%.

Una notevole flessione piega invece il commercio con gli USA le cui esportazioni italiane verso questo paese calano del -12,6%; un altro calo degno di nota è quello che si è manifestato in Spagna con una riduzione di importazioni italiane pari al -9,8%. Portogallo inoltre assiste ad una flessione del -3,0%. Diminuiscono inoltre le vendite nel Regno Unito che manifesta un calo ripido pari al -7,5%.

Se pur i valori assoluti siano più contenuti, si è manifestato nel corso dell'anno 2018 un notevole aumento in Polonia, Giappone e Bulgaria che registrano rispettivamente un aumento del 4,9% ,8,6% e infine 7,2% per la Bulgaria.

Tabella 1.2 Commercio estero di tessuti per paese 2018. Fonte: Confindustria Moda su ISTAT

2.1) Le importazioni

Paesi di origine	Mln. di Euro	Var. %	% sul totale
Totale generale	**1 301**	***-6,7***	***100,0***
di cui:			
Intra UE28	*422*	*-10,4*	*32,4*
Extra UE28	*880*	*-4,8*	*67,6*
I primi 15 fornitori			
Cina	335	*-2,9*	*25,7*
Turchia	258	*-9,4*	*19,9*
Pakistan	106	*-3,0*	*8,1*
Repubblica Ceca	96	*2,5*	*7,4*
Germania	65	*-12,5*	*5,0*
Spagna	48	*-15,1*	*3,7*
Ungheria	39	*2,5*	*3,0*
Romania	36	*36,8*	*2,8*
Regno Unito	29	*-20,5*	*2,2*
Egitto	29	*-9,5*	*2,2*
India	27	*-14,2*	*2,1*
Giappone	26	*15,5*	*2,0*
Francia	23	*-30,1*	*1,8*
Corea del Sud	20	*-25,6*	*1,5*
Malta	15	*-15,9*	*1,2*

2.2) Le esportazioni

Paesi di destinazione	Mln. di Euro	Var. %	% sul totale
Totale generale	**3 028**	***-0,4***	***100,0***
di cui:			
Intra UE28	*1 500*	*-3,4*	*49,5*
Extra UE28	*1 528*	*2,9*	*50,5*
I primi 15 clienti			
Germania	285	*-2,1*	*9,4*
Romania	228	*-5,3*	*7,5*
Francia	210	*0,2*	*6,9*
Cina	191	*3,0*	*6,3*
Tunisia	145	*2,2*	*4,8*
Turchia	136	*3,1*	*4,5*
Hong Kong	133	*6,1*	*4,4*
Stati Uniti	128	*-12,6*	*4,2*
Spagna	124	*-9,8*	*4,1*
Portogallo	117	*-3,0*	*3,9*
Giappone	112	*8,6*	*3,7*
Bulgaria	101	*7,2*	*3,3*
Regno Unito	94	*-7,5*	*3,1*
Polonia	88	*4,9*	*2,9*
Corea del Sud	84	*0,9*	*2,8*

Concentrando l'analisi sui mercati di provenienza dei tessuti importati in Italia , si può osservare un'elevata concentrazione per i Paesi extra-UE che rappresentano circa i $^2/_3$ (67,6%) dell'intero fatturato derivante dall'import mentre il restante terzo importato ha provenienza intra-UE (32,4%).

Tra i Paesi fornitori di materie prime o materiali semilavorati, si riconfermano come nel recente passato Cina e Turchia in testa alla classifica. I primi due supplier rappresentano rispettivamente il 25,7% e il 19,9% del valore import totale: l'Italia acquista dunque in termini di valore circa la metà delle importazioni totali da Cina e Turchia (45,6%). Si evidenzia che tali primati rimangono tali nonostante vi sia stata una notevole riduzione nel periodo di analisi e raccolta dei dati (Gennaio-Ottobre 2018) che ha visto perdere il -9,4% alla Turchia e il -2,9% alla Cina.

Relativamente ai supplier con un'incidenza inferiore al 10%, il Pakistan fa registrare un calo del -3,0%, mentre si evidenziano brusche riduzioni delle importazioni provenienti dalla Germania che ha fatto registrare una riduzione a due cifre perdendo il -12,5% e una notevole e ripida riduzione pari al -15,1% derivante dalla Spagna. In controtendenza, cresce l'import da Repubblica Ceca e Ungheria dove in entrambe si sottolinea un aumento pari al 2,5% nonostante la base di calcolo in termini di valori assoluti sia nettamente differente. Infatti l'Italia importa dalla Repubblica Ceca una quantità di merce pari a un traffico di 96 milioni di euro a differenza dei soli 39 milioni di euro scambiati con l'Ungheria. Interessante notare un incremento molto ripido da parte della Romania che archivia il periodo di analisi con un positivo del +36,8% con un volume di scambi pari ai 36 milioni di euro, rappresentando il 2,8% del totale import.

Un'analisi più dettagliata si può effettuare scorporando dall'industria della tessitura le singole tipologie di tessuti che presentano comportamenti davvero differenti.

Nel periodo di analisi Gennaio-Ottobre 2018 si assiste ad un rilevante aumento delle esportazioni dei tessuti lanieri. Essi sono costituti da tessuti pettinati che registrano un incremento del +4,4% portandosi a quota 738,4 milioni di euro mentre i tessuti cardati hanno visto un aumento del +2,9% per un volume di scambio pari a 392,5 milioni di euro. I tessuti linieri registrano un incremento notevole pari al +8,6%.

Di contro, tutte le altre tipologie di tessuti accusano una flessione delle vendite estere. Arretrano del -3,2% i tessuti a maglia, abbandonando il trend positivo che le aveva caratterizzate fino allo scorso 2017. La stessa flessione si riscontra sui tessuti costituiti di cotone in cui si riscontra un calo che è pari al -3,8%; i tessuti in pura seta, se pur in calo, contengono la flessione al -0,6%.

Spostando l'analisi all'import, nel periodo da Gennaio a Ottobre '18 si evidenziano incrementi di notevole entità per il tessuto liniero che ha registrato un aumento del 16,6% e per l'importazione di tessuto di pura seta che conclude con un balzo del 21,3%. Si nota un risultato positivo anche per le importazioni di tessuto laniero dall'estero con il tessuto pettinato che ha fatto registrare un aumento del 7,5% mentre il tessuto cardato si rivela in aumento più consistente pari a +8,5%.

Al contrario, flessioni a doppia cifra interessano i flussi in ingresso sia di tessuto a

maglia, che cede il -13,3%, sia di tessuto cotoniero, in arretramento del -10,9% (cfr. Figura 1.3).

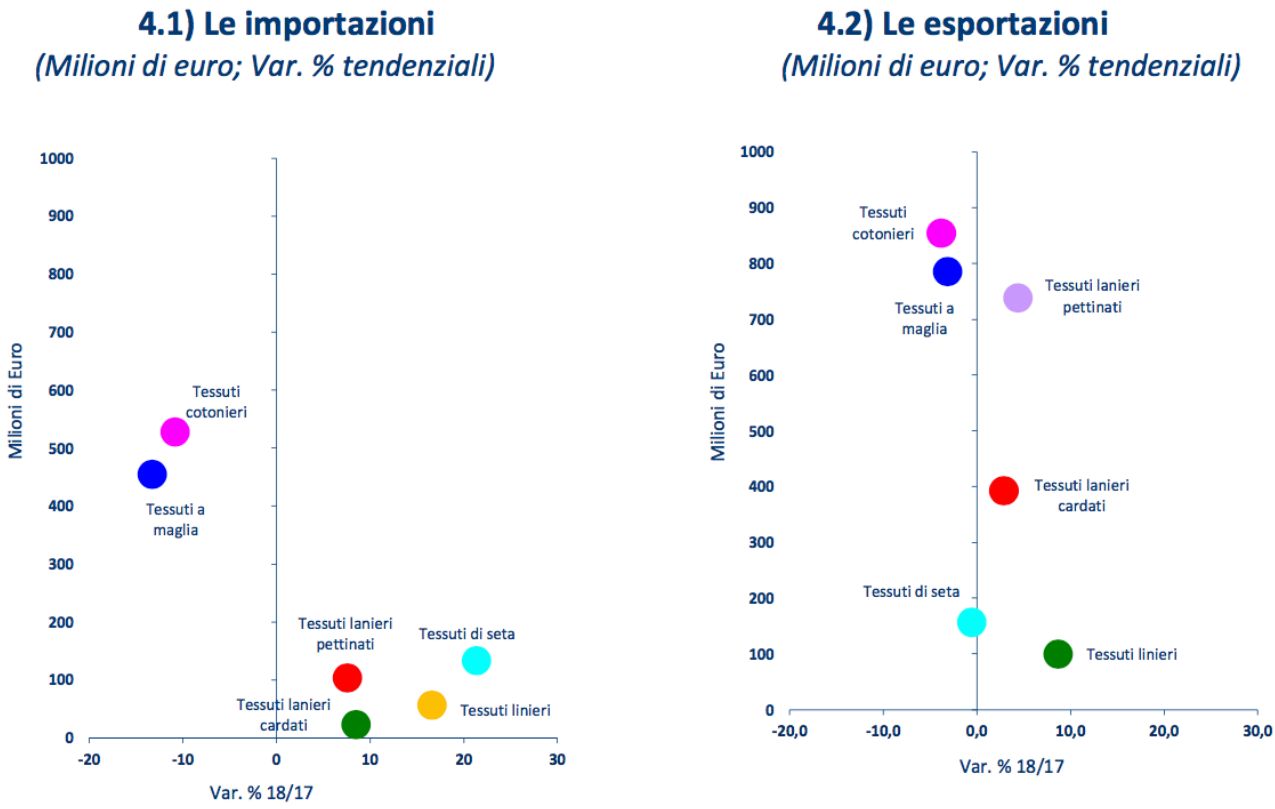


Figura 1.3 Il commercio estero sei tessuti: analisi per comparto. Fonte: Confindustria Moda su ISTAT

Concentrando l'attenzione all'anno 2019 gli analisti economici si attendono una nuova flessione che investirà anche la filiera Tessile-Abbigliamento. Nel caso della tessitura, considerando le aspettative di breve termine monitorate nell'ambito dell'Indagine Congiunturale Confindustria Moda - SMI effettuata al termine dell'anno 2018, gli operatori a campione attivi nel comparto laniero propendevano nel 57% dei casi per una prosecuzione delle condizioni di business sperimentate nell'anno precedente , mentre una quota prossima al 30% circa si attende un peggioramento .

1.4 L'evoluzione del settore tessile dal periodo di crisi

Nonostante si sia notevolmente ridimensionata rispetto alla prima metà del decennio scorso, l'industria tesile in Italia riveste tuttora un ruolo fondamentale all'interno della filiera moda-abbigliamento e rappresenta una delle più importanti realtà manifatturiere italiane.

Su scala europea l'Italia si posiziona attualmente sul podio dei principali produttori del settore tessile e vanta ancora un'elevata specializzazione in tutte le fasi della filiera con solide basi radicate nel territorio nazionale.

Si contano oggi circa 14mila imprese appartenenti al settore tessile-moda che conducono il valore della produzione a più di 20 miliardi di euro: nonostante tali valori sembrino molto importanti sono significatamene inferiori a quelli antecedenti il 2007. In confronto infatti al periodo pre-crisi si contano, in Italia, un numero di imprese nettamente inferiore con una riduzione di 4mila società, un numero di lavoratori calato di 40mila addetti e un valore della produzione inferiore di 3,5 miliardi di euro rispetto al periodo pre-crisi.

La crisi del settore che ha toccato un apice nel 2009 ha manifestato in maniera molto netta il trend al cambiamento già iniziato al termine degli anni '90 del secolo scorso.

Il baricentro dell'industria tessile si è gradualmente spostato verso quei Paesi che richiedevano un costo di manodopera molto contenuto: tali scelte strategiche combinate alla delocalizzazione della produzione a valle hanno condotto il settore tessile italiano in una condizione di business molto particolare e delicata.

Il cambiamento che ha assistito il settore ha colpito principalmente i piccoli produttori e i contoterzisti i quali sono stati facilmente rimpiazzati da aziende estere dove il costo del lavoro non è paragonabile a quello italiano. In particolare si è osservato un danno maggiore nei confronti di tutte quelle imprese che operavano nel settore o fascia del tessile "medio-bassa" in cui la necessità principale per favorire prodotti a prezzi competitivi è quella del "cost saving" ovvero la minimizzazione dei costi. Materie prime, semilavorati e costo del lavoro straniero insieme alla liberalizzazione del commercio mondiale e alle sempre più evidenti pressioni dei paesi più competitivi hanno dunque indirizzato la produzione delle aziende tessili in quella direzione.

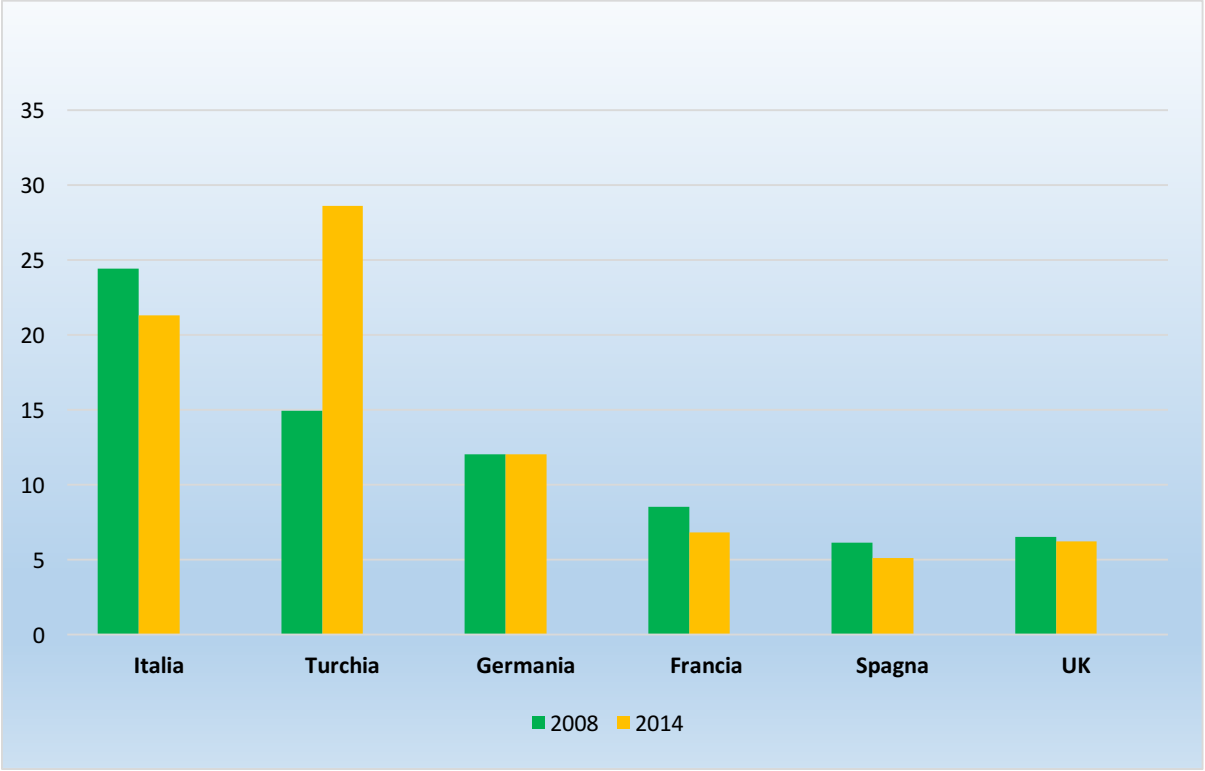


Figura 1.4 Variazione del valore del settore tessile nei principali Paesi europei, dati in miliardi di euro. Fonte: elaborazioni Prometeia su dati Eurostat.

Si può facilmente notare in figura 1.4 come il valore del settore tesile in Italia sia calato da 24,4 miliardi di euro nel 2008 ai 21,3 miliardi di euro nel 2014 (trend negativo che proseguirà anche negli anni successivi). Di notevole impatto è il comportamento della Turchia in cui si osserva un incremento nel valore portato dal settore tessile quasi raddoppiandone il valore, passando da 14,9 miliardi di euro nel 2008 ai 28,6 miliardi nel 2014, a sostegno della delocalizzazione della produzione nei paesi esteri più competitivi. Germania e Regno Unito mantengono invece una evidente stabilità nonostante gli anni di crisi; si osserva invece una flessione, se pur leggera, di Francia e Spagna.

D'altro canto molte imprese tessili si sono destreggiate in modo positivo nel periodo di crisi, sapendo analizzare e cogliere a proprio favore il mutamento in atto nel settore. Si tratta di tutte quelle aziende che si posizionano nel settore della produzione di tessuti o abbigliamento di lusso il cui business non è semplicemente indirizzato alla ricerca di un minor costo del lavoro. Queste imprese hanno reagito alla crisi offrendo, o continuando ad offrire, prodotti di alta qualità accrescendo l'integrazione lungo

tutta la filiera al fine di assicurarsi le materie prime necessarie a garantire la qualità delle lavorazioni e del prodotto finito al fine di servire e soddisfare il cliente in modo efficiente. Al contempo si mossero effettuando investimenti mirati al fine di migliorare, in particolare, tecnologie produttive e razionalizzando i processi produttivi interagendo attivamente con imprese meccano-tessili che si sono manifestate partner fondamentali all'interno della filiera produttiva. L'attenzione sul prodotto si è spostata in direzione del miglioramento della qualità delle materie prime o semilavorati utilizzati e delle prestazioni tecniche richieste dai tessuti e dai filati allo scopo di fornire una sempre più vasta possibilità di impiego e utilizzo. Proseguendo in questa direzione, e con lo sviluppo sempre maggiore di tecnologie 4.0, si sta affacciando al mondo del tessile una nuova tipologia di tessuti detti E-Textile dove si notano alcune imprese italiane impegnate nella progettazione e ricerca. Si tratta, essenzialmente, di particolati tessuti costituiti da fibre conduttive che rilevano gli stimoli esterni forniti dall'ambiente interagendo, tramite sistemi di comunicazione bluetooth per esempio, ad apparecchi elettronici quali PC o smartphone. Nonostante si tratti di un semplice prototipo vi sono già diversi possibili utilizzi in particolare a livello sportivo dove troverebbero impiego al fine di fornire uno strumento di analisi delle prestazioni e delle condizioni di salute di alcuni atleti.

Trascorsi gli anni più intensi di crisi, stabilizzando il settore con la fuoriuscita di tutti quegli operatori che non sono stati in grado di reagire prontamente alla crisi o semplicemente perché troppo deboli per sopportare la nuova fisionomia del settore, il settore tessile italiano può oggi contare su di una base più solida e notevolmente internazionalizzata che si distingue per la presenza di grandi e solidi gruppi di imprese leader, anche e soprattutto in chiave di innovazione tecnologica.

Questa situazione si è raggiunta solo al termine di un lungo percorso intrapreso a partire dalla crisi e non ancora concluso di cui, però, si iniziano ad intravedere i primi risultati. In particolare, si è evidenziata una crescita del fatturato superiore al 2% nell'anno 2017 che rappresenta la miglior performance ottenuta a partire dagli anni della crisi 2007-2009 in cui se ne è manifestata la fase più acuta.

Un'analisi più accurata al fine di valutare il tessuto produttivo tessile, si fonda sulla raccolta e valutazione degli indicatori di redditività che sintetizzano e forniscono informazioni precise e rapide all'interpretazione dei bilanci delle imprese.

Nell'ultimo biennio, il ROI medio ovvero il "Return on Investment" detto anche indice di redditività del capitale investito, calcolato come il rapporto fra risultato operativo e capitale investito netto operativo si è assestato intorno al 7%, dato superiore al valore antecedente al 2007. Tale indice di bilancio coglie l'efficienza economica della gestione caratteristica è necessario al fine di valutare la ripresa del settore.

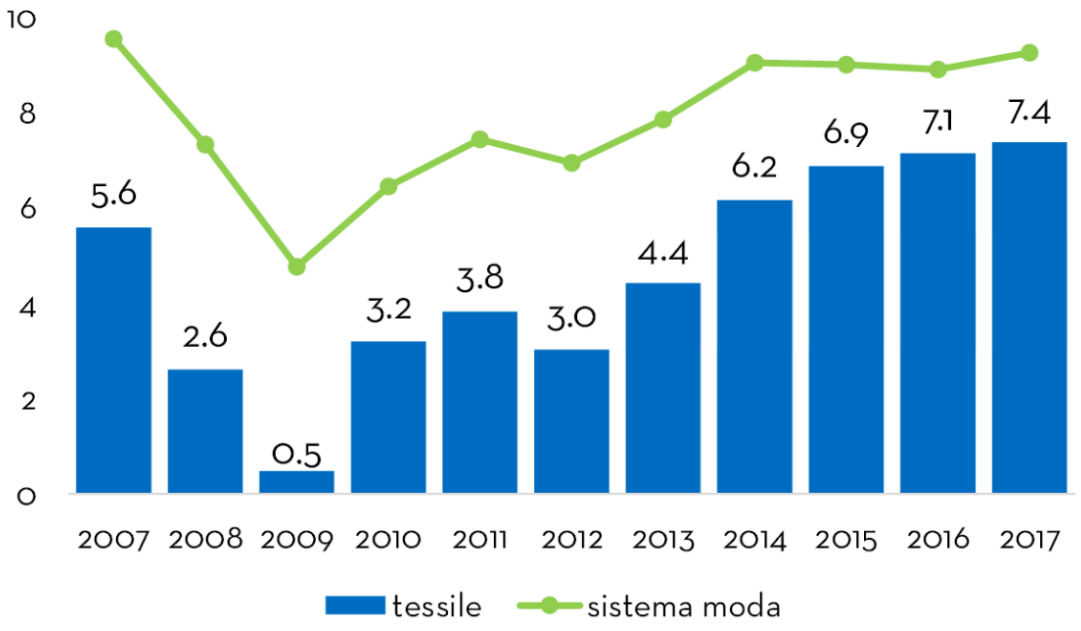


Figura 1.5 L'andamento del ROI 2007-2017 nel settore tessile e nell'intero sistema moda. Fonte: Prometeia MIO (Market Insight Outlook).

L'intero settore moda ha visto, nel corso degli anni, una netta flessione del ROI, trascinata in gran parte dal basso rendimento del settore tessile. Dopo il picco negativo raggiunto nel 2009 il cui valore di ROI ammontava a soli 0,5% per il solo settore tessile negli anni successivi si è manifestato un progressivo incremento dell'indice fino ad assestarsi introno al 7% già al termine del 2015.

1.5 Il ruolo della tessitura negli anni di crisi

Particolare attenzione si rivolge alla tessitura che, proprio come l'intero settore tessile, subisce l'impatto della crisi internazionale. Come già sottolineato, la tessitura svolge un ruolo fondamentale all'interno della filiera tessile, assolvendo il compito di preparazione e lavorazione di filati per la produzione di tessuti.

Spostando l'attenzione ai dati raccolti ed elaborati da federazione Tessile e Moda nel periodo 2005-2009 relativi all'industria della tessitura italiana si può facilmente notare come sia evidente un calo netto del fatturato. Tale andamento si manifesta a partire dall'anno 2008 e subisce una grave flessione verso il basso nel 2009 dove si verifica una riduzione sostanziale del fatturato che decresce del -22,5% come mostrato in Tabella 1.3.

Tabella 1.3 L'industria della tessitura italiana (2005-2009). Fonte: Confindustria Moda su dati ISTAT e indagini interne. (Dati in milioni di Euro).

		2005	2006	2007	2008	2009
Fatturato		8962	9099	9267	8703	6741
Var. %		-	1,5	1,8	-6,1	-22,5
Valore della produzione		7194	7264	7385	6861	5108
Var. %		-	1,0	1,7	-7,1	-25,5
Esportazioni		5760	5772	5607	5063	3714
Var. %		-	0,2	-2,9	-9,7	-26,6
Importazioni		1728	1852	1913	1795	1423
Var. %		-	7,3	3,3	-6,2	-20,7
Saldo commerciale		4033	3913	3694	3288	2291
Consumo apparente		3160	3345	3692	3593	2817
Var. %		-	5,8	10,4	-2,7	-21,6
Indicatori Strutturali (%)						
Esportazioni/Fatturato		64,3	63,4	60,5	58,2	55,1
Importazioni/Consumo apparente		54,6	55,4	51,8	49,9	50,5

Il giro d'affari diminuisce quindi passando da un valore di circa 8,9 miliardi di fatturato nel 2005 a poco più di 6,7 miliardi di euro nel 2009. Lo stesso andamento si è manifestato nel valore della produzione che, dopo un timido aumento nel 2006 e 2007, subisce un crollo del -7,1% nell'anno 2008 ed infine si osserva un picco negativo con una variazione che raggiunge il -25,5%.

Esportazioni ed importazioni manifestano anch'esse un rallentamento rispettivamente del -9,7% e -6,2% nel 2008 e un brusco calo pari al -26,6% e -20,7% nel 2009. Il saldo commerciale si chiude in calo e raggiunge nell'anno 2009 quota 2,3 miliardi di euro a fronte di un dato di partenza di 4033 milioni di euro nel 2005 (una riduzione di quasi -50%). La riduzione dei flussi esteri è solo in parte attribuibile alla riduzione della domanda estera,ne influisce anche alla riduzione di domanda interna.

Nel 2009, la tessitura italiana pesa il 14,9 % del fatturato dell'intero settore Tessile-Moda, mentre l'anno precedente la il peso ammontava al 16,8%.

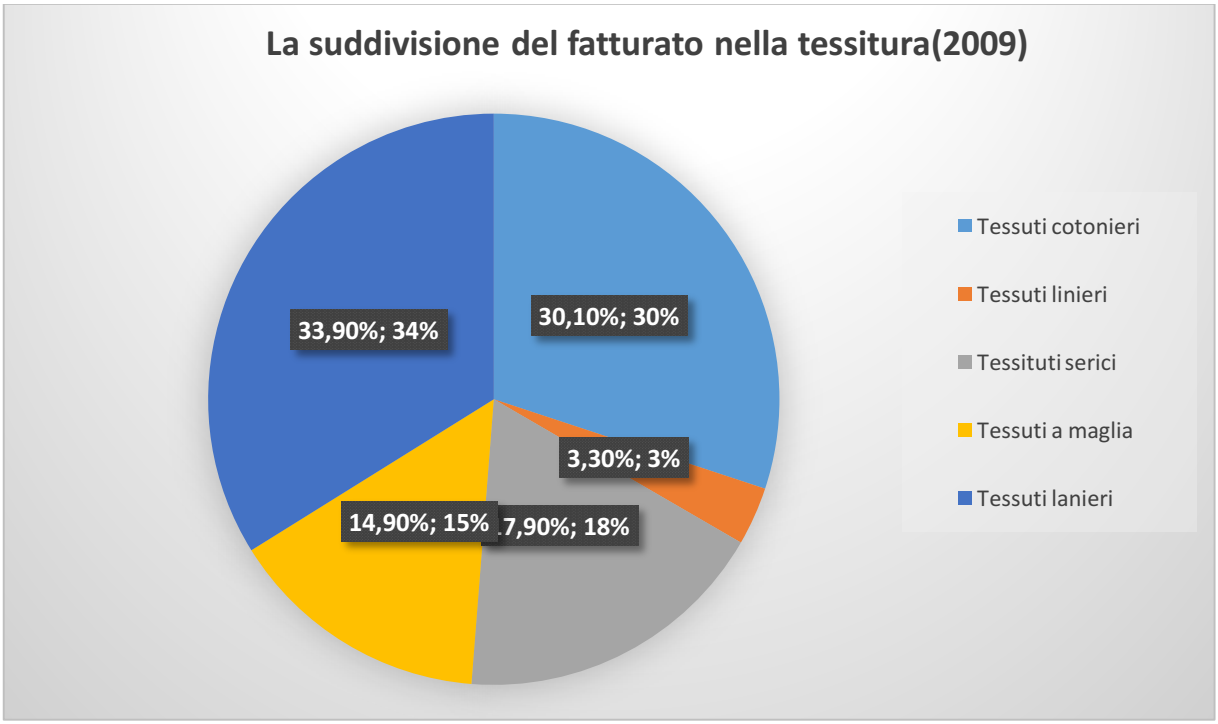


Figura1.6 La suddivisione del fatturato all'interno della tessitura.

L'incidenza sul fatturato della tessitura si suddivide successivamente in tessitura cotoniera, liniera, serica, a maglia e laniera. Si aggiudicano il primato in termini di produzione i tessuti lanieri che rappresentano circa il 34% del fatturato dell'industria;

segue successivamente la produzione di tessuti cotonieri (30%) e serica (17,9%), tessuti a maglia (14,9%) ed infine una porzione marginale si riserva ai tessuti linieri con un'incidenza sul fatturato pari al 3,3%.

Le importazioni italiane, al 2009, ammontano a 1080 milioni di euro, in calo del -21,5% rispetto l'anno precedente. Di tale cifra, il 33,1% ovvero 358 milioni di euro derivano da Paesi Intra-UE e il restante 66,9%, per un totale di 722 milioni di euro di provenienza estera. La variazione delle importazioni ha interessato in misura maggiore i paesi Intra-UE, il cui calo si registra del -31,4%, mentre per i Paesi Extra-UE la diminuzione ammonta a -15,4%.

Le esportazioni invece ammontano a 2735 milioni di euro, in calo anch'esse, con una netta flessione del -28,5%: tale somma si attribuisce per 1448 milioni di euro ai Paesi Intra UE con un peso percentuale del -27,6% e un ammontare pari a 1287 milioni di euro nei confronti dei Paesi Extra-UE con una riduzione del -29,4%.

Concentrando l'analisi in modo dettagliato per Paese, si nota, sempre sulla base dell'elaborazione dei dati Istat di Federazione Tessile e Moda relativa all'anno 2009 un forte calo dell'export pari al -28,5% in aggregato. In particolare si riscontra un calo del -27,6% del mercato intra-UE e un calo ancora maggiore del -29.4% per i Paesi extra-UE. Anche Paesi come la Cina, che in passato si dimostrarono sempre in crescita negli affari con l'Italia, subiscono una contrazione del -19,4% in termini di valori. Tra i paesi di destinazione si osservano flessioni importanti negli USA dove il calo è stato addirittura del -46,1%, seguiti poi da Hong Kong -35,2% e dalla Turchia che ha invece manifestato una performance negativa riducendo i volumi del -33,4% (Tabella 1.4).

Per quanto concerne i Paesi europei invece, si evidenziano cali consistenti sul fronte spagnolo che ha fatto registrare un -31,3% e la Romania, grande partner italiano, che ha chiuso il 2009 con un calo del -27,6%.

Tabella 1.4 Commercio estero di tessuti per paese 2009. Fonte: Confindustria Moda su ISTAT

Importazioni				Esportazioni			
Paesi di provenienza	**Milioni di Euro**	**Var. %**	**% sul totale**	**Paesi di destinazione**	**Milioni di Euro**	**Var. %**	**% sul totale**
TOTALE	**1.080**	***-21,5***	**100,0**	**TOTALE**	**2.735**	***-28,5***	**100,0**
Di cui:				*di cui:*			
Intra UE 27	358	*-31,4*	33,1	*Intra UE 27*	1.448	*-27,6*	52,9
Extra UE 27	722	*-15,4*	66,9	*Extra UE 27*	1.287	*-29,4*	47,1
I primi 10 fornitori				**I primi 10 clienti**			
Cina	309	*-9,9*	28,6	Germania	320	*-23,7*	11,7
Turchia	179	*-6,1*	16,5	Tunisia	277	*-18,9*	10.1
Pakistan	78	*-9,5*	7,2	Francia	221	*-26,7*	8,1
Germania	69	*-28,4*	6,4	Spagna	177	*-31,3*	*6,5*
Repubblica Ceca	66	*-37,0*	6,1	Romania	174	*-27,6*	6,3
Spagna	41	*-21,2*	3,8	Hong Kong	*123*	*-35,2*	*4,5*
India	37	*-25,4*	3,5	Cina	99	*-19,4*	3,6
Francia	34	*-25,8*	3,1	Stati Uniti	*99*	*-46,1*	*3,6*
Bulgaria	24	*2,2*	2,2	Turchia	96	*-33,4*	*3,5*
Tunisia	20	*28,1*	1,9	Portogallo	85	*-24,7*	*3,1*

Dall'analisi per Paese si evidenzia come nel mercato extra-UE l'import di tessuti ha registrato perdite contenute in Paesi quali Cina -9,9% a valore, Turchia - 6,1%, Pakistan -6,5%. Un dato addirittura positivo è quello evidenziato della Bulgaria che chiude con un +2,2%. Al contrario, hanno fatto registrare contrazioni di particolare intensità: Repubblica Ceca (-37%), Germania (-28,4%), Tunisia (-28,1%).

Il tendenziale e progressivo spostamento verso i Paesi extra-UE emerge anche dal confronto delle importazioni ed esportazioni di tessuti nei due mercati UE ed extra-UE nell'arco temporale compreso tra il 2006 e il 2009 (cfr. figura 1.7). Se sul fronte dell'import l'approvvigionamento verso i Paesi extra-UE registra una perdita meno pesante, il mercato Intra-UE, già a partire dagli anni precedenti, sia sul fronte dell'export che dell'import, attesta una situazione maggiormente sfavorevole e destinata ulteriormente a peggiorare.

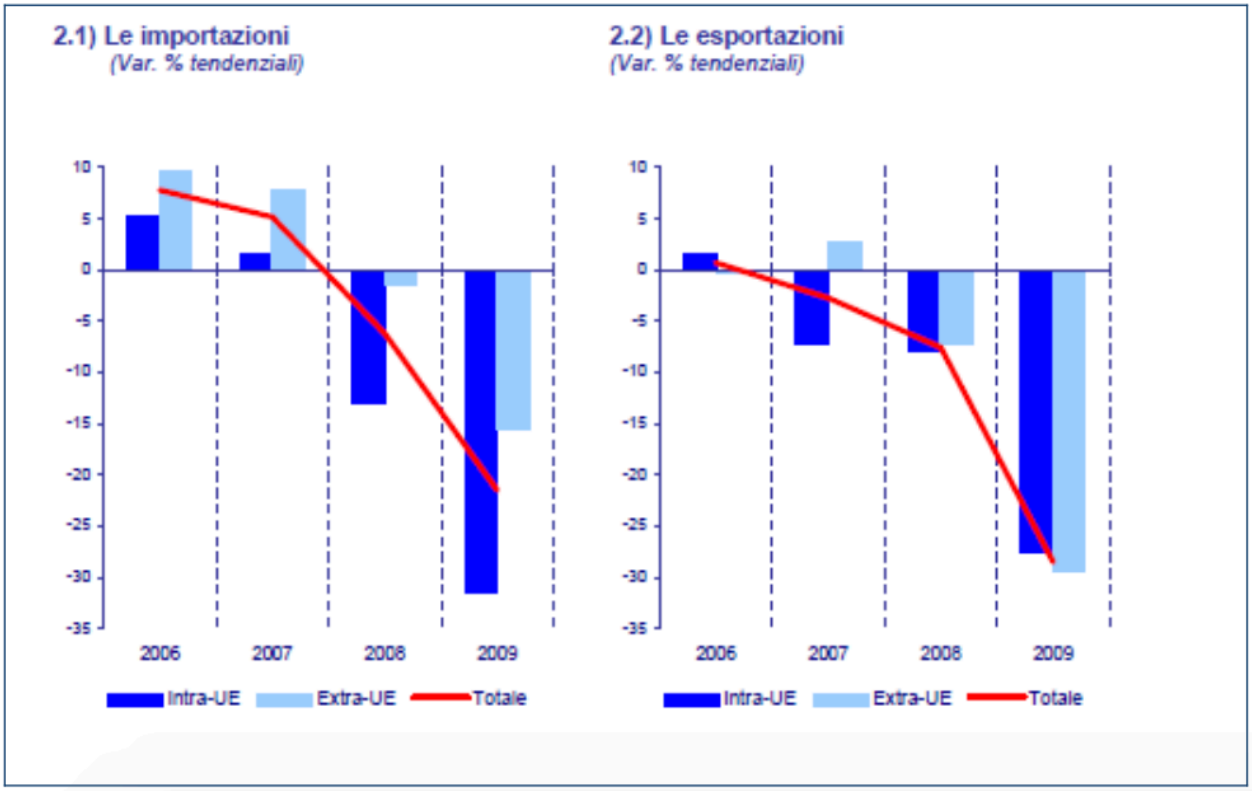


Figura 1.7 Evoluzione delle importazioni ed esportazioni 2006-2009 per i Paesi Intra-UE ed Extra UE. Fonte: Confindustria Moda su ISTAT

2 Il rischio di credito

Negli ultimi anni del XX secolo, molte banche internazionali investirono pesantemente in risorse umane e tecnologiche al fine di migliorare e riorganizzare i metodi di assegnazione, valutazione e gestione del rischio di credito. Tale processo rivoluzionario non si limita ad essere applicato ai metodi di misura del rischio, ma anche ad una delle più tradizionali aree delle banche ovvero quella dei crediti.
Si riferisce ad aspetti come:

- la selezione della controparte
- il grado di indipendenza della risk-unit di una banca
- i criteri per valutare gli obiettivi di business e le performance target
- metodi di misurazione dei risultati ottenuti
- stabilimento di incentivi
- la composizione ottimale di un portafoglio crediti.

La riorganizzazione dei processi si basa sullo sviluppo di nuovi modelli di misura del credito che hanno lo scopo di quantificare in modo appropriato il grado di rischio associato alle diverse esposizioni del credito e permette alle banche di utilizzare in modo più efficiente la capacità di assunzione del rischio.
Si può definire come *rischio di credito* la possibilità che una variazione inattesa del merito di credito di una controparte, nei confronti della quale esiste un'esposizione, generi una corrispondente variazione inattesa del valore di mercato del credito.
Tale definizione incorpora al suo interno tre concetti che necessitano di essere descritti e trattati più nel dettaglio al fine di valutare il rischio di credito in modo accurato.

1. *Rischio di default e rischio di migrazione.* Innanzitutto, il rischio di credito non si limita alla possibilità della manifestazione di un evento di default della controparte: anche un deterioramento del merito di credito costituisce una manifestazione del rischio di credito. Si consideri per esempio un prestito a tasso fisso: se si verifica un deterioramento del merito di credito, il valore di mercato del prestito, determinato dal valore attuale dei flussi di cassa ad essi

associati, sarà ovviamente ridotto. Il motivo risiede nel fatto che il valore attuale dei futuri flussi di cassa dovrebbe essere determinato utilizzando un tasso di sconto che, in aggiunta al tasso "risk-free" (ovvero un tasso di interesse a cui non è associato un rischio, pressoché certo) incorpori uno spread detto "risk-premium" che rifletta la probabilità di default del credito.

Un deterioramento del merito del credito aumenta tale probabilità, e ciò genera automaticamente un corrispondente aumento dello spread e la relativa riduzione del valore attuale del credito. Lo stesso ragionamento si può applicare ad un prestito a tasso variabile in cui lo spread pagato in aggiunta al tasso di mercato variabile si fissa in anticipo. In generale, maggiore è la variazione nello spread e maggiore è la vita residua del debito allora tanto maggiore è la riduzione dell'esposizione al credito dovuta al deterioramento dello stesso e della sua classe di rating.

Per riassumere, il rischio di credito si suddivide in due differenti tipologie il rischio di default e quello di migrazione: il primo rappresenta il rischio di una perdita dovuta all'insolvenza del debitore, mentre la seconda si manifesta come il rischio che risulta dal deterioramento di un credito che "migra" da una classe di rating ad un'altra.

Secondo questa logica, la gestione e rilevazione del rischio di credito non si baserebbe su una semplice distribuzione binomiale dei possibili eventi "default" versus "non-default" ma piuttosto su una distribuzione continua o discreta in cui l'evento di default rappresenta l'evento estremo rispetto a tutti gli eventi in cui il debitore rimane solvente.

2. *Rischio di un evento inatteso.* Un secondo aspetto implicito nella definizione di rischio di credito risiede nel fatto che per considerarsi come rischio, la variazione nel rating del credito della controparte sia non atteso.

   Se una banca ha emesso un prestito conoscendo la futura possibilità della controparte di soffrire un deterioramento nella qualità del credito in termini di profittabilità, solvenza e liquidità, saranno valutate le modalità di concessione e il relativo prezzo. Infatti gli sviluppi che si possono ritenere prevedibili nello *status* del debitore a livello economico-finanziario sono

sempre presi in considerazione nel momento in cui si stabilisce la probabilità di default e il tasso di interesse associato. Il rischio reale è rappresentato dalla possibilità che queste valutazioni si possono rivelare non corrette e si verifichi un deterioramento del credito della controparte imprevedibile a priori.

3. Esposizione creditizia. L'esposizione creditizia o "exposure" è il terzo punto che si deve considerare nell'analisi del rischio di credito. Il rischio di credito non si deve intendere come semplice rischio legato alla concessione di un prestito da parte di una banca ad un ente, azienda o soggetto privato ma bensì include anche i cosiddetti "off-balance-sheet operations" ovvero tutte quelle operazioni che riguardano un'impresa ma che non sono visibili all'interno del suo bilancio. Esempi di tali operazioni sono:

   - forme di garanzia
   - contratti che hanno come oggetto strumenti finanziari derivati scambiati OTC il cui rischio è maggiore in caso di insolvenza
   - contratti di valuta estera con liquidazione finale

   La definizione fornita in precedenza si riferisce al valore di mercato dell'esposizione di un credito che scatena l'insorgere di due problemi.
   Innanzitutto, molte esposizioni creditizie sono registrate al loro valore storico cioè il valore che esse avevano nel momento della sottoscrizione del credito e non il loro valore di mercato. Comunque, una corretta misura del rischio di credito e i suoi relativi effetti richiedono una valutazione bassata sul valore economico dell'esposizione. In secondo luogo, la maggior parte dell'esposizione creditizia di istituti finanziari si costituiscono di asset illiquidi per cui non esiste un mercato secondario; il valore di mercato può essere esclusivamente calcolato in base ad un sistema di valutazione degli asset interno all'azienda stessa.

2.1 Perdita attesa e perdita inattesa

Nell'ambito del rischio di credito è necessario distinguere in perdita attesa e perdita inattesa.

- **Perdita attesa o expected loss (EL)**

  La perdita attesa è il valor medio della distribuzione di probabilità delle perdite future. Molti studi relativi al risk management hanno in passato concentrato l'attenzione su questa componente che, come dice il nome stesso, essendo attesa non dovrebbe rappresentare rigorosamente un rischio. Nella pratica, chi concede il prestito stima ex ante il valore di tale perdita che "protegge" il suo rischio aggiungendo un adeguato spread al tasso di interesse del debito: se il debitore si manifesta insolvente come atteso, il creditore ottiene esattamente la cifra anticipata al momento dell'emissione.

  La stima della perdita attesa nei confronti di un'esposizione creditizia richiede l'analisi di tre fattori:

  i. il valore atteso dell'esposizione nel caso in cui si manifesti il default (EAD – *exposure at default*), una variabile casuale rappresentata dall'attuale esposizione sommata di una quantità pari alla possibile variazione nella dimensione del prestito dal momento in cui si calcola fino alla data di possibile default;
  ii. la probabilità che si verificherà un default nel debitore (PD);
  iii. il tasso della perdita attesa nel caso in cui si verifichi un default (LGD – *loss given default)* valutata come la percentuale dell'esposizione che la banca non sarà più in grado di recuperare nel caso si verifichi l'evento di default; si può altresì calcolare come uno meno il tasso atteso di recupero o *recovery rate* (RR) sull'esposizione.

In conclusione :

$$El = \overline{EAD} * PD * \overline{LGD}$$

- **Perdita inattesa o unexpected loss (UL)**

Il vero e proprio rischio di credito cioè il rischio che la perdita sia maggiore di quanto stimato si chiama perdita inattesa. In generale si può definire come la variabilità della perdita attorno al suo valor medio rappresentato dalla EL.

La differenza tra perdita attesa e perdita inattesa è fondamentale in tutti quei casi in cui si ha a che fare con un portafoglio di esposizioni diversificato. La perdita attesa di tale portafoglio è semplicemente calcolabile come la somma delle perdite attese dei singoli titoli che compongono tale portafoglio mentre la volatilità totale del portafoglio è generalmente minore della somma delle singole volatilità che lo compongono per via della diversificazione del portafoglio stesso che cresce al diminuire della correlazione tra i vari titoli.

In altri termini, mentre la perdita attesa non si può ridurre tramite la diversificazione per esempio costruendo un portafoglio di titoli appartenenti a settori industriali diversi o aree geografiche diverse, la perdita inattesa può essere ridotta con una strategia adatta di diversificazione, distribuendo il rischio tra i settori industriali, aree geografiche, ecc. Ciò significa che una adeguata policy di diversificazione a livello di portafoglio, lasciando inalterato il ritorno atteso totale, riduce sostanzialmente il rischio di credito.

Si noti come la differenza tra perdita attesa e perdita inattesa assume importanza anche a livello economico. Da un lato infatti la perdita attesa dovrebbe fari incrementare le riserve della banca generando un costo in termini di riduzione di profitto. Dall'altro lato invece la perdita inattesa si dovrebbe coprire con capitale appartenente alla banca, infatti come gli azionisti godono di benefici sui risultati attesi, è loro compito anche quello di coprire perdite maggiori a quella attesa con loro fondi.

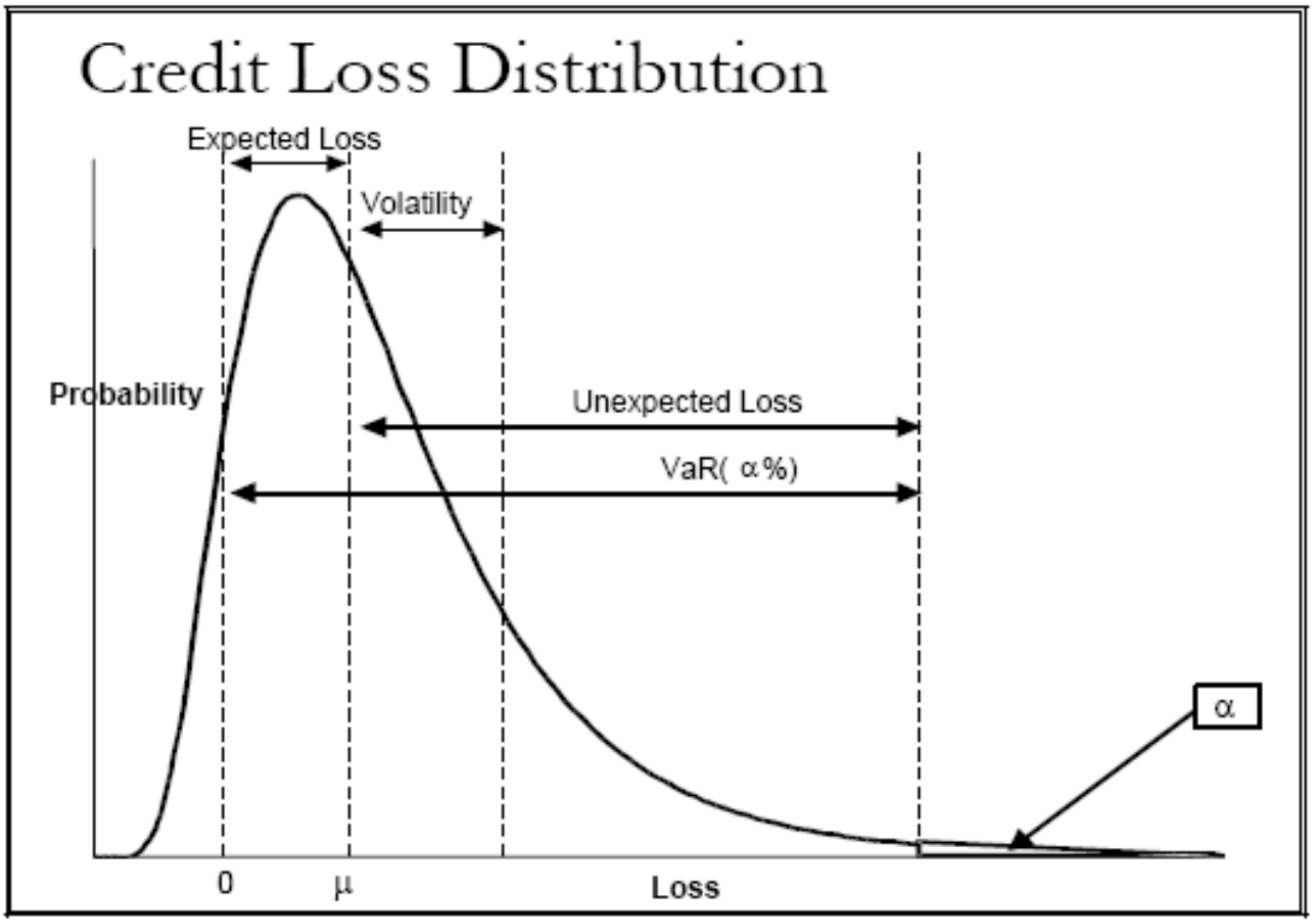


Figura 2.1 Rappresentazione grafica della perdita attesa e la perdita inattesa

2.2 Le principali tipologie di rischio di credito

Il rischio di credito comprende a sua volta i seguenti principali rischi:

- *rischio di default*: si tratta del rischio connesso al default della controparte, che dichiara bancarotta e si trova in uno stato di liquidazione o altrimenti si manifesta un default su di un prestito. In qualità di rischio, conduce ad una perdita uguale al prodotto tra "exposure at default" (EAD) e "loss given default" (LGD);
- *rischio di migrazione*: come spiegato in precedenza, tale rischio si riferisce alla possibilità che un credito si deteriori con il passando da una classe di rating ad una inferiore nel corso del tempo; si conosce anche come rischio di degradamento del credito in quanto il debitore ha un credito classificato in funzione di un rating che potrebbe essere declassato a seguito di una revisione da parte, per esempio, dell'agenzia che ha emesso tale rating;
- *rischio di spread*: si definisce rischio di spread, il rischio associato all'aumento di spread richiesto a chi emette debiti dal mercato; nel caso in cui si verifichi un aumento nel grado di avversione al rischio degli investitori, lo spread collegato ad una determinata probabilità di default potrebbe aumentare;
- *rischio di recupero*: indica il rischio che il tasso di recupero o recovery rate dopo la liquidazione degli asset della controparte, sia minore del valore inizialmente stimato per un errore nella stima dei beni liquidati oppure semplicemente perché il processo di recupero si manifesta più lungo di quanto ci si attendesse;
- *rischio di sostituzione (o pre-regolamento):* indica il rischio che la controparte con cui una banca ha firmato un contratto OTC che ha come oggetto strumenti finanziari derivati, diventi insolvente prima della scadenza del contratto: questo può condurre la banca a "rimpiazzare" tale contratto con un altro a nuove condizioni di mercato;
- *rischio di paese:* indica il rischio che una controparte estera non sia in grado di rispettare le obbligazioni e gli impegni contrattuali presi a causa di eventi

politici o di natura legislativa, come l'introduzione di vincoli di cambio esteri che impediscono il rispetto del contratto.

2.3 Fattori che influenzano o determinano l'insolvenza

Esistono variabili che si possono utilizzare e considerare rilevanti nella valutazione del rischio di credito e della possibile conseguente insolvenza:

- *liquidità relativa dell'attivo*: maggiore è il livello di liquidità e maggiori sono le disponibilità e i mezzi per rimborsare i debiti e per pagare gli oneri finanziari
- *qualità degli utili e trasparenza bilanci*: le imprese in difficoltà tendono a manomettere i bilanci al fine di figurare in salute
- *turnover dell'attivo*: più velocemente ruota l'attivo e più rapidamente si trasforma in liquidita
- *indebitamento:* maggiori sono i livelli di debito e maggiore sarà la probabilità che l'azienda sia in difficoltà
- *quota a breve termine dei default:* più è breve la durata residua dei debiti e maggiore è la probabilità che l'azienda sia in difficoltà
- *redditività e generazione di cash flow:* maggiore è la capacità di generare redditi e flussi di cassa, migliore è la situazione dell'impresa rispetto al servizio del debito
- *riserve da utili:* maggiori sono le riserve da utili reinvestite, maggiore è stata la redditività passata e maggiore è a quota di attività finanziata con risorse interne (si noti come le riserve da utili sono un indicatore della redditività passata cumulata: è una variabile che tiene conto implicitamente dell'età dell'impresa, la sua struttura finanziaria e la politica di dividendi nel tempo
- *variabilità dei redditi e dei cashflow*: maggiore è la volatilità dei redditi o la loro caduta nelle fasi negative del ciclo, maggiore è la probabilità di incontrare difficoltà future ed inattese
- *dimensione*: le grandi imprese hanno accesso ad una più ampia varietà di fonti di risorse finanziarie, possono diversificare più facilmente e, di norma, hanno maggiore potere contrattuale
- *crescita*: genera redditi e cashflow elevati ma richiede al tempo stesso forti e costanti finanziamenti

2.4 I modelli di creidt-scoring

Introduzione

Tra i modelli più ampiamente utilizzati nella valutazione e previsione dell'evento di default di una società, esiste una classe di modelli statistici generalmente noti come "crdit-scoring models". Questi sono modelli multivariati in cui si utilizzano gli stessi e principali indicatori economici e finanziari utilizzati dalle aziende come dati in input, attribuendo un peso a ciascuno di essi che riflette la relativa importanza nella valutazione del default. Il risultato di tale previsione si traduce in un indice di merito di credito espresso come un punteggio o "*score"* che indirettamente misura la probabilità di default del debitore.

Sebbene le tecniche e le metodologie che si basano sui modelli di credit scoring siano state progettate e ideate alla fine degli anni '30 del XX secolo da studiosi come Fisher (1936) e Durand (1941), l'incremento decisivo si raggiuse solo negli anni '60. Si ricordano in particolare gli studi e i risultati raggiunti da Beaver (1967), Altman (1968).

Le più importanti categorie di modelli che di credit scoring che si ricordano sono:

- analisi discriminante
- modelli di regressione (lineare, logit e probit)
- modelli induttivi euristici come le reti neurali e gli algoritmi genetici.

Analisi discriminante

L'analisi discriminante, che fu studiata da Fisher agli inizi del 1936, si basa sulla identificazioni di variabili, tipicamente economico-finanziarie calcolate a partire dai dati di bilancio, che rendono meglio possibile "discriminare" società sane da società anormali o anomale .Le società anomali si possono classificare in vari modi, per esempio società in stato di liquidazione o che presentano indicatori di bilancio al di sotto di alcuni target , sotto processi di ristrutturazione finanziaria o anche imprese per cui i debiti si classificano come "pericolosi" da parte dell'istituto di credito.

Sostanzialmente, l'analisi discriminate si basa su di una tecnica di classificazione che utilizza dati provenienti da un campione di imprese per tracciare un confine che separa le aziende affidabili da quelle insolventi. Tale separazione si basa su una *funzione discriminante.*

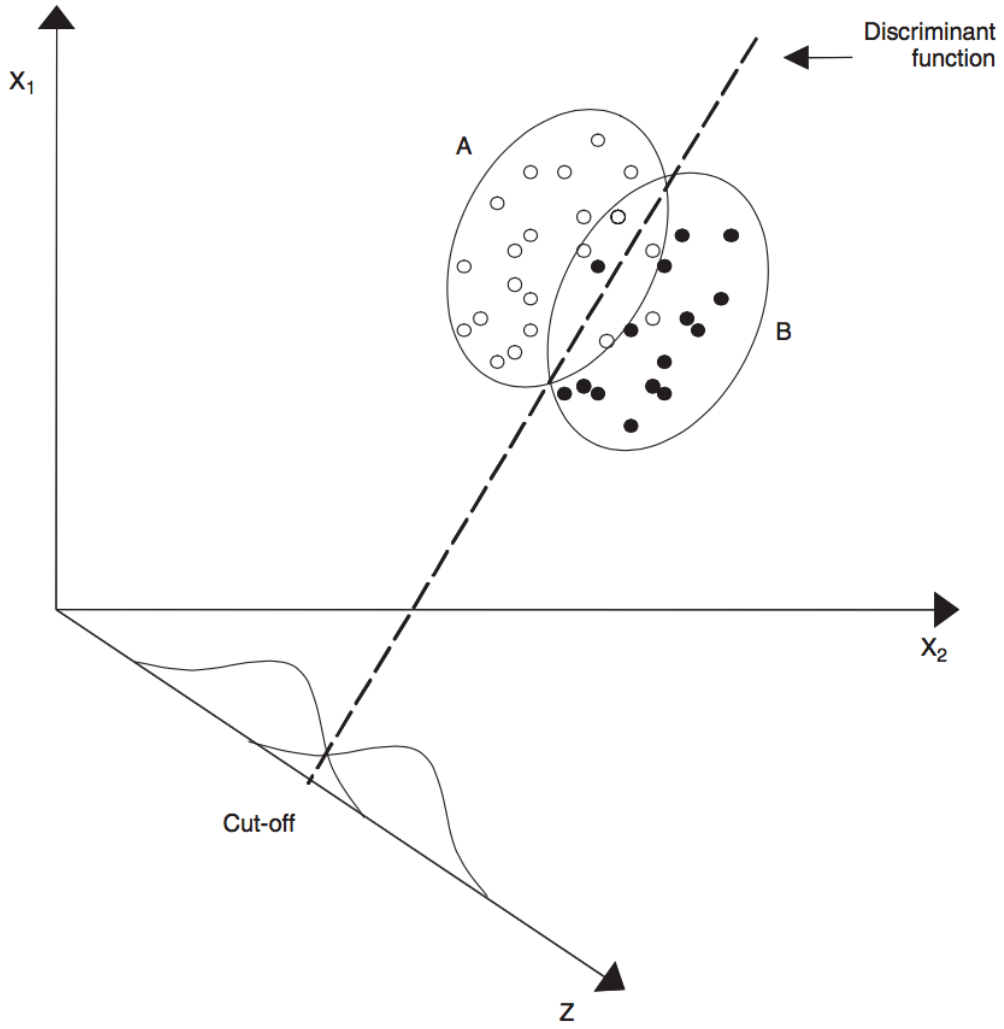


Figura 2.2 La rappresentazione grafica dell'analisi discriminante.

Come si evince dalla Figura 2.2, il modello di Fisher separa, in questo caso semplificato, imprese sane o affidabili contraddistinte dall'insieme A da imprese insolventi di tipo B, descritte da sole due variabili X_1 e X_2. Il punteggio o score generato dalla combinazione delle due variabili originali si valuta sull'asse Z.

In termini più generali, date *n* variabili indipendenti, lo score si può calcolare come:

$$z = \sum_{j=1}^{n} \gamma_j x_j$$

Equazione 2.1

Per la generica *i-esima* società, lo score si calcolerà invece come segue:

$$z_i = \sum_{j=1}^{n} \gamma_j x_{i,j}$$

Equazione 2.2

I coefficienti γ_j si scelgono al fine di separare in modo più netto, cioè di discriminare maggiormente e in modo più chiaro le società sane da quelle anomale. In altri termini, i valori di z_i devono essere tali da massimizzare le distanze tra le medie delle imprese sane e le medie ottenute alle imprese anomale.

I valori così ottenuti dalle imprese in salute saranno fra loro il più possibile simili mentre saranno molto distanti da quelli ottenuti invece dalle imprese anomale. Il fine ultimo di tale procedura è quello di costruire due gruppi differenti di società z_a e z_B.

Si può dimostrare che il concetto prima descritto è soddisfatto se i coefficienti γ sono calcolati nel seguente modo:

$$\boldsymbol{\gamma} = \boldsymbol{\Sigma}^{-1}(\mathbf{x_1} - \mathbf{x_2})$$

Equazione 2.3

dove i valori di *x* sono vettori contenenti le medie delle osservazioni delle società sane e delle società anomale e $\sum$ rappresenta invece la matrice di varianza e covarianza tra le *n* variabili indipendenti.

Si desidera costruire la funzione discriminante

$$z = \sum_{j=1}^{n} \gamma_j x_j = \boldsymbol{\gamma}'\mathbf{x}$$

Equazione 2.4

scegliendo γ in modo da massimizzare la differenza standardizzata tra le medie delle due popolazioni

$$\underset{\boldsymbol{\gamma}}{Max} \frac{|z_A - z_B|}{\sigma_z} = \frac{|\boldsymbol{\gamma}'\mathbf{x}_A - \boldsymbol{\gamma}'\mathbf{x}_B|}{\sigma_z}$$

Equazione 2.5

Al fine di lavorare con i valori assoluti, si utilizza una funzione quadrata *q* dell'Equazione 2.5 e si riscrive il problema come segue:

$$\underset{\boldsymbol{\gamma}}{Max} \frac{(\boldsymbol{\gamma}'\mathbf{x}_A - \boldsymbol{\gamma}'\mathbf{x}_B)^2}{\sigma_z^2} = \frac{(\boldsymbol{\gamma}'\mathbf{x}_A - \boldsymbol{\gamma}'\mathbf{x}_B)^2}{\boldsymbol{\gamma}'\boldsymbol{\Sigma}\boldsymbol{\gamma}} \equiv q(\boldsymbol{\gamma})$$

Equazione 2.6

Prima di procedere, si noti che la funzione in *q* è omogenea e di grado zero in γ. In altri termini si può affermare che $q(k\gamma)=q(\gamma)$ per ogni valore di *k* appartenente all'insieme dei numeri reali. Infatti:

$$q(k\boldsymbol{\gamma}) = \frac{(k\boldsymbol{\gamma}'\mathbf{x}_A - k\boldsymbol{\gamma}'\mathbf{x}_B)^2}{k\boldsymbol{\gamma}'\boldsymbol{\Sigma}k\boldsymbol{\gamma}} = \frac{k^2(\boldsymbol{\gamma}'\mathbf{x}_A - \boldsymbol{\gamma}'\mathbf{x}_B)^2}{k^2\boldsymbol{\gamma}'\boldsymbol{\Sigma}\boldsymbol{\gamma}} = q(\boldsymbol{\gamma})$$

Equazione 2.7

Dopodiché per identificare quel valore di γ che massimizza q, si calcola il gradiente di q e lo si pone uguale a 0:

$$\frac{\partial q}{\partial \boldsymbol{\gamma}} = \frac{2(\boldsymbol{\gamma}'\mathbf{x}_A - \boldsymbol{\gamma}'\mathbf{x}_B)(\mathbf{x}_A - \mathbf{x}_B)\boldsymbol{\gamma}'\boldsymbol{\Sigma}\boldsymbol{\gamma} - 2\boldsymbol{\Sigma}\boldsymbol{\gamma}(\boldsymbol{\gamma}'\mathbf{x}_A - \boldsymbol{\gamma}'\mathbf{x}_B)^2}{(\boldsymbol{\gamma}'\boldsymbol{\Sigma}\boldsymbol{\gamma})^2} = \mathbf{0}$$

Equazione 2.8

da cui si ottiene risolvendo

$$(\mathbf{x}_A - \mathbf{x}_A)\boldsymbol{\gamma}'\boldsymbol{\Sigma}\boldsymbol{\gamma} - \boldsymbol{\Sigma}\boldsymbol{\gamma}(\boldsymbol{\gamma}'\mathbf{x}_B - \boldsymbol{\gamma}'\mathbf{x}_B) = \mathbf{0}$$
$$\boldsymbol{\gamma} = \boldsymbol{\Sigma}^{-1}(\mathbf{x}_A - \mathbf{x}_B)\frac{\boldsymbol{\gamma}'\boldsymbol{\Sigma}\boldsymbol{\gamma}}{(\boldsymbol{\gamma}'\mathbf{x}_A - \boldsymbol{\gamma}'\mathbf{x}_B)}$$

Equazione 2.9

Si noti come il termine $\frac{\boldsymbol{\gamma}'\boldsymbol{\Sigma}\boldsymbol{\gamma}}{(\boldsymbol{\gamma}'\mathbf{x}_A - \boldsymbol{\gamma}'\mathbf{x}_B)}$ nell'equazione precedente sia una quantità scalare. Questa quantità può essere eliminata dalla soluzione senza alterarne il valore considerando che *q(γ)* è omogenea di grado 0. In conclusione quindi si dimostra che

$$\boldsymbol{\gamma} = \boldsymbol{\Sigma}^{-1}(\mathbf{x_1} - \mathbf{x_2})$$

l'Equazione 2.3 è verificata.

2.5 Un esempio di valutazione del rischio di credito tramite l'analisi discriminante: il modello di Altman (1968)

Il modello di scoring maggiormente conosciuto nel contesto del rischio di credito è probabilmente quello sviluppato da Edward Altman nel 1968 per le imprese quotate negli USA. Il campione di imprese che ha condotto alla stesura di tale modello si costituiva di 33 imprese sane e 33 imprese fallite nel periodo compreso fra gli anni 1946-1965. Esso si sviluppa come una funzione di cinque variabili indipendenti e si formula come segue:

$$z_i = 1{,}2 * x_{i,1} + 1{,}4 * x_{i,2} + 3{,}3 * x_{i,3} + 0{,}6 * x_{i,4} + 1{,}0 * x_{i,5}$$

dove:

$$x_1 = \frac{Capitale\ circolante}{Attivo}$$

$$x_2 = \frac{Riserva\ da\ utile}{Attivo}$$

$$x_3 = \frac{EBIT}{Attivo}$$

$$x_4 = \frac{Valore\ di\ mercatio\ di\ P.N.}{Valore\ contabile\ dei\ debiti}$$

$$x_5 = \frac{Ricavi}{Attivo}$$

Il valore della variabile discriminante *z* suggerisce la probabilità che l'azienda risulti insolvente. In questo particolare caso si può distinguere tra tre classi di valori assumibili da *z*:

a. $z < 1{,}8$: alta probabilità di insolvenza
b. $1{,}8 < z < 3$: "area grigia" ovvero imprese poco distinguibili tra anomale o sane
c. $z > 3$: bassa probabilità di insolvenza.

2.6 Dallo score alla probabilità di default

L'analisi discriminante si utilizza a volte per produrre una stima diretta della probabilità di default associata alla singola società che si analizza. Si è dimostrato che se le variabili indipendenti sono distribuite secondo una distribuzione normale multivariata, la probabilità che una società sia anomala può essere calcolata nel seguente modo:

$$PD = p(B|\mathbf{x}_i) = \frac{1}{1 + \frac{1-\pi_B}{\pi_B} e^{z_i - \alpha}}$$

dove z_i e α indicano rispettivamente il punteggio o score ottenuto dalla *i-esima* impresa e il valore del *cut-off* ovvero del limite che separa il gruppo di sane da quello delle anomale, mentre il valore di π_B la probabilità a priori di default o più semplicemente "*prior*", che non dipende strettamente dalle caratteristiche delle singole imprese ma bensì da parametri del settore o mercato.

Poiché ci si riferisce a probabilità e statistica ogni modello non sarà accurato e preciso al 100% ma presenterà degli errori.

Per bassi valori di π_B si intuisce facilmente che la probabilità di default sarà bassa. Al contrario se il valore di π_B si innalza, le probabilità di default di ogni singola osservazione aumenteranno di conseguenza.

2.7 Alcune ipotesi sottostanti l'analisi discriminante

I modelli di analisi discriminate lineari, come quello proposto da Altman, si basano sull'ipotesi che la matrice di varianza e covarianza delle variabili indipendenti (x_j) siano uguali per i due gruppi di società, sane e anomale. Nella pratica, i dati empirici spesso sembrano suggerire il contrario.

Quando si verifica un problema di *eteroschedasticità* (ovvero la varianza dei residui o errori statistici non è indipendente dalle variabili esplicative) tra le osservazioni dei gruppi si possono utilizzare versioni più efficaci dei modelli, che richiedono l'utilizzo e la stima di un numero più elevato di variabili. Se il numero di osservazioni facente parte il campione non sono sufficientemente grandi, le stime potrebbero essere imprecise, e questo comprometterebbe anche, eventualmente, i modelli teorici più sofisticati. In più, le funzioni che si basano su l'analisi discriminante quadratica sono maggiormente difficili da utilizzare nella pratica ed implementare.

Inoltre, come menzionato precedentemente, le formule che convertono lo score nella probabilità di default, assumono che le variabili indipendenti seguono una distribuzione normale multivariata; comunque alcune analisi empiriche hanno riportato che le tale ipotesi non è realistica. Infatti, come molti indicatori economico-finanziari assumono valori compresi tra 0 e 100% (come ad esempio alcuni indici di redditività) , una distribuzione che può assumere valori illimitati cioè compresi tra meno infinito e più infinito come la distribuzione normale è intuitivamente non corretta. Tuttavia, altri test empirici hanno messo in luce che l'analisi discriminante è robusta rispetto a violazioni delle due ipotesi precedenti.

2.8 I modelli di regressione

I modelli di regressioni si possono suddividere in due grandi famiglie: i modelli di regressione lineare e i modelli logit e probit o di regressione logistica.

i. I modelli di regressione lineare

In questa tipologia di modelli, la variabili utilizzate a determinare il default di una società, e i rispettivi pesi, si identificano con una regressione lineare semplice. Il modello si sviluppa e prevede 4 fasi.

1. *Selezione del campione.* Si seleziona un campione formato da un numero sufficiente di osservazioni. Come nel caso di analisi del discriminante esse sono divise in due gruppi, identificate da una variabile di stato binaria o variabile dummy che può assumere esclusivamente il valore 0 oppure 1.
   In particolare se uguale ad 1 significa che la società è anomala, mentre se la variabile assume il valore 0 significa che è sana.
2. *Selezioni di variabili indipendenti.* Per ogni società e ogni osservazione si raccolgono *n* variabili significative. Come nel caso dell'analisi del discriminato tali variabili si identificano spesso con indici di bilancio o indicatori economico-finanziari (leva finanziaria, liquidità, turnover o livelli di investimento).
3. *Stima dei coefficienti.* Si stima un modello utilizzando il metodo dei minimi quadrati ottenendo:

$$y_i = \alpha + \sum_{j=1}^{m} \beta_j x_{i,j} + \varepsilon_i$$

dove β_j indica i coefficienti della regressione.

4. *Stima della probabilità di default.* Il modello sopra descritto si applica al fine di stimare la probabilità di default di una diversa società.

ii. I modelli logit e probit

Nei modelli logit, la relazione lineare si aggiusta di una certa quantità tramite una trasformazione esponenziale, chiamata trasformazione logistica:

$$y_i = f(w_i) = \frac{1}{1 + e^{-w_i}}$$

dove le variabili indipendenti w_i si ottengono come risultato di una regressione lineare del tipo:

$$w_i = \alpha + \sum_{j=1}^{m} \beta_j x_{i,j}$$

Combinando quindi le equazioni sopra scritte e aggiungendo un termine che tenga conto della casualità dell'errore statistico associato alla regressione si ottiene:

$$y_i = \frac{1}{1 + e^{-\alpha - \sum_j \beta_j x_j}} + \varepsilon_i$$

Il range dei valori che si generano tramite una funzione logistica ovvero il codominio di tale funzione è ora limitato all'intervallo (0,1).

Questo garantisce che la variabile dipendente y_i sia sempre compresa tra 0 e 100% e può quindi correttamente interpretarsi come una probabilità di default.

Si noti che esistono trasformazioni alternative a quella logistica che forniscono sempre valori compresi tra 0 e 1. Per esempio si può

utilizzare una funzione di densità cumulata normale e costruire cosi un modello detto “normit” o, in termini più generali “probit”.

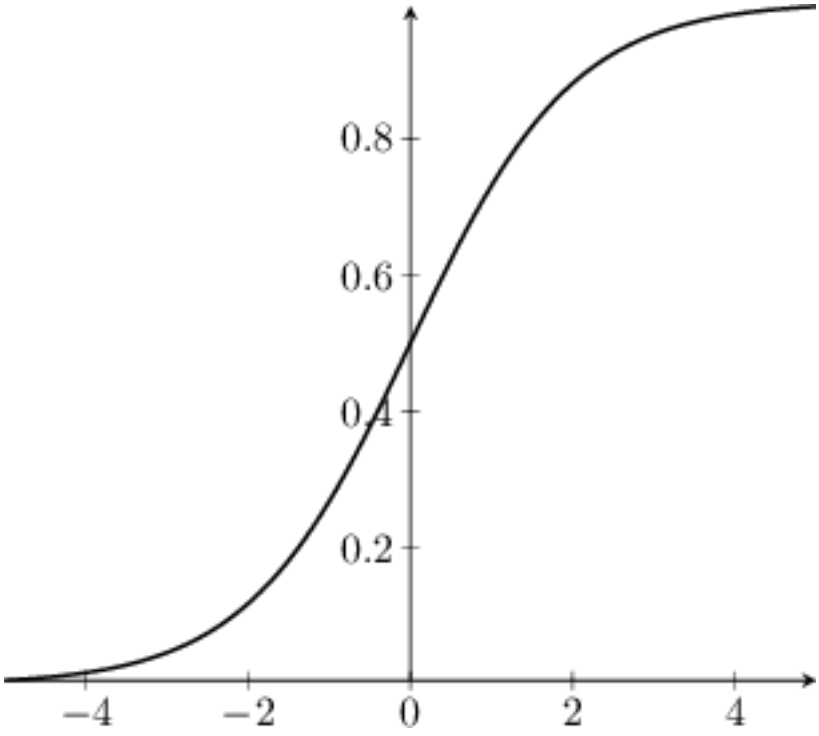


Figura 2.3 Un esempio di funzione logistica

2.9 Modelli induttivi

Reti neurali

I modelli di credit-scoring esaminati nei paragrafi precedenti hanno un denominatore comune: il tentativo di identificare la relazione fondamentale che spieghi la situazione in cui versa una società e successivamente con lo scopo di prevedere l'evento di default. In altri termini, si può affermare che tali modelli si basano su di una caratteristica strutturale che spiega lo stato di salute di un'impresa, e la selezione di variabili rilevanti e significative, scelte in modo calibrato tramite l'impiego di tecniche statistiche. Per esempio i modelli prevedono l'utilizzo di variabili come la leva finanziaria o rapporto di indebitamento oppure ancora il rapporto tra oneri finanziari e risultato operativo, perché si assume che una richiesta eccesiva di debiti oppure una mancanza di risultato operativo possano aumentare il rischio di default. Nel rispetto di tale ipotesi, tali metodi seguono un approccio già definito in precedenza strutturale: essi si basano su assunzioni come per esempio l'uso eccesivo di debito può aumentare la probabilità di default, e si cercano delle conferme di tali ipotesi su un campione di dati.

Le reti neurali utilizzano un approccio induttivo: se, partendo da un campione di dati, si trova una certa regolarità empirica come per esempio un certo valore di x_I al di sotto di un certo cut-off determina il default, tale regolarità si utilizza in seguito per prevedere futuri default di altre imprese. In questo caso, al posto di utilizzare un approccio basato su regole deduttive e determinate, si utilizza un puro metodo empirico basato sull'esperienza.

I modelli strutturali come quelli di analisi discriminante e quelli basati sulle regressioni si fondano su solidi e testati algoritmi che utilizzano test inferenziali per verificare la significatività statistica dei coefficienti. Al contrario, i modelli induttivi si paragonano a "black box" che forniscono soluzioni a problemi complessi in cui risulta difficile stabilire delle regole deduttive che governano un certo fenomeno come il default di una società.

Inoltre, nel caso del rischio di credito, alcuni modelli strutturali si possono "imparare" con gli studi progressivi del problema da parte delle imprese stesse; in altre parole le società, comprese quelle ad elevato rischio di default potrebbero risalire al metodo

(noto) di valutazione del loro stesso stato di saluto e reagire adottando alcune tecniche che possono "mascherare" la condizione in cui vertono realmente. L'influenza dei risultati potrebbe tornare utile nel momento in cui si richiede un ulteriore prestito o si è soggetti ad una valutazione esterna. Ciò a dire che in alcuni casi i modelli strutturali, nonostante siano ben costruiti, possono risultare inutili e non rappresentativi della realtà.

Le reti neurali si basano invece su approccio del tipo scatola nera citato in precedenza. Tali modelli si pongono come obiettivo quello di "mimare" le tecniche di apprendimento umane e la memoria, catturando alcuni aspetti che non possono semplicemente essere introdotti e compresi in un algoritmo di calcolo. Si parla infatti di aspetti e tecniche relative all'intelligenza artificiale.

Un esempio di rete neurale si riporta in figura 2.4 qui sotto.

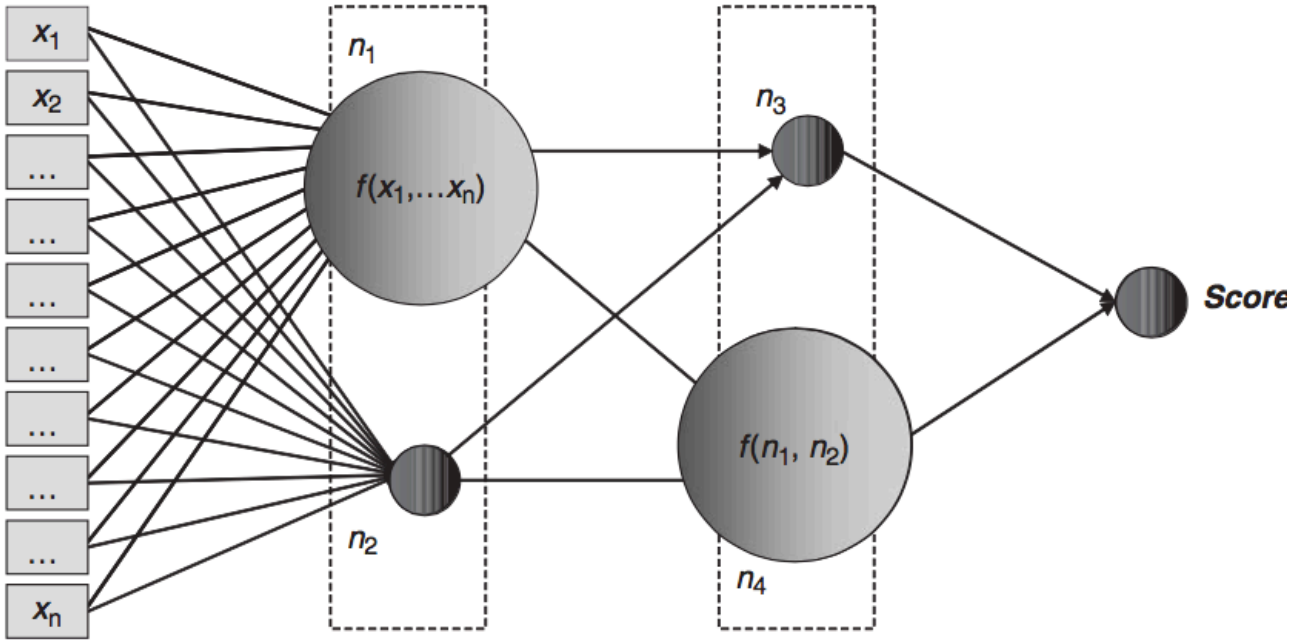


Figura 2.4 Esempio di struttura di una rete neurale

In generale, una rete neurale è costituita da un gran numero di elementi, chiamati “neuroni", che sono collegati tra loro da relazioni elementari chiamate “sinapsi”.
I neuroni sono disposti in “strati"; ogni neurone dello strato più esterno della rete riceve un input di n variabili (che, nel caso dei modelli utilizzati per la previsione di default, possono essere i consueti indicatori economico-finanziari) e li elabora con una funzione lineare o, più spesso, non lineare, il cui risultato viene trasmesso ai neuroni dello strato successivo. Questi neuroni elaborano anche l'ingresso ricevuto con un'ulteriore funzione e trasmettono una nuova uscita al livello successivo nella rete. Dopo uno o più “livelli nascosti", la rete genera un risultato finale. In caso di previsione dei default, il risultato può essere, ad esempio, un punteggio numerico che deve avere un valore il più vicino possibile a 1 per le imprese anomale e il più vicino possibile a 0 per le imprese sane proprio come per i modelli strutturali.
I coefficienti delle singole funzioni elementari che compongono la rete sono stimati (o meglio “appresi" dalla rete) attraverso un meccanismo iterativo. In pratica, i valori dei coefficienti si modificano gradualmente per ottenere risultati il più possibile simili a quelli desiderati (ad esempio, uno per le aziende che subiscono l’evento di default e zero per quelle sane). Il processo di apprendimento di una rete è quindi un graduale tentativo di individuare i corretti pesi da attribuire alle variabili di input e alle sinapsi degli strati nascosti, in modo da ottenere un risultato simile a quello della funzione (sconosciuta) da approssimare.

Algoritmi genetici

Sviluppati da Jhon Holland negli ’60 e ’70 del XX secolo, gli algoritmi genetici, proprio come le reti neurali, si ispirano al comportamento dell’intelligenza umana: tali tecniche si basano sui principi darwiniani della selezione naturale e del concetto di sopravvivenza all’adattamento. Come affermato da Darwin, gli individui che appartengono ad una specie animale competono tra loro per le risorse essenziali, come il cibo e il riparo; inoltre competono continuamente per attirare un compagno. Data una certa popolazione, solo gli individui con buone caratteristiche per interagire con l'ambiente esterno hanno un'alta probabilità di sopravvivere, e di conseguenza di riprodursi. Questo processo evolutivo porta quindi ad un continuo miglioramento

della specie, perché solo gli individui migliori trasferiscono il proprio patrimonio genetico alle generazioni future.

Altri due meccanismi, oltre alla selezione naturale, sono coinvolti nello sviluppo e nel miglioramento della specie. In primo luogo, quando due individui si accoppiano, la ricombinazione genetica detta anche crossover può portare ad una prole con caratteristiche migliori di quelle di ogni individuo nelle generazioni passate; questa maggiore qualità genetica si diffonde poi tra i membri della popolazione attraverso un'ulteriore riproduzione. In secondo luogo, l'ereditarietà genetica contenuta nei cromosomi degli individui può cambiare come risultato di improvvise, anche se rare mutazioni casuali di singoli geni.

Gli algoritmi genetici simulano il processo evolutivo appena descritto. Gli “individui"; necessari per evolvere non sono ovviamente esseri viventi, ma possibili soluzioni ad un problema.

In particolare, supponiamo di voler generare una funzione basata su un insieme di indicatori di bilancio (x1,x2,. . . . , xn), costruiti in modo da assegnare valori elevati alle aziende sane e valori bassi a quelle anomale.

Se per semplicità si considera una funzione lineare del tipo:

$$z = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots\ldots + \alpha_m x_m$$

ogni individuo si rappresenta con un vettore $\alpha_0 \ldots \alpha_m$, che indica sia il peso che il segno algebrico dei vati indicatori di bilancio utilizzati nella costruzione dello score.

Per determinare la soluzione migliore, gli algoritmi genetici operano come segue:

a. una prima popolazione detta di “generazione” di individui viene casualmente generata;
b. l’adattamento di tale popolazione, ovvero la capacità di rappresentare in modo corretto la soluzione al problema, si calcola tramite una funzione di valutazione. Nel caso in cui lo scopo sia quello di prevedere il default, si verifica, per esempio, che alle aziende anormali venga assegnato un punteggio basso e un punteggio elevato a quelle sane;

c. si applica un algoritmo di selezione detto anche “operatore genetico”, che ha lo scopo di selezionare ed individuare quali individui sono in grado di sopravvivere e quali no;
d. si applica un successivo operatore genetico detto “crossover” che permette agli individui sani di riprodursi, in questo modo si genera una seconda soluzione cioè una seconda generazione di individui;
e. l’applicazione di un operatore genetico di “mutazione” introduce la possibilità che si verifichino piccoli cambiamenti in una o più nuove generazioni;
f. l’adattamento di ogni individuo appartiene alla generazione di nuove soluzioni “figlie”. Il processo si itera finché non si raggiunge una soluzione soddisfacente oppure finché il miglioramento da una generazione ad un’altra risulta nullo.

Data una popolazione iniziale di soluzioni, gli algoritmi genetici, applicando ripetutamente regole predefinite, producono soluzioni con un livello di fitness crescente. Tuttavia, il processo di miglioramento della qualità dei risultati non è destinato a durare per sempre: quando sono state prodotte le soluzioni ottimali, che hanno un alto livello di fitness e sono molto simili tra loro, generano (per ricombinazione) una nuova generazione di soluzioni che saranno abbastanza simili ai loro “genitori". Si innesca così un processo di convergenza che porta i nuovi individui a diventare sempre più simili, e spesso addirittura identici.
Tale tecnica non garantisce di identificare la soluzione “ideale” al problema per cui queste tecniche si utilizzano, ma permettono comunque di ottenere “buone” soluzioni molto rapidamente.

2.10 Utilizzo, limiti e problematiche relative all'impiego dei modelli di credit-scoring

I modelli di credit scoring si possono utilizzare per diversi scopi:

i. previsione del default, calcolata cercando sostanzialmente di separare i rischi di credito affidabili da quelli non affidabili;
ii. stimare il livello di rischio del debitore, cioè la PD, assegnando una differente PD ad ogni debitore o ad un gruppo di debitori.

Quando si utilizzano modelli di credit scoring, si deve prestare attenzione ad alcuni problemi. Questo è particolarmente vero se si utilizzano modelli non solo allo scopo di determinare a quali imprese concedere il prestito, ma soprattutto per la stima della probabilità di default. Si illustrano le principali problematiche qui di seguito:

- Bisogna distinguere innanzitutto la differenza che esiste tra la definizione di "anormale" e di "insolvenza". Esistono differenti gradi di insolvenza, che si possono distinguere in ritardo nel pagamento degli interessi sul debito, fino alla liquidazione dell'impresa. La definizione utilizzata per scomporre il campione di stima influisce ovviamente sui risultati del modello, anche quando viene applicato ad una nuova società.
- Un altro problema è legato al fatto che la significatività delle variabili indipendenti utilizzate dal modello di scoring può variare nel tempo, per effetto per esempio del ciclo economico; non si evidenziano ragioni economiche per cui il peso dei singoli indicatori economico-finanziari utilizzati per spiegare il modello di default dovrebbero rimanere invariate.
- Un terzo problema è dovuto al fatto che i modelli di valutazione del credito ignorano numerosi fattori qualitativi, che possono essere molto significativi nel determinare l'insolvenza di un'impresa. Essi comprendono la reputazione dell'impresa (che influenza la sua capacità di accedere a fonti di credito alternative), la fase del ciclo economico, la qualità della gestione e le prospettive del settore cui l'impresa appartiene.

- Proprio perché spesso i dati completi esistono solo per poche aziende che vanno in default mentre è necessario lavorare con campioni di grandi dimensioni per avere stime migliori, c'è spesso il rischio che i campioni di stima siano sbilanciati e includano una percentuale eccessivamente alta di aziende sane. Da questo punto di vista, paradossalmente, una banca con un portafoglio storico di scarsa qualità è in una posizione migliore di una banca che in passato ha subito pochi eventi di default (ed è quindi costretta a lavorare con campioni stimati che sono troppo piccoli o significativamente squilibrati).

3 La regressione logistica

3.1 Introduzione

I metodi che sfruttano la regressione si stanno affermando parte sempre più integrante di molte analisi di dati, elaborazioni o studi che hanno lo scopo di descrivere la relazione esistente tra una variabile di risposta e una o più variabili esplicative. In alcuni casi è possibile che la variabile di risposta si rappresenti tramite una variabile cosiddetta discreta, ovvero che può assumere un numero limitato di valori. Negli ultimi anni il modello di regressione logistica si presenta come la risposta e l'approccio standard, in moti casi e applicazioni, a questo tipo di problema.

Prima di approfondire la tecnica di regressione logistica, si sottolinea che l'obiettivo di questo tipo di analisi è lo stesso di ogni altra tecnica che preveda la costruzione di modelli statistici: trovare il miglior modello di adattamento possibile e ragionevole al fine di descrivere la relazione che esiste tra una variabile risposta o dipendente e una o più variabili indipendenti dette anche predittori o variabili esplicative. La regressione lineare è il più semplice di questi modelli e si utilizza in tutti i casi in cui la variabile indipendente si presenta come una variabile continua.

La caratteristica principale che distingue un modello di regressione logistica da uno di regressione lineare si basa sul fatto che la variabile di uscita o di risposta nel caso di regressione logistica sia una variabile *binaria o dicotomica.* Le variabili di questo tipo si configurano come variabili che possono assumere solo due valori mutuamente esclusivi, in questo caso 0 oppure 1. La differenza tra regressione logistica e lineare si riflette sia sulle scelte dei paramenti che sulle assunzioni necessarie al modello. Per quanto concerne le metodologie utilizzate nei due diversi modelli di regressione, esse seguono gli stessi principi e assunzioni.

Sfruttando i metodi statistici, si è in grado, attraverso l'utilizzo della regressione logistica, di calcolare la probabilità con cui una osservazione o dato in ingresso appartenga a una classe o meno.

3.2 Principio di funzionamento

Come precedentemente spiegato, la regressione logistica si utilizza nei casi in cui non sia possibile l'impiego di altre tecniche di regressione come quella lineare per esempio, in quanto la variabile che si sta cercando di predire è una variabile binaria che identifica l'appartenenza o meno ad un gruppo o una classe. L'obiettivo è classificare un'osservazione.

Si riporta un esempio di studi che si svolgono utilizzando la regressione logistica:

- Determinare se una transazione bancaria è fraudolenta o sicura
- Capire se un tumore è maligno o benigno
- Classificare un'azienda come sana o anomala nella valutazione del rischio di credito

Tabella 3.1 Esempio di casi in cui si utilizza la regressione logistica

Classificazione osservazioni

$y \in \{0,1\}$	**1 o classe positiva**	**0 o classe negativa**
Transazione bancaria	Fraudolenta	Sicura
Tumore	Maligno	Benigno
Azienda sana o anomala	Anomala	Sana

La regressione logistica è di fatto un'analisi predittiva che si utilizza per valutare la relazione esistente tra una variabile dipendente e una o più variabili indipendenti, stimando delle probabilità attraverso una funzione logistica. Le probabilità si tramutano poi in valori binari per poter fare una previsione.

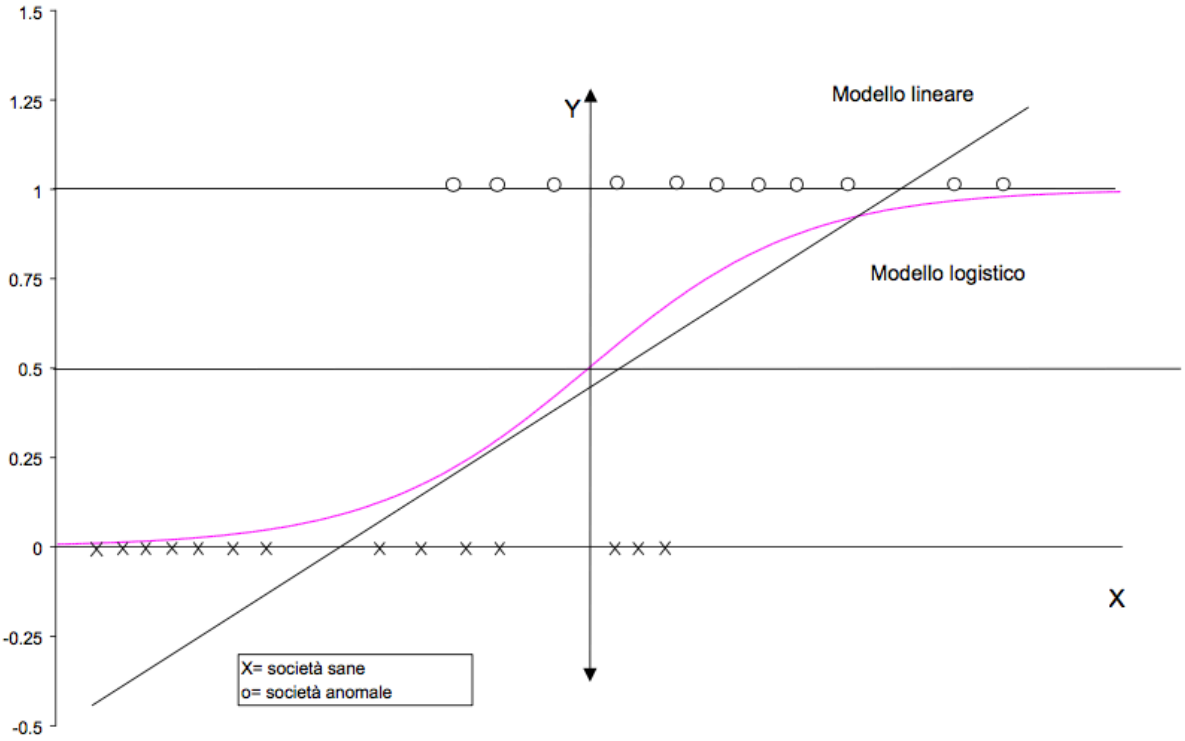


Figura 3.1 La differenza tra il modello lineare e quello logistico

Si ipotizzi dunque che esista una variabile che riesca a sintetizzare lo stato di salute dell'impresa e in particolare se questa sia o possa diventare insolvente.

Si indichi tale variabile con y': tale variabile non si può osservare a priori e si calcola come $y_i' = \alpha+\beta X_i+\mu_i$ *i-esima* variabile.

Si è in grado di osservare, invece, una realizzazione di tale variabile dicotomica:

$$f(x) = \begin{cases} 1, se\ y^* > 0 & se\ \mu > -\alpha - \beta x \\ 0, se\ y^* \leq 0 & se\ \ \mu \leq -\alpha - \beta x \end{cases}$$

La probabilità che $y_i=1$ coincide con la probabilità che $\mu_i > -\alpha - \beta x_i$ =1-F($-\alpha - \beta x_i$) dove F rappresenta la funzione di distribuzione cumulativa di μ_i.

$$\text{Prob}\ (y_i = 1)\ = \text{Prob}\ (u_i\ >\ -\ \alpha\ -\ \beta x_i)$$

$$= 1 - F\ (-\ \alpha\ -\ \beta x_i)$$

Si assume inoltre una distribuzione logistica delle probabilità.

$$\text{Prob}(y_i = 1) = P_i = 1 - F(-\alpha - \beta X_i)$$
$$= \frac{1}{1+e^{-(\alpha+\beta X_i)}} = \frac{e^{(\alpha+\beta X_i)}}{1+e^{(\alpha+\beta X_i)}}$$

$$\text{Prob}(y_i = 0) = 1 - P_i = \frac{1}{1+e^{(\alpha+\beta X_i)}}$$

Di conseguenza:

$$\frac{P_i}{1-P_i} = e^{(\alpha+\beta X_i)}$$

Applicando ad entrambi i membri della precedente equazione la funzione logaritmo si ottiene:

$$\ln\left[\frac{P_i}{1-P_i}\right] = \alpha + \beta X_i$$

che equivale a:

$$\ln\left[\frac{f_A(X_i)}{f_S(X_i)}\right] = \alpha + \beta X_i$$

dove *f* rappresenta la funzione di densità di probabilità delle popolazioni Sane e Anomale (S e A).

La funzione di verosimiglianza è la seguente:

$$L = \prod_{y_i=0} F(-\beta' x_i) \prod_{y_i=1} F(1-\beta' x_i)$$

La forma funzionale di F dipende dalle ipotesi che si effettuano su u_i , nel caso in cui si ipotizzi che gli errori seguano una distribuzione logistica allora si ottiene una funzione logit in cui :

$$F(-\beta' x_i) = \frac{1}{1+\exp(\beta' x_i)} = \Pr ob(y_i = 0 \mid x_i)$$
$$1 - F(-\beta' x_i) = \frac{\exp(\beta' x_i)}{1+\exp(\beta' x_i)} = \Pr ob(y_i = 1 \mid x_i)$$

da cui si ricava la funzione di verosimiglianza è

$$L = \prod_{i=1}^{n} \left(\frac{1}{1+\exp(\beta' x_i)} \right)^{(1-y_i)} \left(\frac{\exp(\beta' x_i)}{1+\exp(\beta' x_i)} \right)^{y_i}$$

Nel caso in cui si ipotizzi che la distribuzione cumulata siano di tipo Normali, si ottiene un altro tipo di modello detto Probit o Normit in cui si assume che la distribuzione dei residui sia di tipo Normale con media nulla e varianza costante.
Si riporta qui sotto la formulazione della distribuzione cumulata

$$F(-\beta' x_i) = \int_{-\infty}^{-\beta' x_i / \sigma} \frac{1}{(2\pi)^{1/2}} \exp(-\frac{r^2}{2}) dr$$

Poiché la distribuzione Normale e quella logistica hanno andamenti simili ad eccezione delle code, si utilizza principalmente il modello logit perché si privilegia sia dal punto di vista operativo e matematico in quanto più semplice.
La stima del logit si effettua con il criterio di massima verosimiglianza, con stima lineare per esempio con il metodo Newton-Raphson.
La probabilità condizionata che la variabile dipendente assuma il valore 1 data l'osservazione delle varie variabili indipendenti non è lineare, ma si comporta invece

come lineare la relazione logaritmo dell'odds ovvero il logit, con le variabili indipendenti.

Si definisce odds il rapporto tra probabilità che sono tra di loro complementari

$$odds = \frac{prob(y_i \mid x_i)}{1 - prob(y_i \mid x_i)}$$

$$Ln(odds) = \log it\,(prob(y_i \mid x_i)) = Ln\left[\frac{prob(y_i \mid x_i)}{1 - prob(y_i \mid x_i)}\right] = \beta' x_i$$

come spiegato in precedenza, una volta che si applica la funzione logaritmo all'odds si ottiene il logit.

La funzione di log-verosimiglianza sarà:

$$\ln(L) = \sum_{i=1}^{n} \{y_i \ln[prob(y_i \mid x_i] + (1 - y_i)\ln[1 - prob(y_i \mid x_i]\} =$$

$$= \sum_{i=1}^{n} \left\{ y_i \ln\left[\frac{prob(y_i \mid x_i)}{1 - prob(y_i \mid x_i)}\right] + \ln[1 - prob(y_i \mid x_i)] \right\} =$$

$$= \sum_{i=1}^{n} \left\{ y_i \beta' x_i + \ln\left[1 - \frac{e^{\beta' x_i}}{1 + e^{\beta' x_i}}\right] \right\} =$$

$$= \sum_{i=1}^{n} \left\{ y_i \beta' x_i + \ln\left[\frac{1}{1 + e^{\beta' x_i}}\right] \right\} = \sum_{i=1}^{n} \left\{ y_i \beta' x_i - \ln\left[1 + e^{\beta' x_i}\right] \right\}$$

Ipotizzando un comportamento lineare e utilizzando una sola variabile indipendente ed esplicativa x_i:

$$\beta' x_i = \beta_1 + \beta_2 x_i$$

Il sistema di equazioni di massima verosimiglianza si ricava calcolando la derivata prima del logaritmo della funzione di massima verosimiglianza rispetto ai coefficienti e ponendola uguale a zero per massimizzarne il valore assunto:

$$\frac{\partial Ln(L)}{\partial \beta_1} = \sum_{i=1}^{n} \left\{ y_i - \frac{e^{\beta_1+\beta_2 x_i}}{1+e^{\beta_1+\beta_2 x_i}} \right\} = 0$$

$$\frac{\partial Ln(L)}{\partial \beta_2} = \sum_{i=1}^{n} \left\{ y_i x_i - \frac{e^{\beta_1+\beta_2 x_i}}{1+e^{\beta_1+\beta_2 x_i}} x_i \right\} = 0$$

Se si definisce p_i la probabilità dell'evento:

$$\frac{e^{\beta_1+\beta_2 x_i}}{1+e^{\beta_1+\beta_2 x_i}}$$

allora:

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} p_i$$

$$n_a = \sum_{i=1}^{n} p_i$$

In definitiva si può affermare che la somma delle probabilità dell'evento corrisponde alle somme degli eventi osservati.

3.3 Test sulla regressione logistica

Dopo aver stimato i coefficienti di un modello, si procede di norma nella valutazione della significatività delle variabili utilizzate nel modello. Ciò spesso implica formulazione e test di ipotesi statistici al fine di determinare se le variabili indipendenti siano "*significative"* in relazione alla variabile di risposta.

Un approccio per valutare la significatività del coefficiente di una variabile all'interno di un modello si basa sulla valutazione del modello con e successivamente senza la variabile in questione. Se il modello costruito utilizzando una generica variabile indipendente aggiunge più informazioni e fornisce un'interpretazione più accurata della variabile risposta, allora è necessario inserire tale variabile nel modello.

La procedura da seguire in questo caso si basa sul confronto tra i valori osservati in uscita dal modello con la variabile indipendente e cioè comprenderla nel modello e i valori di variabile dipendete ottenuti senza essa. La funzione matematica che si deve utilizzare per comparare i modelli dipende dal particolare tipo di problema. Se i valori predetti dal modello in uscita utilizzando una variabile sono migliori o in un certo senso più accurati, allora si può definire come significativa tale variabile.

Il metodo generale per valutare la significatività delle variabili è quello che si utilizza in tutti i modelli di regressione come quello lineare: in questo caso si utilizza l'analisi della varianza.

Ciò che accomuna l'analisi e la valutazione della significatività statistica nella regressione lineare e in quella logistica è il fatto di confrontare i valori ottenuti dalla variabile risposta con i valori forniti dal modello con l'utilizzo o meno della variabile indipendente. Nella regressione logistica, il confronto tra valori osservati e predetti si basa sulla funzione log-verosimiglianza descritta nel paragrafo precedente.

Tale confronto si basa sulla seguente espressione, che si definisce come rapporto di verosimiglianza:

$$G = -2\ln\left[\frac{likelihood\ \ del\ \ \text{mod}ello\ \ senza\ la\ \ \text{var}iabile\ x}{likelihood\ \ del\ \ \text{mod}ello\ \ con\ la\ \ \text{var}iabile}\right] = -2\ln\left[\frac{\left(\frac{n_a}{n}\right)^{n_a}\left(\frac{n_s}{n}\right)^{n_s}}{\prod_{i=1}^{n} prob(y_i \mid x_i)^{y_i}\left[1 - prob(y_i \mid x_i\right]^{(1-y_i)}}\right]$$

La statistica *G* è definita da alcuni autori (come McCullagh e Nelder (1983)) *devianza* e gioca un ruolo fondamentale nella valutazione della bontà di adattamento.
L'ipotesi iniziale è che il modello non sia più accurato quando si inserisce la variabile indipendente. *G* assume una forma asintotica come una variabile chi-quadro e si confronta il p-value con il livello di significatività scelto α, se minore allora si rifiuta l'ipotesi inziale della non significatività della variabile.

Un altro indice che si trova in letteratura è il cosiddetto LRI ovvero "*Likelihood ratio index"* pensato da McFadden e appartenente alla categoria degli pseudo-R^2.
Tale indice si costituisce come rapporto fra il valore di log-verosimiglianza dell'intero modello con tutte le variabili di stima utilizzate e il valore della funzione di log-verosimiglianza calcolata utilizzando solo la costante. Tale test si effettua per verificare se tutti i coefficienti delle variabili siamo nulli o meno.

$$LRI = 1 - \frac{Ln(L)}{Ln(L_0)}$$

Spesso si trova rappresentato come complemento ad uno del rapporto descritto in precedenza.
L'indice assume valori compresi tra 0 e 1. Se pur il valore 0 può essere assunto dal rapporto non è altrettanto per il valore unitario. Infatti, salvo il caso estremo in cui la probabilità dell'evento fosse 100% se y=1 e 0 se y=0 in cui la verosimiglianza assume il valore 1 e il logaritmo di 1 varrebbe 0 portando così l'indice uguale a 1.
La log-verosimiglianza è un numero negativo, tanto più negativo quanto il modello è insoddisfacente (il modello con la sola costante dovrebbe avere la log-verosimiglianza più negativa del modello con le variabili); quindi il rapporto tra le funzioni di log-

verosimiglianze dovrebbe essere tanto più piccolo quanto più il modello con le variabili è efficace, portando il LRI a valori elevati.

I valori che l'indice LRI assume sono generalmente piccoli intorno ad un range che comprende i valori (0,2-0,4):valori che sono piccoli rispetto a quelli ottenuti dall' R^2.

Ogni coefficiente del modello si valuta come rapporto tra il valore stimato del coefficiente risultante dalla regressione e facente parte del modello e il suo errore standard relativo. Scelto un determinato livello di significatività si può facilmente determinare se il coefficiente stimato sia significativo o meno. Tale tecnica prende il nome di test di Wald.

$$Wald\ test = \frac{\beta_j}{S.E.(\beta_j)}$$

Se si fosse invece interessati a costruire un intervallo di confidenza per il coefficiente β_j si può facilmente scrivere come:

$$\beta_j \pm z_{1-\alpha/2} * S.E.(\beta_j)$$

dove $z_{1-\alpha/2}$ indica il punto 100(1-α/ 2)% della distribuzione Normale standard.

Per quanto concerne invece l'errore standard relativo ai coefficienti, esso si riferisce alle derivate parziali seconde della funzione di log-verosimiglianza ordinate in una matrice di cui si calcola l'inversa.

3.4 Un esempio di regressione logistica: il modello di Ohlson (1980)

Si riportano al fine di esempio tre modelli pubblicati da J.Ohlson nel 1980.Tali modelli si riferiscono ad un un campione non bilanciato di imprese industriali statunitensi degli anni '70. Il primo modello si pome come obiettivo la stima della probabilità di default un anno prima che il default si manifesti. Il secondo modello stima la invece la probabilità di default due anni prima del default ed il terzo modello stima la probabilità che il default avvenga nel primo o nel secondo anno prima del default.
I modelli si costituiscono delle seguenti variabili:

- $SIZE = \ln\left({attivo}/{livello\ dei\ prezzi\ impliciti\ nel\ PIL}\right)$
- $TLTA = {debiti\ totali}/{attivo}$
- $WCTA = {capitale\ circolante}/{attivo}$
- $OENEG = \begin{cases} 1\ se\ il\ totale\ debiti > attivo \\ 0\ altrimenti \end{cases}$
- $NITA = {ris.netto\ di\ bilancio}/{attivo}$
- $FULT = {flusso\ di\ cassa}/{debiti\ totali}$
- $INTWO = \begin{cases} 1\ se\ il\ ris.netto\ di\ bilancio < 0 \\ 0\ altrimenti \end{cases}$
- $CHIN = misura\ la\ variazione\ del\ risulatto\ netto\ di\ bilancio$

I risultati dei modelli ottenuti si illustrano nella tabella qui sotto riportata , dove si concentra l'attenzione in particolare sui coefficienti del modello e il loro segno algebrico (Tabella 3.2).

Tabella 3.2 I coefficienti di modelli studiati da Ohlson

	MODELLO 1	MODELLO 2	MODLLO 3
Costante	-1,32	1,84	1,13
SIZE	-0,407	-0,519	-0,478
TLTA	6,03	4,76	5,29
WCTA	-1,43	-1,71	-0,99
CLCA	0,0757	-0,297	0,062
NITA	-2,37	-2,74	-4,62
FUTL	-1,83	-2,18	-2,25
INTWO	0,285	-0,78	-0,521
OENEG	-1,72	-1,98	-1,91
CHIN	-0,521	0,4218	0,212

Si noti come, seppur i coefficienti dei modelli siano significativi dal punto di vista statistico, alcuni di essi non si presentano con il segno economico "corretto".

Infatti, variabili come *SIZE* presentano il segno negativo correttamente in quanto maggiore è il valore di tale variabile e minore sarà la probabilità di default (in accodo con l'idea che maggiore sarà l'attivo dell'impresa rispetto al livello dei prezzi impliciti nel PIL e minore sarà la probabilità con cui si verificherà un default).

Discorso analogo per le variabili come TLTA, ma in questo caso la correlazione tra il coefficiente della variabile e la probabilità di default sarà opposto. Infatti maggiore sarà il valore di questa variabile e maggiore sarà la probabilità di default vista la presenza di un segno positivo nel coefficiente. In questo caso si applica invece un ragionamento per cui più grande è il rapporto fra debiti totali e attivo e più grande sarà la probabilità di default.

Si noti inoltre come si presentano quattro errori nel segno dei coefficienti.

- In MODELLO 2 il segno algebrico di CLCA negativo non è corretto, in quanto maggiore è il valore di tale variabile e più grande è la probabilità di default. La presenza del segno negativo indicherebbe che se un'impresa presenta livelli

maggiori di attività correnti su passività correnti rispetto ad un'altra sarebbe più meno incline al default;

- In MODELLO 2 e in MODELLO 3 entrambi si presentano con un segno negativo di fronte alla variabile INTWO, in questo caso maggiore è la variabile binaria e più grande è la probabilità di default. Il segno negativo indicherebbe che una società con risultato netto negativo ha una probabilità di default minore di una con risultato netto positivo;
- In MODELLO 2 e in MODELLO 3 i coefficienti della variabile CHIN sono positivi, ciò comporta il fatto che se aumenta la variazione di risultato netto in due anni consecutivi allora la probabilità di default aumenterebbe per via del segno positivo del coefficiente. Questo ovviamente rappresenta un errore in quanto se si manifestasse un aumento del risultato netto da un anno a quello successivo, questo comporterebbe una riduzione della probabilità di default e non un suo aumento.

Si sottolinea che per quanto un modello sia accurato e preciso, un coefficiente con segno economicamente non corretto non deve essere incluso nel modello. Nel caso di errori di segno il modello si dovrebbe rifiutare.

4 Analisi ed elaborazione dei dati

4.1 Scelta del settore

Nella valutazione del rischio di credito di un settore industriale tramite un modello di scoring, riveste un ruolo fondamentale la valutazione e la selezione di un'industria particolare. Il settore, infatti, si presenta come particolarmente soggetto ad eventi di default (il significato di questo termine si descriverà successivamente). Tali eventi si sono manifestati in passato e si presentatane ancora tuttora con una certa periodicità. Varie erano le possibilità nell'individuazione del settore; in questo lavoro di tesi si è scelto il macro-settore tessile. Poiché il settore tessile si presenta come uno dei settori manifatturieri più grandi e sviluppati in Italia, che conta un numero di imprese molto elevato, è stato necessario individuare in modo più dettagliato un micro-settore all'interno del tessile italiano.

Per prima cosi si è valutato il codice ATECO ovvero il codice che classifica le aziende italiane per l'attività economica svolta. Al fine di procedere in modo mirato all'interno di un settore si è stabilito che le imprese appartenenti a tale settore dovessero avere tutte lo stesso codice ATECO identico alla 3° cifra. Il codice è infatti costituito da sei cifre, che spiegano progressivamente nel dettaglio quale sia l'attività economica svolta dall'impresa così classificata. Il codice ATECO del settore tessile è 130000: la prima cifra cioè il numero 1 classifica l'impresa come appartenente al settore manifatturiero, mentre la seconda cifra cioè il numero 3 si riferisce al fatto che l'impresa opera nell'industria del tessile. Come già sottolineato, vista la vastità del settore tessile, si è scelto di analizzare la sola tessitura. Le aziende che lavorano e operano al fine di produrre tessuti e classificate quindi come facenti parte la tessitura presentano invece codice ATECO 132000: in questo caso l'ultima cifra si riferisce al settore manifatturiero tessile della sola fase di tessitura.

Nella classificazione con tale codice ATECO si sono rilevate un certo numero di imprese soggette a particolari eventi. Il numero di imprese selezionate e di cui sono state scaricate le informazioni ammonta a 451, tutte imprese italiane che si sono classificate con codice ATECO 132000 tessitura. Di tale imprese, prima di scaricare i dati di bilancio si è valutato nel dettaglio quali eventi avevano subito in passato.

La valutazione di questi eventi è stata fatta riconducendoli alla seguente classificazione che smista gli eventi in 4 diversi flag ognuno dei quali le classifica secondo lo stato di salute dell'impresa:

- ***Flag=1: società anomala***
  - *Concordato preventivo*
  - *Fallimento*
  - *Amministrazione giudiziaria*
  - *Chiusura del fallimento*
  - *Altre cause*
  - *Liquidazione giudiziaria*
  - *Motivo non precisato*
  - *Stato di insolvenza*
  - *Sequestro giudiziario*
  - *Concordato fallimentare*
  - *Amministrazione controllata*
  - *Cancellazione per comunicazione piano di riparto*
  - *Amministrazione straordinaria*
  - *Chiusura per fallimento o liquidazione*
  - *Decreto cancellazione tribunale*
  - *Scioglimento per atto dell'autorità*
  - *Sequestro conservativo di quote*
  - *Bancarotta*
- ***Flag=2: società sana in condizioni particolari***
  - *Liquidazione volontaria*
  - *Scioglimento e liquidazione*
  - *Scioglimento*
  - *Chiusura della liquidazione*
  - *Cessazione di ogni attività*
  - *Cancellata d'ufficio ai sensi art. 2490 c.c. (bilancio di liquidazione)*
  - *Liquidazione*
  - *Scioglimento e messa in liquidazione*
  - *Chiusura per liquidazione*

- *Scioglimento senza messa in liquidazione*
- *Cessazione dell'attività nella provincia*
- *Cessazione d'ufficio*

- ***Flag=3: società sana in condizioni particolari***
  - *Fusione mediante incorporazione in altra società*
  - *Scissione*
  - *Trasferimento sede all'estero*
  - *Fusione mediante costituzione di nuova società*
  - *Cessione azienda*
- ***Flag=4: società sana in condizioni particolari***
  - *Cessata*
  - *Cancellata dal registro delle imprese*
  - *Trasferimento in altra provincia*
  - *Cancellata d'ufficio a seguito istituzione cciaa di Fermo, di Monza*

Con *flag=0* si indica invece una società che non presenta particolari segnalazioni e si può quindi considerare come società sana.
Una volta individuato un numero "ragionevole" di società che presentano uno di questi flag, considerando anche tutte le società appartenenti al settore, si può procedere allo scaricamento di tutti i dati necessari a costruire un modello di scoring. Si noti come, fra tutti e quattro i flag, quelli che sono più importanti e significativi al fine di costruire un modello di credit scoring sono il flag 1 e il flag2.

4.2 Il campione di imprese

Il campione di imprese che si è utilizzato è costituito da 451 imprese tutte classificate con codice ATECO 132000 (tessitura). I dati necessari allo svolgimento dell'analisi del rischio di credito di questo settore si costituiscono sia di dati anagrafici sia di dati di bilancio.

I dati anagrafici riguardano:

- la ragione sociale ovvero la denominazione dell'impresa;
- l'anno di costituzione;
- il comune, la provincia e la regione della sede legale dell'impresa;
- il codice ISTAT di comune, regione e provincia (il codice ISTAT è un codice necessario ad indentificare e classificare l'impresa secondo l'area geografica);
- la forma giuridica (S.P.A.; S.R.L; S.A.P.A.; S.A.S. ecc.);
- il codice fiscale dell'impresa;
- se il bilancio sia consolidato o meno;
- se il bilancio rispetta i principi contabili internazionali IFRS;
- se il bilancio sia abbreviato o dettagliato;
- il codice ATECO 2007 dell'impresa;
- lo stato giuridico;
- se l'impresa risulta quotata o no;
- l'eventuale procedura che l'impresa ha subito;
- la data di inizio di conclusione della procedura (se presente);

I dati di bilancio si riferiscono invece a tutte le voci che costituiscono lo stato patrimoniale ed il conto economico.

Si è scelto di introdurre nel campione tutte le imprese appartenenti al settore della tessitura con informazioni il più possibile recenti; di conseguenza si sono selezionate le imprese il cui ultimo bilancio risalisse al 2018 oppure al 2017, visti i pochi bilanci depositati che risultavano aggiornati al 2018. Infatti, delle 451 imprese solo 185 presentano l'ultimo bilancio al 2018.

Al fine di avere un numero di osservazioni accettabili, si è scelto di utilizzare una base di analisi costituita dalla serie storica di bilanci 2010-2018 oppure 2010-2017, ove presenti le informazioni relative a tali anni (non tutte le imprese si presentano con una serie così ampia di bilanci).

Il numero totale di osservazioni raccolte ammonta a 3679, se si considera che derivano da 451 imprese, in media per ogni impresa ci sono 8 osservazioni ovvero 8 bilanci disponibili.

Il campione si costituisce quasi in totalità da società di capitali ad eccezione di una sola società di persone presente sotto forma giuridica si S.A.S. e due società consortili sotto forma di S.C.A.R.L.P.A. (cfr. Tabella 4.1).

Tabella 4.1 LA suddivisone del campione in base alla forma giuridica.

Forma giuridica	*Numero di imprese*	*Peso percentuale*
S.P.A.	119	26,39%
S.R.L.	329	72,95%
S.A.S.	1	0,22%
S.C.A.R.L.P.A.	2	0,44%
Totale	*451*	*100%*

Si sottolinea come nessuna delle imprese presenti nel campione, tra le società per azioni, sia quotata.

Le imprese sono generalmente di piccole dimensioni, riflettendo la struttura frammentata del settore tessile: il fatturato medio ammonta a circa 8 milioni di euro mentre il numero medio di dipendenti si aggira intorno ai 30.

Si è inoltre effettuata un'analisi per la valutazione dell'area geografica a cui ogni impresa appartiene secondo questa suddivisone:

- nord-ovest a cui appartengono le imprese la cui sede si trova in: Valle d'Aosta, Piemonte, Liguria e Lombardia;
- nord-est: Trentino Alto Adige, Friuli Venezia Giulia, Emilia Romagna, Veneto;

- centro: Toscana, Umbria, Marche, Abruzzo, Molise e Lazio;
- sud e isole: Campania, Puglia, Basilicata, Calabria, Sicilia e Sardegna.

In termini numerici si può osservare la suddivisone delle osservazioni in area geografica in Tabella 4.2.

Tabella 4.2 Suddivisione del campione per macro area geografica

Macro Area geografica	*Numero di osservazioni (numero di bilanci)*	*Peso percentuale*
Nord-ovest	1974	53,65%
Nord-est	394	10,71%
Centro	1222	33,22%
Sud e isole	89	2,42%
Totale	*3679*	*100%*

Una volta ottenute tutte queste informazioni, si sono scaricate le informazioni di bilancio relative ad ogni impresa. Per ogni azienda e per ogni osservazione del campione (corrispondente ad un bilancio), sono state scaricate tutte le voci di bilancio, sia quelle relative allo stato patrimoniale sia quelle relative al conto economico, di tutti gli anni disponibili per la suddetta azienda.

Tutti i dati utilizzati in questo lavoro di tesi sono stati scaricati dal database AIDA ("Analisi Informatizzata delle Aziende").

4.3 Descrizione delle procedure di elaborazioni dati

4.3.1 Inserimento dati

I dati anagrafici e di bilancio delle imprese sono stati esportati dal database AIDA in formato file Excel. Tali dati sono stati successivamente copiati su un altro file al fine di rendere più agevole l'analisi. In tale file ogni riga rappresenta un'osservazione, in cui sono presenti i dati necessari all'identificazione dell'impresa come la ragione sociale, la forma giuridica, il codice fiscale ecc. e tutte le informazioni di bilancio relative a tale anno. Ogni colonna rappresenta una voce di bilancio disposte in ordine: prima sono state riportate tutte le voci dell'attivo dello stato patrimoniale, successivamente quelle del passivo ed infine quelle relative al conto economico.

Una volta ricopiati tutti i dati, si deve valutarne l'accuratezza.

Tra i vari dati riportati, oltre alle informazioni anagrafiche e a quelle di bilancio, sono stati riportati anche gli eventi di default. In questo caso, se l'azienda ha subito uno degli eventi riportati tra quelli presenti in *flag=1* e *flag=2*, allora si individuava tale azienda inserendo semplicemente un valore pari ad uno all'interno di una colonna per la classificazione della società come "anomala". Tale operazione è stata fatta sia per identificare se la società avesse subito un particolare evento e quindi inserire tanti uno quanti i bilanci della società ed inoltre in un'altra colonna si è inserito il valore uno solo per individuare l'anno in cui l'evento si è manifestato. In questo modo si ottengono due colonne che mettono in luce la differenza tra il default della società e il default della società in uno specifico anno, fondamentali per le analisi successive.

Si sottolinea come in moltissimi bilanci, non tutte le informazioni relative ad ogni voce di bilancio fossero disponibili. Spesso infatti si raccoglievano solo le voci "principali", mentre i dati relativi alle voci "secondarie" erano assenti. In questo caso si sostituisce il valore "n.d." ovvero dato non disponibile con il valore zero.

4.3.2 Valutazione degli errori nei dati

Un lavoro fondamentale che si è svolto riguarda la "pulizia" dei dati del campione scaricati dal database. Sfruttando un altro foglio Excel collegato al primo in cui si riportavano le medesime informazioni anagrafiche e di bilancio si è potuto procedere alla correzione di alcuni errori. In tale in file, le voci di attivo, passivo e conto economico sono state raggruppate e aggregate fra di loro in macro voci al fine di valutare se esistessero differenze tra attivo e passivo per esempio. Più nel dettaglio la suddivisone dell'attivo dello stato patrimoniale è la seguente:

- attività correnti costituite da rimanenze, crediti esigibili e liquidità
- totale immobilizzazioni costituite da immobilizzazioni immateriali, materiali e finanziarie.

La somma delle attività correnti e del totale immobilizzazioni fornisce l'attivo netto.

Per quanto riguarda l'aggregazione delle voci del passivo invece la suddivisone è la seguente:

- passività correnti costituite da debiti entro l'esercizio e ratei/risconti
- capitali permanenti costituiti invece da patrimonio netto, fondi rischi e TFR e dai debiti oltre l'esercizio.

Il controllo nel caso dello stato patrimoniale si effettua sulla differenza tra attivo netto e passivo netto, se diversa da zero allora si rileva un errore. Grazie alle aggregazioni delle macro voci, in alcuni casi l'errore si evidenzia rapidamente e si può correggere agevolmente. In altri casi, invece, l'individuazione di errori e la successiva correzione si rivela un lavoro molto arduo: spesso infatti i dati estratti non erano corretti e ciò si può affermare dal fatto che le voci di bilancio non coincidevano.

La stessa procedura è stata eseguita per il conto economico, al fine di valutare anche in questo caso la presenza di eventuali errori nei dati scorporando le varie voci di bilancio. Dal valore della produzione si passa al valore aggiunto operativo sottraendo

consumi e servizi esterni, successivamente all'EBITDA (o MOL) sottraendo il costo del lavoro. Successivamente si ottiene l'EBIT sottraendo al MOL gli ammortamenti e svalutazioni, si giunge poi al risultato corrente una volta considerati proventi e oneri finanziari. Si raggiunge il risultato rettificato ante imposte sottraendo le partite straordinarie ed in fine sottraendo le imposte si ottiene il risultato rettificato. Sommando o sottraendo le rivalutazioni si ottiene il risultato netto o utile di esercizio. Il controllo di errori nel caso del conto economico si effettua quando, una volta scorporate tutte le voci di conto economico la differenza tra l'ultima voce ottenuta e il valore dell'utile netto di bilancio riferito al foglio precedente risulti diversa da zero. Nel caso di un errore infatti, la differenza non nulla deve essere "aggiustata". Tale lavoro ha richiesto molto tempo in quanto spesso l'errore non è facilmente individuabile. Si porta al fine di esempio, un caso che si è ripetuto con frequenza, quello relativo alla mancanza di dettagli nei bilanci scaricati dove non vi era suddivisione fra ammortamenti di immobilizzazioni materiali e quelli immateriali, ma si specificava solamente il valore totale delle immobilizzazioni. Nel caso di errori è stato necessario valutare singolarmente le voci di bilancio al fine di capire quale fosse la causa.

4.3.3 Elaborazione dei dati

Solo dopo aver sistemato e corretto i dati ed integrato quelli mancanti, si è potuto procedere alla elaborazione e ai calcoli necessari per ottenere i dati finali da inserire all'interno del modello di regressione. Generalmente i modelli di credit scoring utilizzano principalmente indicatori di bilancio che rivestono il ruolo di variabili indipendenti, insieme ad alcune variabili di tipo anagrafico come l'età dell'impresa e l'area geografica per esempio.
In questo lavoro di tesi si sono calcolati 42 indicatori di bilancio, sfruttando ovviamente i dati di bilancio di ogni società per ogni anno. Ad ogni osservazioni corrisponde dunque un indice di bilancio.
Tal indicatori si suddividono in 3 grandi categorie che sono:

- *indicatori di redditività*
- *indicatori di produttività e struttura operativa*
- *indicatori di liquidità e struttura finanziaria*

Alla prima categoria di indicatori appartengono indicatori come: EBITDA/Ricavi, EBIT/Ricavi, Utile corrente/Ricavi, Risultato netto rettificato/Ricavi, EBITDA/AN, ROA, ROE, ROE ante imposte.
Alla classe di indicatori di produttività e struttura operativa appartengono invece indicatori come: Valore Aggiunto Operativo/Ricavi, Consumi/costi operativi, Servizi esterni/costi operativi, Costo del lavoro/costi operativi, Ammortamenti materiali/costi operativi ecc.
Infine, tipici indicatori di liquidità e struttura finanziaria sono: AC/PC, AC-mg/PC, Liquidità/PC, Patrimonio netto/AN, Capitale circolante/AN, Patrimonio netto tangibile/AN, OFN/EBIT, OFN/EBITDA, ln(Ricavi), ln(AN) e molti altri.
Una volta calcolati tutti gli indicatori di bilancio per ogni osservazione, si procede con il calcolo del 5° e del 95° percentile, per ogni tipologia di indicatore. Tale passaggio risulta fondamentale nell'analisi egli outlier, ovvero di valori assunti dagli indicatori tropo elevati o troppo piccoli che potrebbero compromettere i risultati delle regressioni. Infatti, per ognuno dei 42 indicatori, se i valori da essi assunti fossero minori del valore corrispondente al 5° percentile, ad essi si attribuiva il valore del 5°percentile; se invece il valore dell'indicatore calcolato è maggiore del 95° allora si sostituisce tale valore con quello del 95°. In questo modo la discrepanza tra i dati si riduce notevolmente e le osservazioni sono tra di loro maggiormente confrontabili, si evita inoltre il rischio che alcuni indicatori assumano valori troppo elevati dovuti al fatto che in assenza di dati è stato inserito il valore 0 oppure 1 (in caso la voce di bilancio che presenta l'assenza di dato fosse a denominatore nel calcolo dell'indicatore).
Successivamente, tramite l'utilizzo di semplici funzioni Excel, si sono calcolate diverse correlazioni. In particolare, è stata costruita una matrice di correlazione dove si è valutata la correlazione tra ogni coppia di indicatore: tale passo risulta fondamentale nella successiva elaborazione dei modelli al fine di capire quali siano gli indicatori maggiormente adatti a farne parte.

Un altro calcolo fondamentale nella valutazione di quali indicatori utilizzare nei modelli di credit scoring riguarda la valutazione della correlazione tra gli indicatori e i flag. Infatti, per ogni indicatore, è stata calcolata la correlazione tra essi e il flag sulla società (il flag assume il valore 1 per ogni anno di osservazione se la società ha manifestato un evento di default, indipendentemente dall'anno in cui si è manifestato) e la correlazione tra indicatore e flag anno (il flag in questo caso assume il valore 1 solo in riferimento all'anno in cui si è verificato l'evento di default). Risulta fondamentale il segno della correlazione al fine di capire se l'indicatore sia adatto a rappresentare correttamente il default. Infatti il segno deve rispettare il significato economico dell'indicatore:

- se maggiore è il valore dell'indicatore e minore è la probabilità di default allora il segno di tale indicatore deve essere negativo come per esempio l'indice di liquidità oppure indicatori come EBIT/Ricavi;
- se maggiore è il livello dell'indicatore e maggiore è la probabilità di default allora il segno di tale indicatore deve essere positivo per esempio si può citare OFN/EBITDA oppure Debiti finanziari/valore aggiunto.

In questo lavoro di tesi non tuti gli indicatori si sono presentati con il segno economicamente corretto. Il segno algebrico dalla correlazione in alcuni casi non è corretto, e ciò è probabilmente dovuto alla mancanza di dati e alla loro bassa accuratezza. Si noti come sono stati utilizzati esclusivamente indicatori che presentavano il segno della correlazione corretto con quello che ci si aspetta dal significato economico-finanziario di ognuno di essi.

Costruzione dei modelli di scoring

Per quanto concerne la costruzione dei modelli, si è ritenuto necessario innanzitutto valutare quali indicatori inserire nel modello e in particolare da quali indicatori incominciare. Per capire quale fossero i primi indicatori da utilizzare è stato prima necessario calcolare l'accuracy di ogni indicatore ovvero una sorta di precisione dello stesso. A tal fine si è utilizzato un foglio Excel in cui si è seguita la seguente procedura:

- inserimento del numero di osservazione, valor dell'indicatore di cui si intende calcolare l'accuracy e il flag relativo all'osservazione (flag=1 se la società ha presentato un default);
- utilizzo di una colonna che che rappresenta il modello "perfetto" costituita da soli 0 e 1, si ordinano tali valore al fine di ottener prima tutti valori nulli che determinato il fatto che la società sia sana e successivamente tanti valori unitari quanti sono i bilanci delle società in default;
- tramite un ordinamento sincronizzato delle colonne che contengono le informazioni relative all'osservazione, al valore dell'indicatore e al flag, indicando come variabile principale di ordinamento l'indicatore, che si ordina in modo crescente o in modo decrescente in funzione del suo significato economico;
- tramite delle differenze percentuali relative alle osservazioni totali si calcola l'accuracy: si ripete tale operazione per ogni indicatore.

I passaggi che si sono utilizzati per calcolare l'accuracy degli indicatori sono stati ripetuti anche per l'accuracy degli indicatori calcolati in base al flag anno oltre a quelli relativi al flag società, con l'unica differenza consiste nel fatto che il modello perfetto prevede ovviamente un numero di osservazioni (bilanci) in default minore in quanto si considera solo l'anno in cui si verifica l'evento di default e non tutte le osservazioni relative a tale società. Se l'indicatore presenta un valore di accuracy elevata allora significa che le osservazioni e i relativi valori degli indicatori delle società in default, una volta ordinati, si trovano in fondo al foglio di calcolo.

Terminato il calcolo dell'accuracy, inizia la costruzione del modello vero e proprio. Partendo dall'indicatore che ha la precisione più elevata, si aggiunge successivamente un secondo indicatore che deve rispettare le seguenti condizioni:

- l'indicatore deve presentare un valore di accuracy elevato
- l'indicatore non deve essere fortemente correlato con l'indicatore precedentemente inserito.

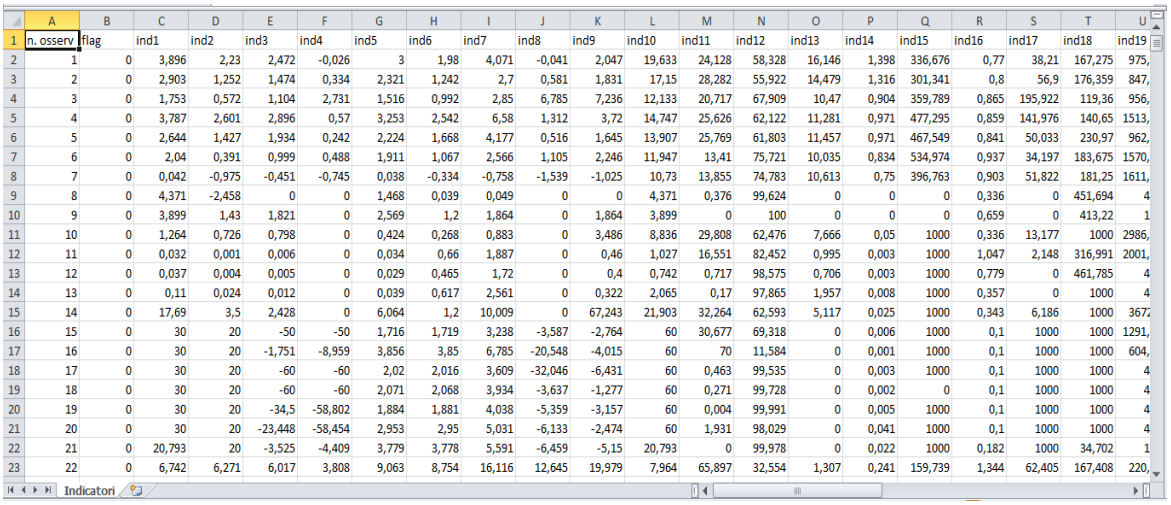
Una volte che si è selezionato l'indicatore che rispetta le due precedenti condizioni, si sfrutta lo strumento della regressione logistica al fine di determinare quale siano i

coefficienti di tali indicatori che costituiscono il modello e il valore della cosiddetta intercetta.

La regressione logistica è stata effettuata utilizzano il software statistico "*R*". Le informazioni necessarie a costruire un modello di regressione logistica detto anche logit sono gli indicatori che si intende inserire nel modello e il flag di società o di anno, ovvero quella variabile binaria già descritta in precedenza che assume il valore unitario se l'azienda ha subito un default.

Più nello specifico si riportano di seguito i passaggi necessari e l'algoritmo necessario al funzionamento del software statistico.

1. Impostare la struttura del file Excel con gli indicatori nel seguente modo:



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	n. osserv	flag	ind1	ind2	ind3	ind4	ind5	ind6	ind7	ind8	ind9	ind10	ind11	ind12	ind13	ind14	ind15	ind16	ind17	ind18	ind19
2	1	0	3,896	2,23	2,472	-0,026	3	1,98	4,071	-0,041	2,047	19,633	24,128	58,328	16,146	1,398	336,676	0,77	38,21	167,275	975,
3	2	0	2,903	1,252	1,474	0,334	2,321	1,242	2,7	0,581	1,831	17,15	28,282	55,922	14,479	1,316	301,341	0,8	56,9	176,359	847,
4	3	0	1,753	0,572	1,104	2,731	1,516	0,992	2,85	6,785	7,236	12,133	20,717	67,909	10,47	0,904	359,789	0,865	195,922	119,36	956,
5	4	0	3,787	2,601	2,896	0,57	3,253	2,542	6,58	1,312	3,72	14,747	25,626	62,122	11,281	0,971	477,295	0,859	141,976	140,65	1513,
6	5	0	2,644	1,427	1,934	0,242	2,224	1,668	4,177	0,516	1,645	13,907	25,769	61,803	11,457	0,971	467,549	0,841	50,033	230,97	962,
7	6	0	2,04	0,391	0,999	0,488	1,911	1,067	2,566	1,105	2,246	11,947	13,41	75,721	10,035	0,834	534,974	0,937	34,197	183,675	1570,
8	7	0	0,042	-0,975	-0,451	-0,745	0,038	-0,334	-0,758	-1,539	-1,025	10,73	13,855	74,783	10,613	0,75	396,763	0,903	51,822	181,25	1611,
9	8	0	4,371	-2,458	0	0	1,468	0,039	0,049	0	0	4,371	0,376	99,624	0	0	0	0,336	0	451,694	4
10	9	0	3,899	1,43	1,821	0	2,569	1,2	1,864	0	1,864	3,899	0	100	0	0	0	0,659	0	413,22	1
11	10	0	1,264	0,726	0,798	0	0,424	0,268	0,883	0	3,486	8,836	29,808	62,476	7,666	0,05	1000	0,336	13,177	1000	2986,
12	11	0	0,032	0,001	0,006	0	0,034	0,66	1,887	0	0,46	1,027	16,551	82,452	0,995	0,003	1000	1,047	2,148	316,991	2001,
13	12	0	0,037	0,004	0,005	0	0,029	0,465	1,72	0	0,4	0,742	0,717	98,575	0,706	0,003	1000	0,779	0	461,785	4
14	13	0	0,11	0,024	0,012	0	0,039	0,617	2,561	0	0,322	2,065	0,17	97,865	1,957	0,008	1000	0,357	0	1000	4
15	14	0	17,69	3,5	2,428	0	6,064	1,2	10,009	0	67,243	21,903	32,264	62,593	5,117	0,025	1000	0,343	6,186	1000	367
16	15	0	30	20	-50	-50	1,716	1,719	3,238	-3,587	-2,764	60	30,677	69,318	0	0,006	1000	0,1	1000	1000	1291,
17	16	0	30	20	-1,751	-8,959	3,856	3,85	6,785	-20,548	-4,015	60	70	11,584	0	0,001	1000	0,1	1000	1000	604,
18	17	0	30	20	-60	-60	2,02	2,016	3,609	-32,046	-6,431	60	0,463	99,535	0	0,003	1000	0,1	1000	1000	4
19	18	0	30	20	-60	-60	2,071	2,068	3,934	-3,637	-1,277	60	0,271	99,728	0	0,002	0	0,1	1000	1000	4
20	19	0	30	20	-34,5	-58,802	1,884	1,881	4,038	-5,359	-3,157	60	0,004	99,991	0	0,005	1000	0,1	1000	1000	4
21	20	0	30	20	-23,448	-58,454	2,953	2,95	5,031	-6,133	-2,474	60	1,931	98,029	0	0,041	1000	0,1	1000	1000	4
22	21	0	20,793	20	-3,525	-4,409	3,779	3,778	5,591	-6,459	-5,15	20,793	0	99,978	0	0,022	1000	0,182	1000	34,702	1
23	22	0	6,742	6,271	6,017	3,808	9,063	8,754	16,116	12,645	19,979	7,964	65,897	32,554	1,307	0,241	159,739	1,344	62,405	167,408	220,

Indicatori

è importante che non vi siano celle vuote, né righe né colonne. La prima colonna indica il numero dell'osservazione, la seconda colonna indica invece il flag della società e quelle successiva rappresentano invece tutti gli indicatori utilizzati in questo lavoro di tesi.

2. Il software richiede che il file Excel contenente gli indicatori sia salvato in formato "CSV delimitato dal separatore di elenco".
3. Importare il file Excel degli indicatori in R mediante la seguente istruzione: **Database <- read.csv2 ("Indicatori.csv")**

4. La regressione logistica si definisce tramite il comando:

   modello <- glm(flag ~ ind_+ind_+ind_..., data=Database, family=binomial(link=logit))## link logit

   dove flag indicala variabile binaria che rappresenta se è avvenuto il default e ind_ è la notazione utilizzata per inserire l'indicatore nel modello. Ogniqualvolta si desideri inserire un altro indicatore è necessario utilizzare il segno di somma. Le successive istruzioni sono necessarie per specificare che la regressione è logistica.
5. Visualizzare i risultati della regressione in termini di coefficienti, intercetta e significatività statistica e altri paramenti della regressione utilizzando questo comando:

   summary(modello).

Una volta ottenuti i coefficienti degli indicatori e valutata la loro significatività, se ne analizza il segno algebrico. Nel modello si devono presentare solo indicatori con il segno algebrico corretto.

Successivamente si utilizza la stessa procedura impiegata per la valutazione dell'accuracy degli indicatori anche per il modello. In questo caso si utilizza il logit come un indicatore costituito dalla combinazione lineare degli altri indicatori.

Infatti il logit si costruisce come:

$$logit = costante + \beta_1 * ind_1 + \beta_2 * ind_2 + .. + \beta_n * ind_n$$

si calcola quindi per ogni osservazione il logit utilizzando i coefficienti risultanti dalla regressione e il valore degli indicatori calcolato. Si inserisce il valore di logit nel foglio di calcolo dell'accuracy e si ordina in modo crescente rispetto ad esso insieme alle variabili di flag e di osservazione. In questo caso è fondamentale l'ordinamento crescente del valore del logit in quanto un valore "elevato" di tale punteggio è indice di elevata probabilità di default, mentre un valore "ridotto" indica che la società sia sana.

Si inseriscono via via più indicatori nel modello fino a quando questi non aumentano più l'accuracy generale oppure finché non risultano significativi.

Non si devono inserire indicatori con segni algebrici non corretti ed è necessario valutare con attenzione il segno di un indicatore risultante dalla regressione.

Quando un indicatore non si presenta con il segno corretto o l'accuracy non aumenta più si interrompe l'inserimento degli indicatori.

Si ricorda come sia fondamentale inserire indicatori non correlati o per lo meno non fortemente correlati con quelli già presenti nel modello in quanto non solo non aggiungono informazione e la precisione del modello non aumenta ma soprattutto è possibile che non risultino più significativi gli altri indicatori già inseriti ed in particolare quello con elevata correlazione ad esso perda di significatività.

Tutte le procedure appena descritte si sono ripetute anche per la costruzione di modelli in cui il flag fosse relativo all'anno e non alla società.

Si noti che non si son utilizzati esclusivamente indicatori di bilancio o variabili di natura economico-finanziarie; infatti sono state utilizzate variabili come l'area geografica di appartenenza dell'impresa suddividendo il campione in nord-over, nord-est, centro, sud e isole. L'accuracy di tale indicatore si è pero rilevata bassa e, inserendo tale variabile nel modello, l'accuracy genere non è incrementata.

Inoltre, si è utilizzata anche una variabile relativa all'età dell'impresa calcolata come differenza tra l'anno relativo all'ultimo bilancio disponibile e la data di costituzione dell'impresa; anche in questo caso si è calcolata l'accuracy dell'indicatore età ma anch'esso non si è rilevato significativo nella costruzione del modello.

5 Illustrazione dei modelli di scoring

Si presentano in questo lavoro di tesi sette modelli di credit risk scoring: quattro si riferiscono al default della società e tre sono invece relativi al default anno ovvero si considera solo il bilancio in cui si è verificato il default.

Le procedure seguite per la costruzione dei modelli sono quelle riportate nel capitolo precedente.

Si descrive ora nel dettaglio ogni modello, riportando l'output statistico e il valore di accuracy calcolato per ognuno di essi.

5.1 I Modello società

Il primo modello società è costituito da 5 variabili di bilancio e si presenta come il modello con maggiore accuracy nella classe dei modelli per società. L'accuracy di tale modello infatti vale 72,00%.

Si riporta di seguito l'output statistico.

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.692117   1.032458   2.607  0.00912 **
ind35        0.016508   0.002659   6.208 5.35e-10 ***
ind7        -0.011940   0.002036  -5.865 4.49e-09 ***
ind24       -0.019586   0.004576  -4.280 1.87e-05 ***
ind42       -0.401789   0.062956  -6.382 1.75e-10 ***
ind9        -0.011525   0.005018  -2.297  0.02164 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1677.0  on 3678  degrees of freedom
Residual deviance: 1264.3  on 3673  degrees of freedom
AIC: 1276.3

Number of Fisher Scoring iterations: 7
```

Figura 5.1 I modello società

Le variabili utilizzate si riportano nella seguente tabella:

Tabella 5.1 Codifica degli indicatori del I modello società

Indicatore	Variabile economico-finanziaria
Ind35	$Passività\ correnti\ /_{Ricavi}$
Ind7	ROE
Ind24	$P.N.tangibile\ /_{Attivo\ netto}$
Ind42	$\ln(Ricavi)$
Ind9	$Val.aggiunto\ operativo\ /_{Ricavi}$

Il modello logit risultate dall'utilizzo di questi indicatore è dunque:

$$logit = 2{,}692117 + 0{,}016508 * \frac{PC}{Ric} - 0{,}011940 * ROE - 0{,}019586 * P.N.tang/AN - 0{,}401789 * \ln(Ric) - 0{,}011525 * Val.agg\ op/Ric$$

Si noti come i segni algebrici rispettino siano in accordo al significato economico-finanziario degli indicatori e del loro rapporto con la funzione logit. Infatti le variabili che aumentano la probabilità di default se sono elevate compaiono con un segno positivo mentre quelle che la riducono compaiono con il segno negativo.

Si sottolinea inoltre come tutte le variabili che compongono questo modello siano significative a livello α=5%: infatti il rapporto fra ogni coefficiente stimato e la relativa deviazione standard risultano essere maggiori del valore target relativo all'area di una normale standard al 95% (1-α) pari a 1,96. In modo analogo si osserva come il valore del *p-value* sia minore al 5% per ogni variabile.

Si noti inoltre come le variabili utilizzate nel modello siano tutte significative anche a livello α=1% a meno di *ind9* , poiché i valori della *zcalc* sono maggiori di 2,58, valore che corrisponde al quantile di una distribuzione normale standard 1-α=99%.

5.2 II Modello società

Il secondo modello costruito si costituisce di 5 variabili economico-finanziarie; l'accuracy di tale modello è del 70,68%.

Si riporta di seguito l'output statistico.

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.2796269  0.9756476   2.337  0.01946 *
ind42       -0.3888941  0.0611367  -6.361 2.00e-10 ***
ind35        0.0154442  0.0026705   5.783 7.32e-09 ***
ind22       -0.0231952  0.0045797  -5.065 4.09e-07 ***
ind8        -0.0058207  0.0022130  -2.630  0.00853 **
ind30        0.0019626  0.0006013   3.264  0.00110 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1677.0  on 3678  degrees of freedom
Residual deviance: 1270.6  on 3673  degrees of freedom
AIC: 1282.6

Number of Fisher Scoring iterations: 7
```

Figura 5.2 II modello società

Le variabili utilizzate si riportano nella seguente tabella:

Tabella 5.2 Codifica degli indicatori del II modello società

Indicatore	Variabile economico-finanziaria
Ind35	$Passività\ correnti\ /_{Ricavi}$
Ind8	$ROE\ ante\ imposte$
Ind22	$P.N./_{Attivo\ netto}$
Ind42	$\ln(Ricavi)$
Ind30	$Oneri\ finanziariiunto\ /_{Auto\ finan.lordo}$

Il modello logit risultate dall'utilizzo di questi indicatore è dunque:

$$logit = 2{,}2796269 - 0{,}3888941 * \ln(Ric) + 0{,}0154442 * \frac{PC}{Ric} - 0{,}0231952 * \frac{P.N.}{AN} - 0{,}0058207 * ROE\ ante\ imposte + 0{,}0019626 * \frac{\text{OF}}{\text{Autof}}.\text{lordo}$$

Si noti come i segni dei coefficienti siano corretti dal punto di vista concettuale.

Tutte le variabili sono statisticamente significative ad un livello di α=5% e, ad eccezione del coefficiente dell'intercetta, tutte le restanti variabili sono significative anche un livello di α=1%.

5.3 III Modello società

Il terzo modello società si presenta come una combinazione lineare di 4 variabili economico-finanziarie; l'accuracy di questo modello risulta essere del 68,45%.

Si riporta di seguito l'output statistico.

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  4.6145788  0.8632739   5.345 9.02e-08 ***
ind42       -0.4830963  0.0581937  -8.302  < 2e-16 ***
ind36        0.0014875  0.0003904   3.810 0.000139 ***
ind22       -0.0257102  0.0044885  -5.728 1.02e-08 ***
ind8        -0.0127628  0.0016775  -7.608 2.78e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1677.0  on 3678  degrees of freedom
Residual deviance: 1310.6  on 3674  degrees of freedom
AIC: 1320.6

Number of Fisher Scoring iterations: 6
```

Figura 5.3 III modello società

Le variabili utilizzate si riportano nella seguente tabella:

Tabella 5.3 Codifica degli indicatori del III modello società

Indicatore	Variabile economico-finanziaria
Ind42	$\ln(Ricavi)$
Ind36	$Debiti\ finanz.stimati\ /_{Valore\ aggiunto}$
Ind22	$P.N./_{A.N.}$
Ind8	$ROE\ ante\ imposte$

Il modello logit risultate dall'utilizzo di questi indicatore risulta essere:

$$logit = 4{,}6145788 - 0{,}4830963 * \ln(Ric) + 0{,}0014875 * Debiti\ finanz.stimati\ /_{Valore\ aggiunto} - 0{,}0257102 * P.N./_{A.N.} - 0{,}0127628 * ROE\ ante\ imposte$$

Anche questo output statistico relativo alla regressione logistica sulle varabili del III modello, evidenza la significatività statistica e la correttezza del segno dei coefficienti del modello.

Si sottolinea che in questo modello tutti i coefficienti risultano significativi al 5%, all'1% e anche all'1‰. In fatti tutti i valori di *zcalc* sono abbondantemente maggiori di 1,96 e 2,58 *e* di conseguenza anche i *p-value* sono notevolmente inferiori a 0,05 e 0,01, in alcuni casi sono inoltre tendenti a zero.

5.4 IV Modello società

L'ultimo modello che si presenta relativamente al default società, si configura come un modello costituito da 5 variabili economico-finanziarie. L'accuracy di tale modello è del 65,98%.

Si riporta di seguito l'output statistico.

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.4724187  0.9241806   3.757 0.000172 ***
ind25       -0.0225133  0.0041576  -5.415 6.13e-08 ***
ind41       -0.4131024  0.0635291  -6.503 7.90e-11 ***
ind27        0.0016864  0.0006192   2.723 0.006461 **
ind37        0.0056231  0.0017205   3.268 0.001082 **
ind8        -0.0100387  0.0022655  -4.431 9.38e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1677.0  on 3678  degrees of freedom
Residual deviance: 1331.2  on 3673  degrees of freedom
AIC: 1343.2

Number of Fisher Scoring iterations: 6
```

Figura 5.4 IV modello società

Le variabili utilizzate si riportano nella seguente tabella:

Tabella 5.4 Codifica degli indicatori IV modello società

Indicatore	Variabile economico-finanziaria
Ind25	$\frac{\text{P. N. tangibile}}{\text{Debiti} + \text{P. N.}}$
Ind41	$\ln(Attivo\ netto)$
Ind27	Oneri finanziari netti/EBITDA
Ind37	Debiti finanziari stimati/Ricavi

Ind8	$ROE\ ante\ imposte$

Il modello logit risultate dall'utilizzo di questi indicatore risulta essere:

$$logit = 3{,}4724187 - 0{,}0225133 * \frac{P.N.tangibile}{debiti + P.N.} - 0{,}4131024 * \ln(AN) + 0{,}0016864 * \frac{OFN}{EBITDA} + 0{,}0056231 * \frac{Deb\ finanz\ stiamti}{Ric} - 0{,}0100387 * ROE\ ante\ imposte$$

Dall'osservazione dell'output è evidente come, anche in questo modello, tutti i coefficienti dono corretti dal punto di vista algebrico e significativi dal punto di vista statistico anche ad un livello di significatività α=1%.

5.5 I Modello anno

Si analizzano ora i modelli di scoring costruiti con il flag anno.

Si ricorda che per flag anno si intende come anomala l'unica osservazione relativa al singolo anno in cui si verifica l'evento di default e non si considerano come anomali tutti i bilanci di quella società.

Il modello sotto riportato ha un'accuracy pari al 71,87%.

Si riporta l'output statistico relativo al primo modello di scoring.

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  6.818013   2.181058   3.126  0.00177 **
ind42       -0.817719   0.156463  -5.226 1.73e-07 ***
ind28        0.002090   0.001133   1.845  0.06497 .
ind23       -0.024189   0.012493  -1.936  0.05284 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 348.31  on 3678  degrees of freedom
Residual deviance: 288.62  on 3675  degrees of freedom
AIC: 296.62

Number of Fisher Scoring iterations: 9
```

Figura 5.5 I modello anno

Le variabili utilizzate si riportano nella seguente tabella:

Tabella 5.5 Codifica degli indicatori I modello anno

Indicatore	Variabile economico-finanziaria
Ind42	$\ln(Ricavi)$
Ind28	$Oneri\ finanziari\ netti / EBIT$
Ind23	$Risultato\ utile / A.N.$

Il modello logit risultate dall'utilizzo di questi indicatore è il seguente:

$$logit = 6{,}818013 - 0{,}817719 * \ln(Ric) + 0{,}002090 * \frac{OFN}{EBIT} - 0{,}024189 * Ris\ utile/AN$$

Si noti come il segno dei coefficienti sia corretto e rispecchia il significato economico-finanziario dei vari indicatori.

Si sottolinea inoltre che i coefficienti relativi alle le variabili $Oneri\ finanziari\ netti/EBIT$ e $Risultato\ utile/A.N.$ non sono significativi ad un livello α=5%; infatti rapportando il valore stimato dei coefficienti ai relativi errori standard si ottengono valori di *zcalc* inferiori a 1,96 in valore assoluto (rispettivamente 1,845 e -1,936) e di conseguenza i valori del p-value corrispondenti sono maggiori al 5% (rispettivamente 6,497% e 5,284%). Si può dunque affermare che tali variabili siano sì significative , ma ad un livello di significatività α=10%.

5.6 II Modello anno

Si presenta ora il secondo modello relativo al flag anno: si costituisce da due soli variabili. L'accuracy di tale modello è del 69,88%.

Si riporta di seguito il relativo output statistico.

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  6.5122326  2.4103946   2.702   0.0069 **
ind41       -0.8271214  0.1713203  -4.828 1.38e-06 ***
ind27        0.0043835  0.0009986   4.390 1.14e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 348.31  on 3678  degrees of freedom
Residual deviance: 290.85  on 3676  degrees of freedom
AIC: 296.85

Number of Fisher Scoring iterations: 8
```

Figura 5.6 II modello anno

Le variabili utilizzate si riportano nella seguente tabella:

Tabella 5.6 Codifica degli indicatori II modello anno

Indicatore	Variabile economico-finanziaria
Ind41	$\ln(Attivo\ netto)$
Ind27	$Oneri\ finanziari\ netti / EBITDA$

Il modello logit risultate dall'utilizzo di questi indicatore è il seguente:

$$logit = 6{,}5122326 - 0{,}8271214 * \ln(AN) + 0{,}0043835 * \frac{OFN}{EBITDA}$$

in questo modello, costituito da due sole variabili, si nota come i coefficienti delle variabili stimate ln (Attivo netto) e Oneri finanziari netti/EBITDA siano entrambi corretti dal punto di vista del segno algebrico e molto significativi. Infatti il p- value di entrambi i coefficienti risulta essere molto minore al 5% e all'1%.

5.7 III Modello anno

Si presenta ora l'ultimo modello relativo al flag anno: si costituisce di tre variabili. L'accuracy di tale modello è del 70,80%.

Si riporta di seguito il relativo output statistico.

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  6.490223   2.252240   2.882  0.00396 **
ind42       -0.791870   0.160649  -4.929 8.26e-07 ***
ind25       -0.016457   0.008463  -1.945  0.05182 .
ind27        0.003168   0.001077   2.942  0.00326 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 348.31  on 3678  degrees of freedom
Residual deviance: 284.05  on 3675  degrees of freedom
AIC: 292.05

Number of Fisher Scoring iterations: 9
```

Figura 5.7 III modello anno

Le variabili utilizzate si riportano nella seguente tabella:

Tabella 5.7 Codifica degli indicatori III modello anno

Indicatore	Variabile economico-finanziaria
Ind42	$\ln(Attivo\ netto)$
Ind27	$Oneri\ finanziari\ netti / EBITDA$
Ind25	$Patrimonio\ netto\ tang / Debiti + Patrimonio\ netto$

Il modello logit risultate dall'utilizzo di questi indicatore risulta essere:

$$logit = 6{,}490223 - 0{,}791870 * \ln(Ric) - 0{,}016457 * \frac{P.N.tang}{Debiti + P.N.} + 0{,}003168 * \frac{OFN}{EBITDA}$$

Si può notare come i coefficienti delle variabili utilizzate siano algebricamente corretti e significativi dal punto di vista statistico a livello di significatività α=5%
a meno del coefficiente della variabile Patrimonio netto tangibile /(debiti +patrimonio netto) che presenta un p-value leggermente maggiore al 5%, pari infatti a 5,182%. Si può dunque affermare che tale coefficiente sia significativo a livello α=10%.

6 Analisi delle performance e degli errori dei modelli

6.1 Performance dei modelli

In questo capitolo si intende descrivere e fornire una spiegazione dell'analisi delle performance dei modelli di scoring. In particolare si intende analizzare gli errori dei modelli costruiti evidenziando dunque le società che sono state classificate in modo non corretto dal modello di regressione logistica.
Si classificano dapprima le società in quattro gruppi o cluster. Si riporta nella seguente tabella la suddivisone.

Tabella 6.1 Classificazione degli errori nei modelli

		Classificazione	
Situazione reale		**Società sana**	**Società anomala**
	Società sana	Corretto	Errore di II tipo
	Società anomala	Errore di I tipo	Corretto

Ovviamente se il modello classifica una società che non ha subito un evento di default in una delle osservazioni riportate, come società sana in funzione delle variabili utilizzate in quel modello, allora la classificazione è corretta. Un'analisi identica si effettua quando la società risulta essere anomala e il modello correttamente la classifica come anomala.
Vi sono invece due casi di classificazione errata, si tratta infatti dei casi in cui il modello classifica una società come anomala quando in realtà tale società risulta sana e come

sana una società che invece risulta essere anomala. Nel primo caso di società anomala classificata come sana si parla di errore di I tipo, mentre nel secondo caso di parla di errore di II tipo.

Un secondo metodo di valutazione della precisone del modello e dell'analisi degli errori è l'accuracy detta anche *curva Gini, Power curve o Lorenz curve*.
Si riporta un esempio di seguito in figura 6.1.

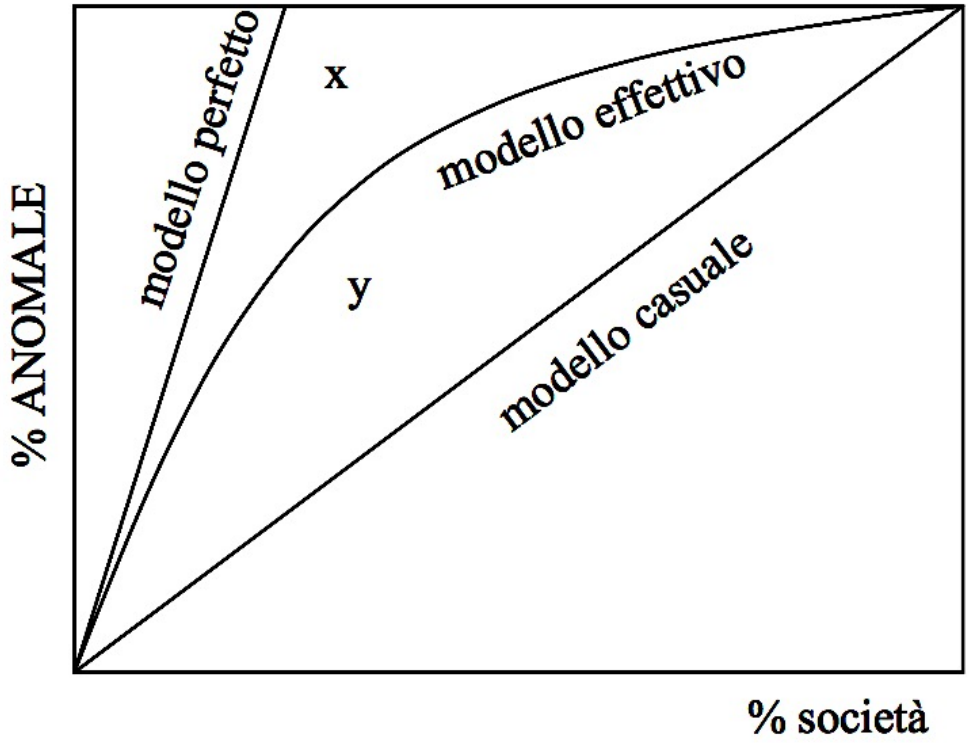


Figura 6.1 Curva Gini

Come si osserva dalla figura 6.1 il "modello perfetto" identifica in modo corretto tutte le società anomale. Il modello casuale invece attribuisce a tutte le società la stessa probabilità di default, di conseguenza non fornisce alcun tipo di informazione utile. Il modello effettivo è invece il modello risultate dall'analisi dei coefficienti ottenuti mediante la regressione logistica.

Si può calcolare l'accuracy, detta anche *"Gini coefficient"* cioè il rapporto fra le osservazioni che ricadono nella zona delimitata da *y* e il totale delle osservazioni *x* e *y*:

$$Accuracy = \frac{y}{x + y}$$

Il valore del rapporto è compreso tra 0 e 1.

Un altro strumento utilizzato nella valutazione delle performance dei modelli è la cosiddetta curva ROC ovvero *Receiving Operating Characteristic.* Concettualmente tale curva è simile all'accuracy. Si riporta di seguito una figura illustrativa.

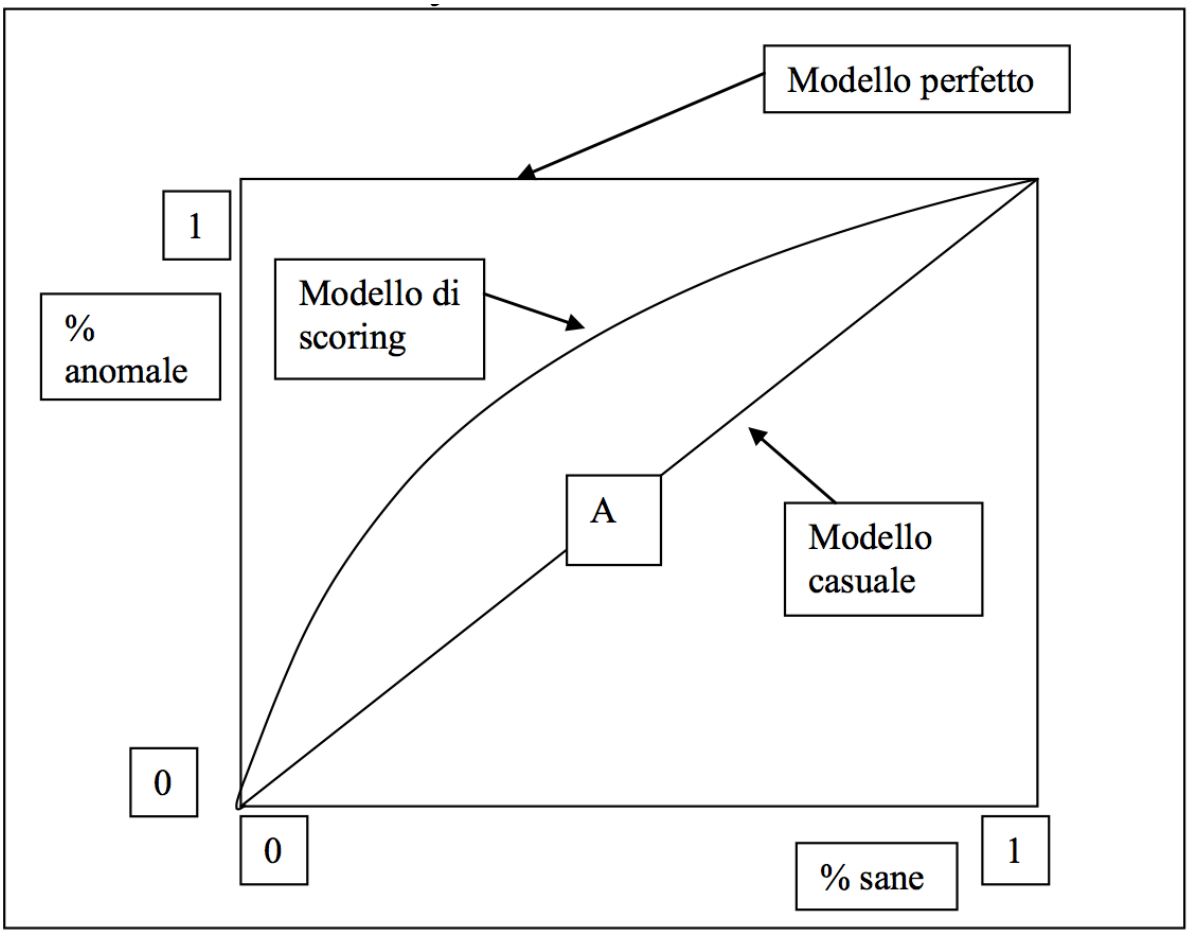


Figura 6.2 La curva ROC.

La retta diagonale indica un modello di scoring prettamente casuale mentre il segmento superiore che delimita il grafico e contrassegnato come modello perfetto indica invece il modello senza alcun errore. L'area di grafico indicata con la lettera A è invece la parte del grafico compresa tra il modello effettivo e la parte inferiore del

grafico. L'area tra il segmento inferiore ed il modello casuale vale 0.5. Se il sistema di scoring fosse perfetto l'area A sarebbe uguale ad 1. I sistemi realistici hanno quindi un'area compresa tra 0.5 ed 1 e sono tanto migliori quanto più sono prossimi ad 1. Esiste una banale equazione che relaziona l'accuracy precedentemente descritta alla curva ROC ed è la seguente:

$$Accuracy = 2 * ROC - 1 = 2 * (ROC - 0{,}5)$$

Si riportano di seguito in figura la classificazione degli errori descritta in precedenza nell'utilizzo del modello ROC.

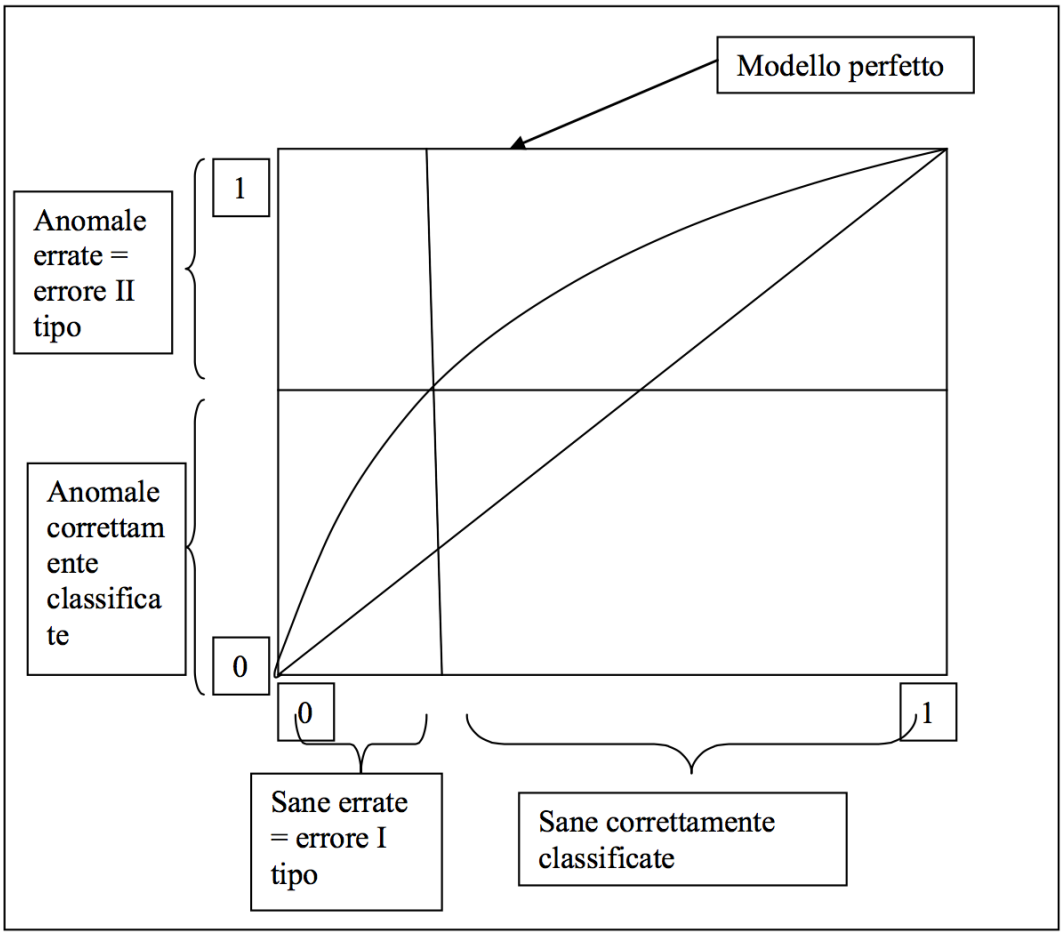


Figura 6.3 La valutazione degli errori nella curva ROC.

Come si può notare in figura, si evidenziano i quattro cluster in cui si suddivide il campione. Sane correttamente classificate nel riquadro a destra, anomale correttamente classificate nel riquadro in basso a sinistra, anomale errate in quanto società sane che sono state classificate come anomale (errore del II tipo) nel riquadro in alto a sinistra ed infine una piccola area sempre in basso a sinistra dove ricadono le osservazioni di società anomale classificate invece come sane (errore del I tipo).

Si riportano di seguito le curve di accuracy dei due migliori modelli di cui si intende valutare performance ed errori di classificazione.

- I modello società: accuracy pari al 72%.

$$logit = 2{,}692117 + 0{,}016508 * \frac{PC}{Ric} - 0{,}011940 * ROE - 0{,}019586 * P.N.tang/AN - 0{,}401789 * \ln(Ric) - 0{,}011525 * Val.agg\ op/Ric$$

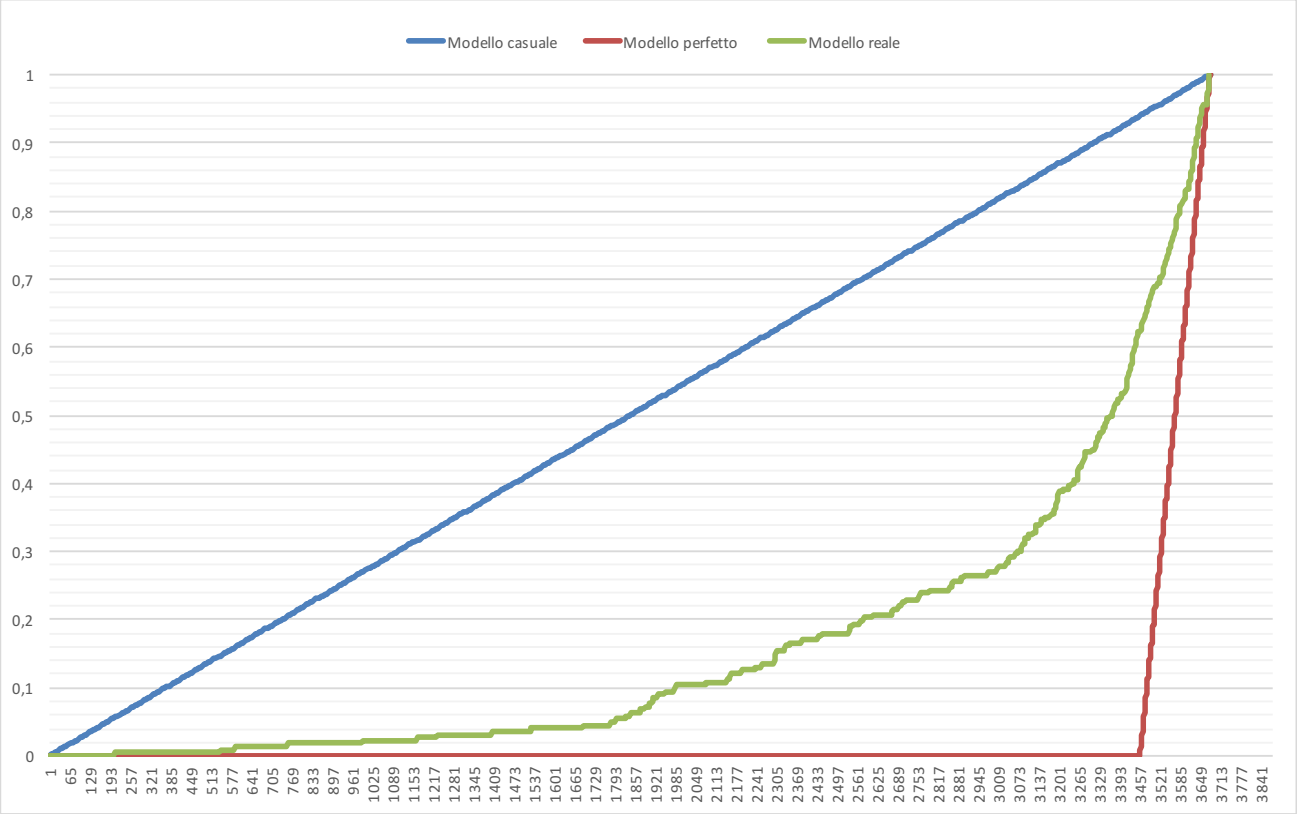


Figura 6.4 Curva accuracy I modello società

- I modello anno: accuracy pari al 71,87%.

$$logit = 6{,}818013 - 0{,}817719 * \ln(Ric) + 0{,}002090 * \frac{OFN}{EBIT} - 0{,}024189 * Ris\ utile/AN$$

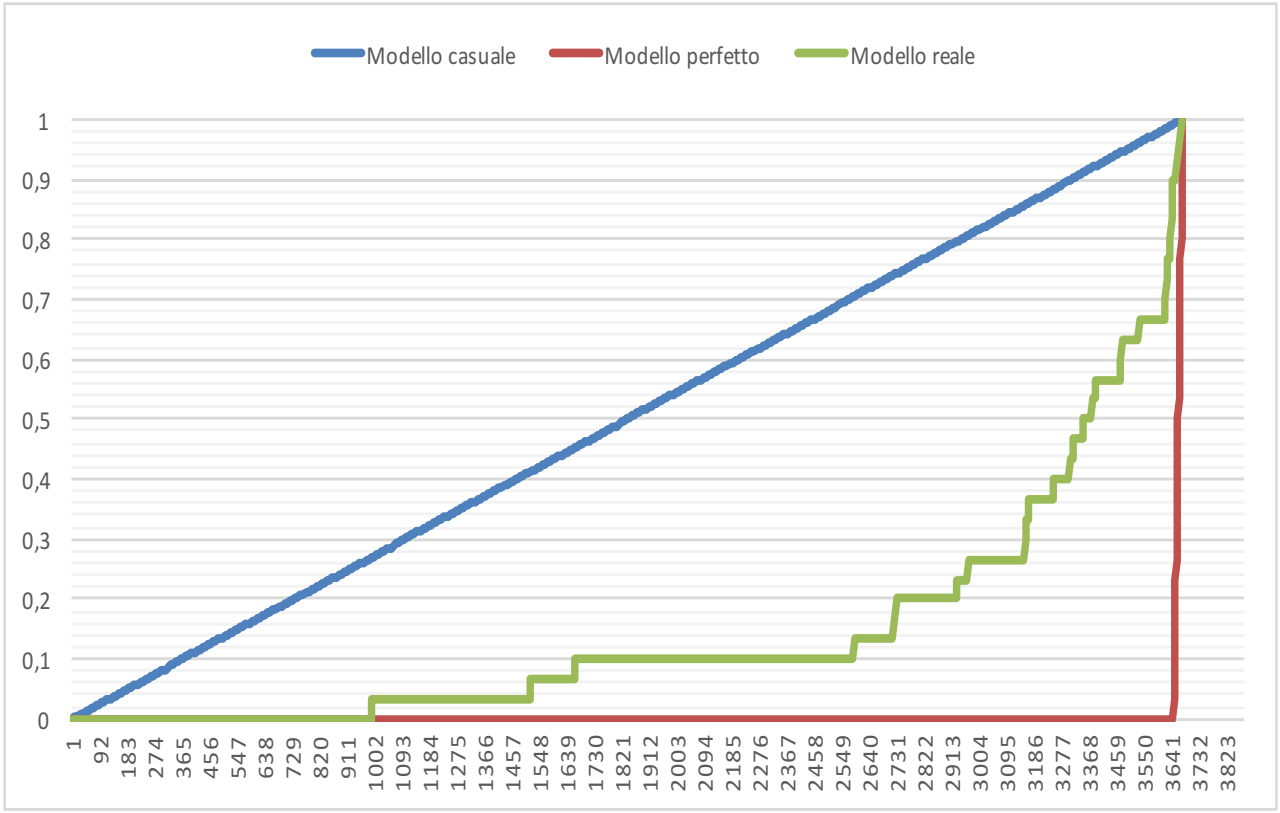


Figura 6.5 Curva accuracy I modello anno

6.2 Valutazione degli errori dei modelli

Una volta analizzate le performance dei modelli si sono inoltre analizzati gli errori di classificazione dei modelli di scoring, errori di I o II tipo sopra descritti.
Per valutare gli errori dei modelli si sono suddivise le osservazioni totali in quattro gruppi distinti: società sane correttamente classificate, società sane erroneamente classificate, società anomale erroneamente classificate e società anomale correttamente identificate.
Successivamente si sono valutate le differenze tra le medie di variabili non facenti parte del modello e, comunque, poco correlate con esso. Si suddivide in due differenti analisi lo studio degli errori per il miglior modello società e per il miglior modello anno. In ogni caso si utilizzano diverse variabili per valutare la differenza tra le medie di società sane correttamente ed erroneamente classificate e la differenza tra le medie di società anomale correttamente ed erroneamente classificate.

6.3 Analisi errori I modello società

Al fine di valutare gli errori nel miglior modello società si sono analizzate le seguenti variabili:

- AC/PC

MEDIA SANE CORRETTE	204,20
MEDIA SANE ERRATE	122,93
MEDIA ANOMALE ERRATE	201,89
MEDIA ANOMALE CORRETTE	115,05

Si può notare come, usando l'indicatore AC/PC sia evidente la differenza tra le medie di società sane correttamente classificate e società sane erroneamente classificate (204,20 vs. 122,93) ed inoltre la differenza tra le società anomale errate e quelle

correttamente classificate dal modello (201,89 vs. 115,05). La differenza tra le classi si evidenzia inoltre tramite un istogramma di seguito riportato.

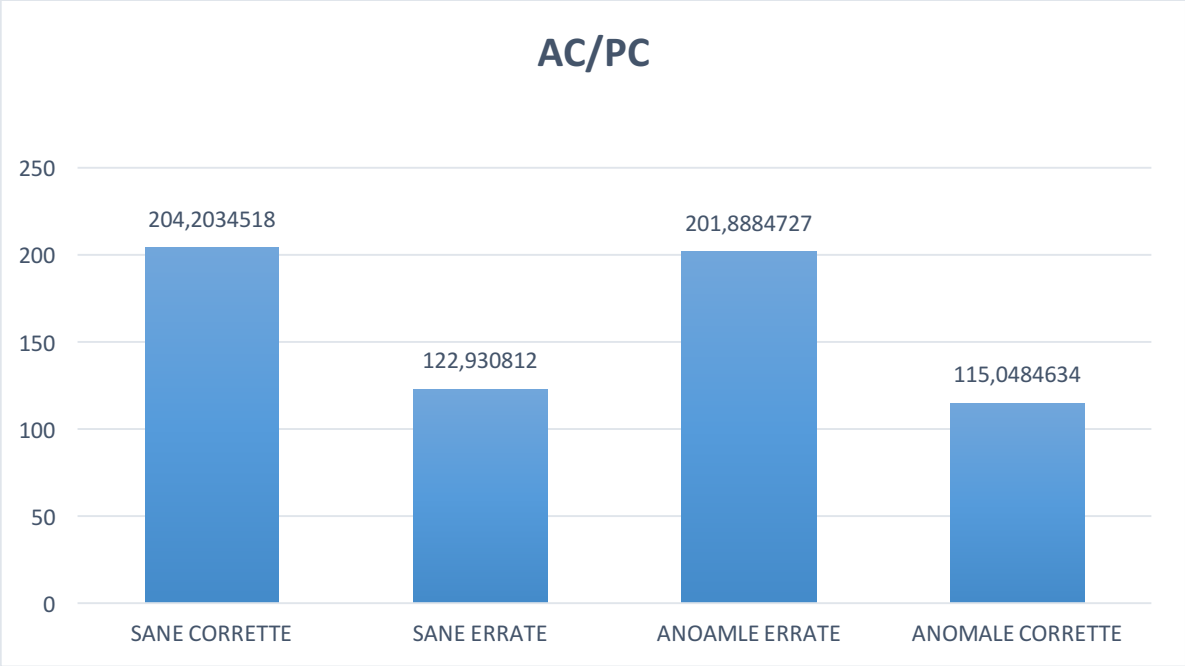


Figura 6.6 Differenza tre le medie dei gruppi di società con l'indicatore AC/PC.

- OFN/EBIT

MEDIA SANE CORRETTE	72,98
MEDIA SANE ERRATE	201,99
MEDIA ANOMALE ERRATE	70,06
MEDIA ANOMALE CORRETTE	288,03

Un altro indicatore utilizzato per l'analisi degli errori del modello default società è il rapporto tra gli oneri finanziari netti e l'EBIT.

Si nota in questo caso come sia accentuata la differenza tra le medie del gruppo di società sane correttamente classificate e sane anomale (72,98 vs. 201,99) e tra le società anomale errate e le società anomale correttamente classificate (70,06 vs. 288,03).

S i riporta la classificazione anche dal punto di vista grafico.

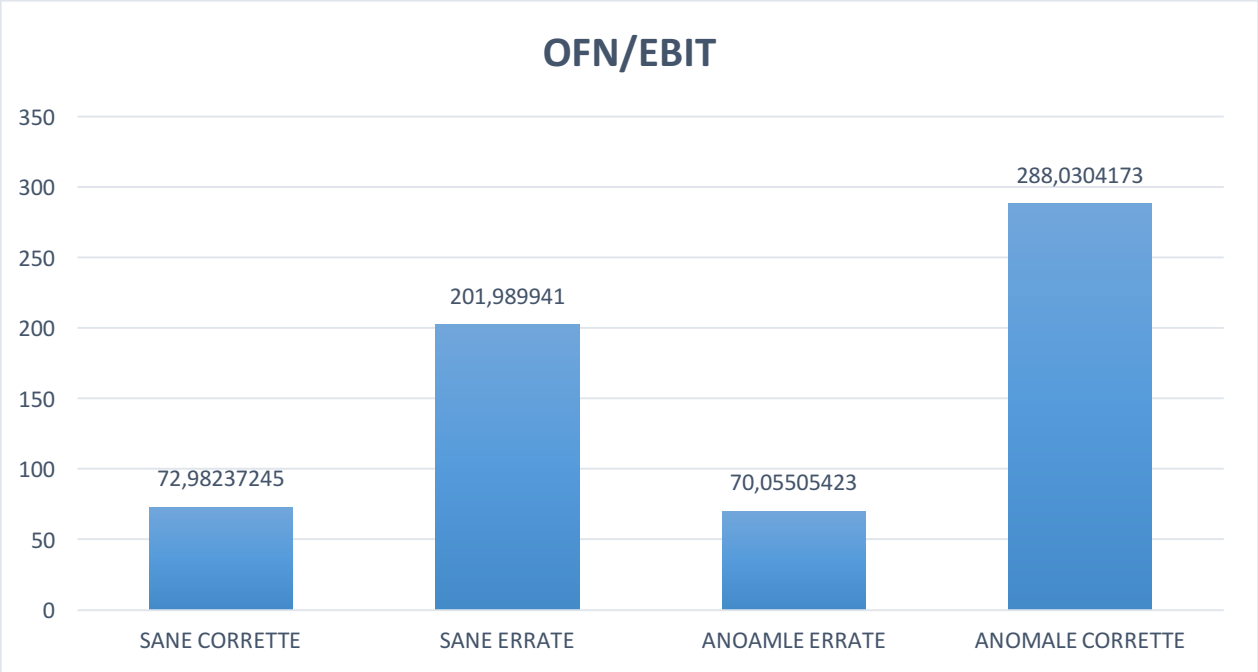


Figura 6.7 Differenza tre le medie dei gruppi di società con l'indicatore OFN/EBIT.

- Debiti totali/EBITDA

MEDIA SANE CORRETTE	67,28
MEDIA SANE ERRATE	327,58
MEDIA ANOMALE ERRATE	99,55
MEDIA ANOMALE CORRETTE	532,77

Si mette ora in evidenza le differenze tra le medie dei quattro gruppi di società sfruttando l'indicatore Debiti totali/EBITDA.

La media del valore dell'indicatore Debiti totali/EBITDA delle società sane correttamente classificate dal modello ammonta a 67,28 mentre quella delle società sane erroneamente classificate ammonta a 327,58. La media invece di tale variabile per le società anomale errate risulta essere pari a 99,55 contro il valore di 532,77 media delle società anomale correttamente classificate dal modello.

Si riporta di seguito un istogramma che evidenzia i concetti appena spiegati.

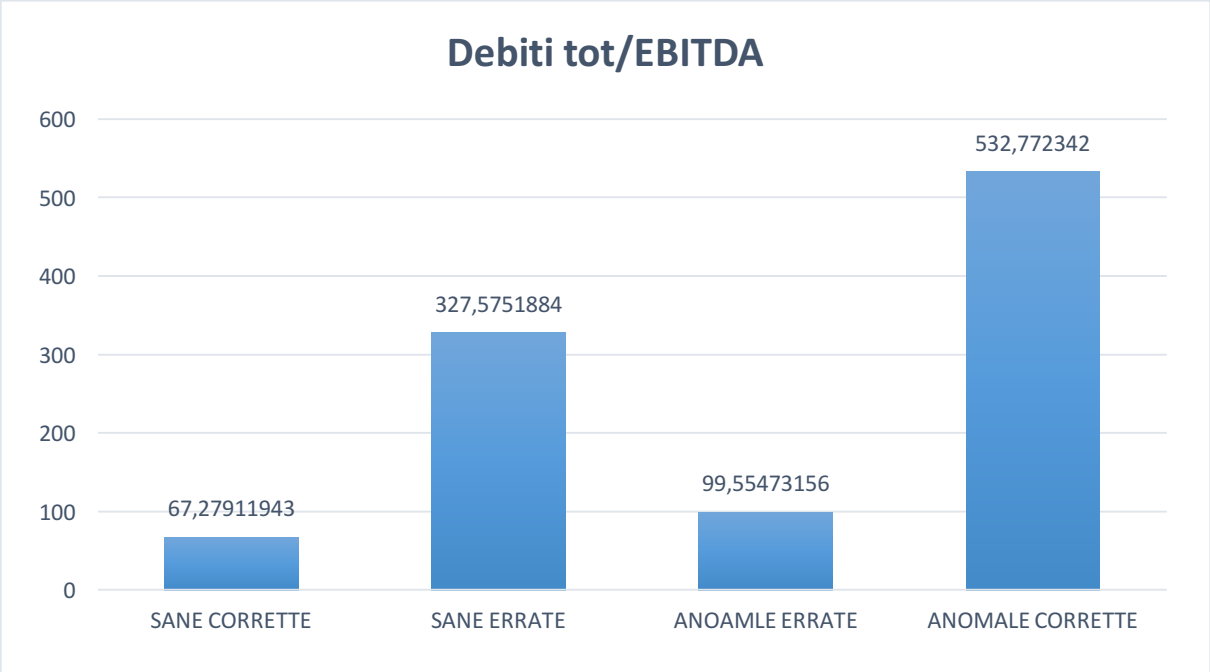


Figura 6.8 Differenza tre le medie dei gruppi di società con l'indicatore Debiti totali/EBITDA.

- EBITDA-servizio debito/AN

MEDIA SANE CORRETTE	4,52
MEDIA SANE ERRATE	-4,27
MEDIA ANOMALE ERRATE	2,73
MEDIA ANOMALE CORRETTE	-7,49

Un altro indicatore utilizzato per l'analisi degli errori del modello default società è il rapporto tra i debiti totali e l'EBITDA.

Si può notare in questo caso come ancor più siano separate le differenze tra le medie dei quattro gruppi di società. Infatti la media delle società sane correttamente identificate ammonta a 4,52 mentre quelle delle società sane erroneamente identificate assume valore negativo e in particolare risulta pari a -4,27. La differenza tra le medie dei due gruppi è dunque molto accentuata.

Un ragionamento analogo si può effettuare sulla differenza tre le medie delle società anomale: la media delle società anomale errate ammonta a 2.73 mentre quelle delle società anomale correttamente identificate dal modello ammonta invece a -7,49.
Si riporta qui di seguito il grafico che mette in luce le differenze tra i gruppi di sane anomale.

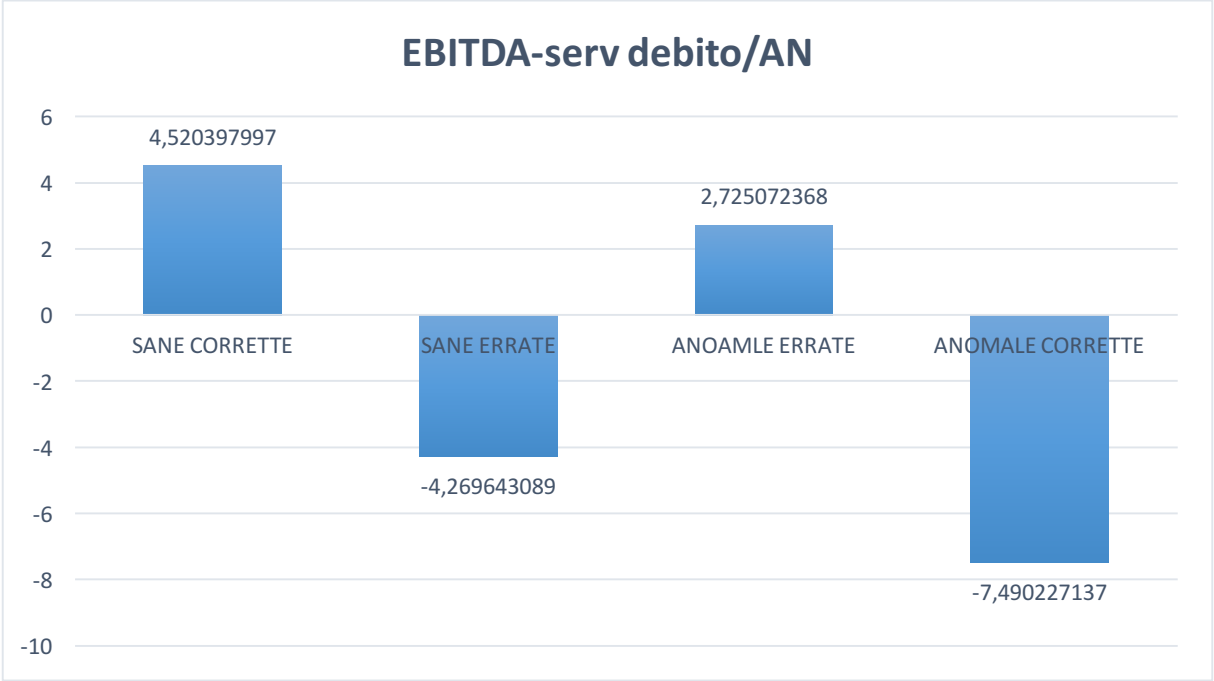


Figura 6.9 Differenza tre le medie dei gruppi di società con l'indicatore EBITDA-servizio debito/AN.

6.4 Analisi errori I modello anno

Per valutare gli errori dei miglior modello anno, invece, si procede in modo leggermente diverso: in questa caso si sono analizzati separatamente indicatori per la valutazione della differenza tra le medie di società sane corrette e società sane anomale, tra media di società errate corrette e media delle società anomale errate.

- AC/PC

MEDIA ANOMALE CORRETE	246,66
MEDIA ANOMALE ERRATE	114,89

Si può notare come la media dell'indicatore di liquidità AC/PC sia nettamente diversa tra i due gruppi di società. Infatti la media dell'indicatore per le società anomale corrette vale 246,66 mentre la media dello stesso per quelle classificate come sane dal modello vale invece 114,89.

Si riporta qui di seguito il grafico che mette in luce le differenze tra i gruppi di anomale correttamente classificate e anomale erroneamente classificate.

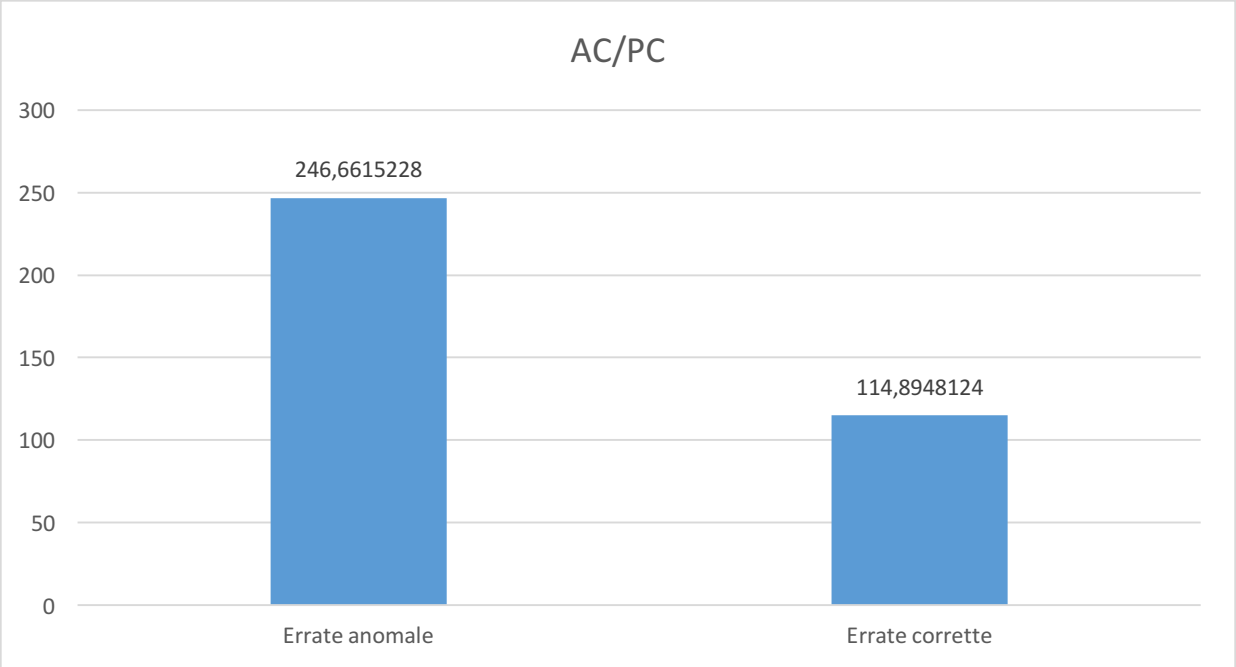


Figura 6.10 La differenza tra medie nelle società anomale errate ed anomale corrette sfruttando l'indicatore AC/PC.

- Debiti finanziari stimati/VA

MEDIA ANOMALE CORRETE	114,23
MEDIA ANOMALE ERRATE	220,85

In questo caso si è utilizzato invece l'indicatore Debiti finanziari stimati/Valore aggiunto per spiegare la differenza tra i gruppi di società anomale erroneamente classificate e società anomale correttamente classificate. La netta differenza tra le medie dei due gruppi è facilmente apprezzabile 114,23 è la media delle anomale corrette contro 220,85 la media invece delle società anomale errate.
Si riporta anche in questo caso il grafico relativo.

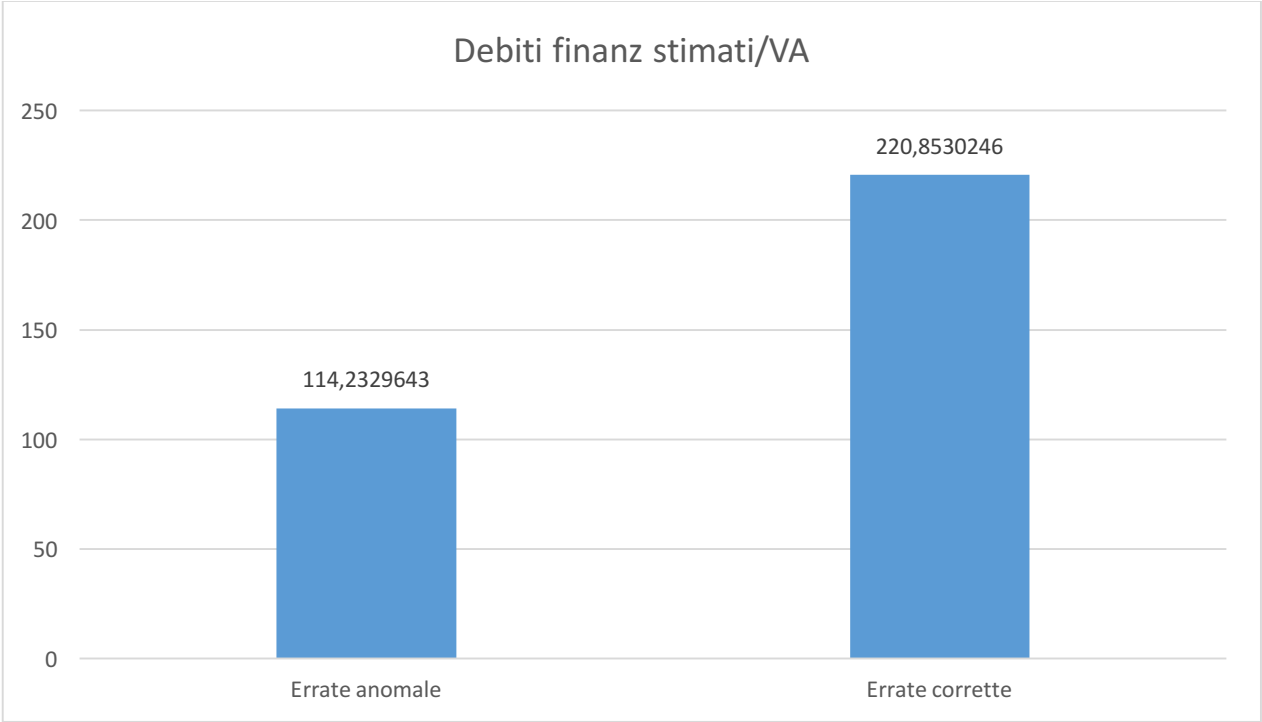


Figura 6.11 Differenza tra le medie di società anomale correttamente ed erroneamente classificate con l'indicatore debiti finanziari stimati/VA.

- Debiti totali/EBITDA

MEDIA SANE CORRETE	131,88
MEDIA SANE ERRATE	143,47

Si utilizza l'indicatore debiti totali/EBITDA per valutare invece la differenza tra medie di società sane che sono state correttamente classificate e società sane invece erroneamente classificate. In questo caso la differenza tra i due gruppi non è molto notevole, fatto dovuto principalmente al numero limitato di osservazioni.
Si riporta il grafico relativo alla differenza tra i due gruppi.

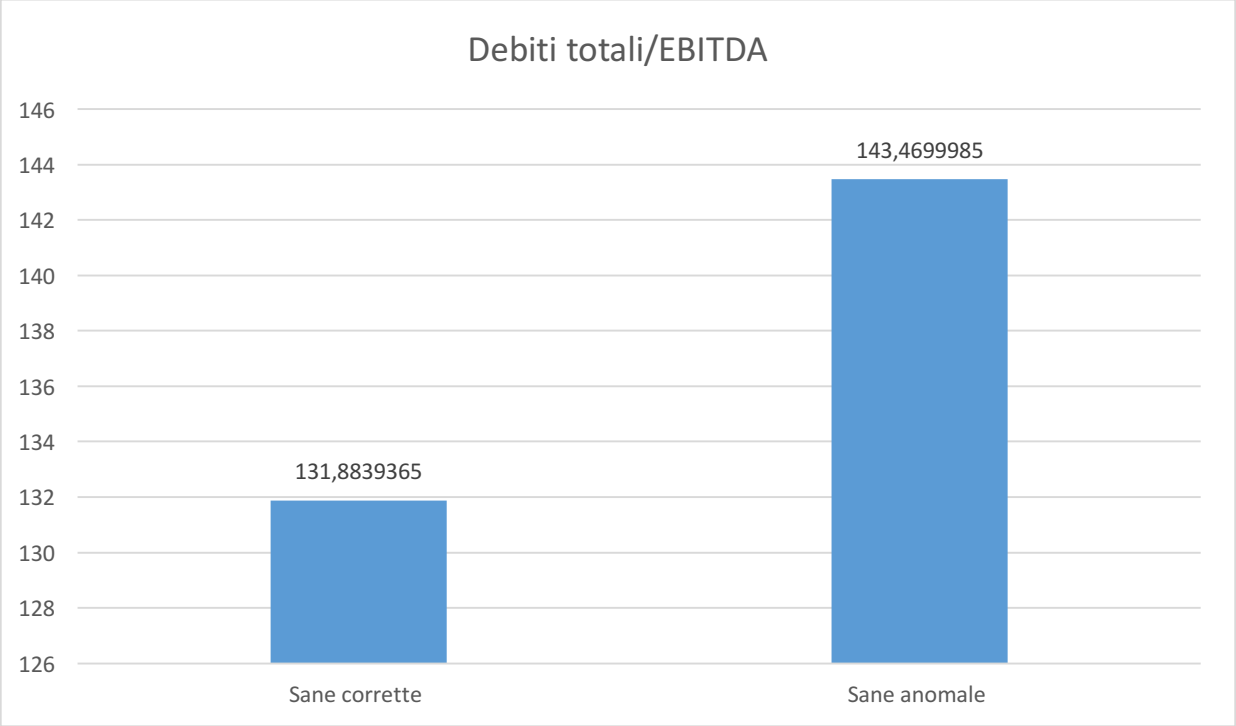


Figura 6.12 La differenza tra le medie di società sane correttamente classificate e società sane anomale usando l'indicatore Debiti totali/EBITDA.

Conclusioni

In questo lavoro di tesi si sono costruiti ed analizzati modelli di risk scoring al fine di valutare la rischiosità creditizia nel settore tessile e in particolare nella tessitura italiana. A tal fine si è deciso di utilizzare la statistica per la costruzione dei modelli e in particolare lo strumento della regressione logistica attuata tramite il software statistico *"R"*.

I sette modelli costruiti, quattro relativi al default della società e tre relativi invece al default anno, si basano solamente su variabili di tipo economico finanziarie, indici di bilancio ricavabili esclusivamente da voci di stato patrimoniale e conto economico. Si sono utilizzate anche altre variabili non appartenenti a tale categoria, variabili di tipo demografico come la zona geografica di appartenenza della società (Nord-Ovest, Nord-Est, Centro, Sud e isole) ed inoltre la variabile anagrafica relativa all'età dell'impresa ma entrambe le variabili non sono risultate significative alla fine della costruzione del modello.

Inoltre si sottolinea la scarsa qualità dei dati in ingresso provenienti dal database AIDA, per i quali molti valori risultavano assenti, alcune informazioni mancanti ed errori nei dati di bilancio. Tale aspetto, insieme alla mancanza di un numero elevato di osservazioni della tessitura, ha portato alla costruzione di modelli con un accuracy discreta ma che potrebbe essere migliorata. Il miglioramento si potrebbe ottenere inserendo non solo variabili di tipo economico-finanziario, anagrafiche, demografiche ecc. ma inserendo anche variabili di relative per esempio al livello di qualità dei prodotti, al livello di competitività sui prezzi e il grado di competitività sulle quantità.

Per la valutazione della bontà ei modelli si è utilizzata l'accuracy, una sorta di precisone percentuale rispetto al totale delle osservazioni: ciò significa che l'accuracy è tanto maggiore quanto il modello identifica correttamente le osservazioni raggruppandole nelle classi di appartenenza di società sane o società anomale. Si ricorda però che l'accuracy è una statistica di tipo non parametrico cioè dipendente essenzialmente dall'ordinamento crescente o decrescente delle osservazioni, nettamente differente dalla statistica parametrica. Quest' ultima infatti ipotizza di conoscere la legge di probabilità che governa la popolazione ameno

di alcuni parametri come la media e la varianza ad esempio. Nella statistica non parametrica invece non si effettuano alcun tipo di ipotesi a priori sulle caratteristiche della popolazione (o comunque si effettuano ipotesi meno restrittive rispetto a quelle tipiche della statistica parametrica).

SItografia

- http://www.schededigeografia.net/Italia/Economia/industria_tessile_abbigliamento_italia.htm

- https://it.wikipedia.org/wiki/Industria_tessile

- https://www.google.it/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=2ahUKEwjGyK3ys4DlAhUMPVAKHWvlBugQFjAAegQIARAC&url=https%3A%2F%2Fwww.sistemamodaitalia.com%2Fit%2Farea-associati%2Fcentro-studi%2Fitem%2Fdownload%2F11440_07f3cff442f5c95f99598384e817a361&usg=AOvVaw0-r9Nnp7vrH5GOyog4yVbm

- http://www.treccani.it/enciclopedia/l-industria-tessile_%28Il-Contributo-italiano-alla-storia-del-Pensiero:-Tecnica%29/

- http://www.treccani.it/enciclopedia/tessitura

- https://www.prometeia.it/atlante/tessile-italiano-a-10-anni-crisi

Bibliografia

David W. Hosmer, Stanley Lemeshow, “Applied Logistic Regression”, Second Edition,John Wiley & Sons Inc., 2000

Andrea Resti, Andrea Sironi, “Risk Management and Shareholders’ Value in Banking: From risk Measurement Models to Capital Allocation Policies, John Wiley & Sons Ltd, England, 2007