

# CERVED GROUP SpA

Rating Validation

---

## Indice

### Parte I - Approccio Teorico

1.1 - Introduzione .....	4
1.2 - Approccio metodologico .....	5
1.3 - Calcolo dell'Accuracy Ratio e del Area Under Curve .....	8
1.3.1 - Il calcolo del Accuracy Ratio e la rappresentazione del Cumulative Accuracy Profile .....	8
1.3.2 - Il calcolo del Area Under Curve e la rappresentazione del Receiver Operating Characteristic .....	9
1.4 - Backtesting .....	11
1.4.1 - La metodologia Bootstrap .....	11
1.4.1.1 - La distorsione .....	12
1.4.1.2 - L'errore standard.....	13

### Parte II - Approccio Empirico

2.1 - Descrizione dei dati.....	16
2.2 - Modelli di valutazione (Testing Rating Accuracy) .....	16
2.3 - Backtesting.....	18

Bibliografia	19
--------------	----

---

## **Parte I**

### **Approccio Teorico**

---

## 1.1 - Introduzione

Le **agenzie di rating** esprimono pareri sul merito creditizio di un determinato emittente. In altri termini valutano la probabilità di fallimento dell'emittente, rispetto ai suoi obblighi finanziari (rating dell'emittente). Questi pareri, o rating, sono prevalentemente fondati su informazioni riguardanti i flussi di reddito e il bilancio (in particolare l'indebitamento) dell'entità oggetto del rating, nonché i risultati finanziari passati, piani strategici e prospettive di sviluppo. Essi si limitano a dare indicazioni sulla situazione in un dato momento e devono essere pertanto confermati o rivisti periodicamente per tenere conto dei recenti sviluppi, economici e di altro tipo. I rating classificano gli emittenti in categorie che corrispondono a gradi più o meno elevati di rischio di fallimento. Di conseguenza le agenzie di rating usano scale di merito di credito, nell'ambito delle quali il discrimine fondamentale è tra le categorie dette "**investment grade**" e "**speculative grade**", in funzione del rischio associato.

I rating sono basati sia su dati pubblicamente disponibili sia su informazioni che non sono accessibili al pubblico. Essi esercitano un'influenza notevole sui mercati finanziari. Le ragioni sono due. In primo luogo, pur essendo basati su valutazioni complesse, possono essere facilmente e immediatamente compresi dagli investitori, indipendentemente dal loro grado di competenza o profilo. In secondo luogo, le agenzie di rating godono di buona reputazione e sono viste dai partecipanti al mercato come organismi che forniscono analisi imparziali. Il rating è quindi considerato elemento fondamentale che condiziona la capacità di ottenere finanziamenti esterni sui mercati mobiliari. Il rating di un emittente determina i tassi d'interesse che dovrà offrire per ottenere un finanziamento esterno. Inoltre i rating sono utilizzati sempre più spesso in ambito contrattuali nelle clausole relative all'interruzione di una linea di credito, all'accelerazione del rimborso del debito o alla modifica di altre condizioni di finanziamento.

Tenuto conto del ruolo e dell'influenza che le agenzie di rating svolgono sui mercati, le autorità di vigilanza hanno riscontrato la necessità di regolare l'attività di rating, soprattutto per quanto attiene alla validazione dei rating. Le agenzie non solo devono fondare i loro rating su un'analisi rigorosa delle informazioni disponibili e su un controllo permanente dell'integrità delle loro fonti di informazione, ma anche avvalersi di procedure in grado di verificare l'attendibilità e la bontà di tali rating.

---

## 1.2 - Approccio metodologico

Il Comitato di Basilea ha formulato alcune linee guida relative alle caratteristiche che un sistema di rating deve possedere. Queste ultime riguardano la numerosità delle classi, la definizione di insolvenza, le modalità di determinazione del rating, gli organi coinvolti nel processo di *rating assignment*, le relative funzioni e il relativo grado di indipendenza. Importante è anche il tema dedicato alla valutazione della qualità di un sistema di rating. In generale, tale valutazione dovrebbe riguardare sia la **qualità degli input** che alimentano il sistema, sia **l'affidabilità del processo** mediante il quale gli input vengono elaborati, sia la **correttezza degli output** prodotti. La maggioranza dei criteri di valutazione, qualitativi e quantitativi, riguardano i dati che da un sistema di rating possono essere ricavati: tassi di insolvenza e tassi di migrazione.

Alcuni criteri sono i seguenti:

- i tassi di insolvenza dovrebbero risultare crescenti in modo monotono al peggiorare della classe di rating;
- i tassi di insolvenza di una data classe di rating dovrebbero essere stabili nel tempo, a indicazione del fatto che tutte le informazioni rilevanti sono incorporate nel rating;
- le posizioni che conservano la stessa classe di rating in un dato orizzonte temporale dovrebbero essere alte in percentuale; in altri termini, la diagonale della matrice di transizione dovrebbe presentare valori più alti di quelli delle altre celle;
- i tassi di migrazione verso classi vicine dovrebbero essere più elevati di quelli verso classi più lontane;
- i debitori che divengono insolventi dovrebbero essere classificati in classi basse (peggiori) dal maggior numero possibile di anni.

A fianco di questi criteri relativamente semplici, che possono essere verificati disponendo di una serie storica sufficientemente ampia di dati, in letteratura sono stati proposti alcuni criteri più sofisticati per valutare la qualità di un sistema di rating.

Un primo criterio per valutare la capacità discriminante, ossia il *model power* di un modello di scoring o di rating è quello della *contingency table* (Tabella 1). Si tratta di una matrice che sintetizza, all'interno di quattro diversi quadranti, le informazioni relative a:

---

- numero di imprese correttamente valutate solventi dal modello ( $N_1$ );
- numero di imprese erroneamente valutate solventi, corrispondente al numero di errori del primo tipo ( $N_2$ );
- numero di imprese erroneamente valutate insolventi, corrispondente al numero di errori del secondo tipo ( $N_3$ );
- numero di imprese correttamente valutate insolventi ( $N_4$ ).

		Stato Attuale	
		Solventi	Insolventi
Valutazione Modello	Solventi	$N_1$ Valutazioni Corrette	$N_2$ Errore di 1° tipo (Alpha)
	Insolventi	$N_3$ Errore di 2° tipo (Beta)	$N_4$ Valutazioni Corrette

**Tabella 1: Contingency Table**

Sulla base dei risultati riportati da una *contingency table* è possibile calcolare degli indicatori di performance del modello di scoring o rating sottostante. Tra questi i principali sono quattro:

- **Sensitivity:** percentuale di imprese insolventi correttamente identificate  $N_4/(N_2+N_4)$
- **Specificity:** percentuale di imprese sane correttamente identificate  $N_1/(N_1+N_3)$
- **Errore Alpha:** percentuale di imprese insolventi erroneamente classificate come sane  $N_2/(N_2+N_4)$
- **Errore Beta:** percentuale di imprese sane erroneamente classificate come insolventi  $N_3/(N_1+N_3)$

Il numero di errori del primo tipo e del secondo tipo dipende in modo cruciale dal valore del *cut-off* utilizzato per discriminare fra imprese sane e insolventi, ossia nel distinguere tra classi di rating *pass* e classi *fail*. Così, un valore soglia più elevato, ossia più selettivo, tende a produrre un maggior numero di errori del secondo tipo e un minor numero di errori del primo tipo. E' dunque evidente che la qualità di un modello di scoring o di un sistema di rating non può essere valutata considerando uno solo degli indicatori sopra riportati e quindi occorre analizzare e soprattutto minimizzare congiuntamente i valori assunti almeno dagli errori Alpha e Beta.

In realtà la *contingency table*, per quanto efficace come strumento di analisi del potere discriminante di un modello che segue una logica dicotomica (sano o anomalo), risulta meno efficace per valutare la performance di un sistema di rating, caratterizzato dalla presenza di più classi. Il motivo è legato al fatto che un sistema di rating non presenta un'unica soglia di discriminazione tra imprese solventi e imprese insolventi, ma numerose soglie che separano fra loro le diverse classi del sistema. L'analisi del *model power* dei sistemi di rating richiede dunque il ricorso a strumenti di analisi più completi. Fra questi, due assumono particolare rilevanza: la curva **Receiver Operating Characteristic** (ROC) e il **Cumulative Accuracy Profile** (CAP).

Il Comitato di Basilea ha previsto inoltre che il modello di rating sia sottoposto a una valutazione periodica basata su una logica di *backtesting*.

Le procedure di *backtesting* impiegate da Cerved Group utilizzano tecniche di ricampionamento casuale a partire dal campione di sviluppo per ovviare ad alcune difficoltà proprie delle attività di verifica (impossibilità tecnica di disporre di osservazioni su unità non interessate dal processo di rating, relativa scarsità di eventi di default) e successivamente calcolano gli indicatori ROC e CAP.

Cerved Group ha impostato un ricampionamento basato su procedure *bootstrap* stratificate per assicurarsi una maggior precisione nelle stime: infatti mentre la metodologia *bootstrap* permette la convergenza della distribuzione empirica di uno stimatore alla vera distribuzione, attraverso il campionamento stratificato si ottiene una riduzione della varianza di ciascun gruppo o strato con la conseguente riduzione della varianza di tutta la simulazione.

Infine sempre adottando la metodologia *bootstrap* si sono individuati degli indicatori in grado di valutare la bontà della stima ottenuta dal *backtesting*: la distorsione e l'errore standard.

La **distorsione** indica la differenza tra il valore atteso dello stimatore e il vero valore di un parametro da stimare, mentre l'**errore standard** indica la deviazione standard stimata dell'errore del metodo di stima o misurazione impiegato.

---

### 1.3 - Calcolo dell'Accuracy Ratio e dell'Area Under Curve

Il nuovo accordo di Basilea ha enfatizzato l'importanza di procedure statistiche di validazione come esplicitamente espresso nella sezione 252 paragrafo 599 nel 3° BIS Consultative Paper. Pertanto, Cerved Group ha deciso di sottoporre il proprio sistema di rating ad analisi periodiche volte a verificarne la qualità, cioè a capire se i giudizi espressi *ex ante* dal sistema di rating sono coerenti *ex post* con i risultati ottenuti. Per verificare la bontà del modello si ricorre a strumenti grafici di analisi, quali il Cumulative Accuracy Profile e il Receiver Operative Characteristic, e indicatori sintetici di tali rappresentazioni come l'Accuracy Ratio e l'Area Under Curve.

#### 1.3.1 - Il calcolo dell'Accuracy Ratio

La prima misura della performance di un sistema di rating o scoring è rappresentata dal rapporto di Gini. Questa curva misura la coerenza fra le previsioni di un sistema di rating in termini di *ranking* delle probabilità di insolvenza o del merito creditizio dei soggetti valutati e il *ranking* effettivo delle insolvenze. In pratica le imprese oggetto di valutazione vengono ordinate dalla peggiore alla migliore; quindi sull'asse delle ascisse è riportata in ordine crescente di qualità la classe di rating, mentre sull'asse delle ordinate viene posta la relativa frazione di imprese divenute insolventi, ossia la percentuale cumulata di insolvenze. Si ottiene così la curva indicata con CAP, corrispondente a un sistema di rating "reale". Il sistema di rating viene confrontato graficamente con altre due curve relative rispettivamente al modello perfetto (in blu) e al modello *random* (in rosso). La curva relativa al modello perfetto si ottiene ipotizzando un sistema di rating (teoricamente) perfetto cioè in grado di assegnare alla classe peggiore tutti i soggetti che diverranno insolventi. Il modello *random* o sistema di rating casuale, rappresentato da una retta diagonale, è invece il risultato che si ottiene da un sistema di rating privo di ogni valore informativo, ossia le insolvenze sono distribuite in modo uniforme tra tutte le classi. Il sistema "reale" si colloca nel nostro caso sempre tra le due curve corrispondenti a un sistema di rating perfetto e a un sistema casuale. Un sistema di rating reale è tanto migliore quanto più la curva CAP si avvicina a quella corrispondente al sistema perfetto.

Sulla base di questa logica si è costruito un indice sintetico di performance del sistema, denominato rapporto di Gini o Accuracy Ratio (AR), ottenuto come

---

rapporto tra l'area compresa tra la curva CAP, relativa al sistema in esame, e quella corrispondente a un sistema casuale e l'area compresa tra la curva relativa al sistema perfetto e quella corrispondente a un sistema casuale.

Il rapporto di Gini può assumere valori compresi tra 0 e 1. Quanto maggiore è il rapporto, tanto maggiore è il potere discriminante del sistema di rating in esame.



**Figura 1: Cumulative Accuracy Profile**

### **1.3.2 - Il calcolo dell'Area Under Curve e la rappresentazione del Receiver Operating Characteristic**

La curva ROC è una rappresentazione grafica mediante la quale è possibile identificare la combinazione di errori del primo e del secondo tipo corrispondenti a tutti i valori assunti da un livello di *cut-off*  $k$  che separa i soggetti solventi da quelli insolventi, ossia distingue tra classi di rating *pass* e classi *fail*. È quindi una rappresentazione grafica del *false alarm rate* sulle ascisse e dell'*hit rate* sulle ordinate per tutti i punti di *cut-off*. Il *false alarm rate* individua la frequenza con cui si sono scorrettamente classificati come insolventi dei soggetti solventi, mentre l'*hit rate* identifica la frequenza con cui si sono correttamente identificati i soggetti insolventi. Un sistema di rating/scoring risulta pertanto migliore quanto maggiore è l'indicatore *hit rate* e quanto minore è il *false rate*.

Quanto maggiore è l'inclinazione di tale curva tanto minore è il numero di falsi allarmi rispetto al numero di imprese correttamente identificate come insolventi, e tanto migliore è dunque la performance del modello.

---

Nei grafici viene inoltre confrontata la curva ROC con la curva corrispondente a un modello naif, privo di alcun valore informativo, corrispondente alla diagonale rossa. Un modello perfetto che non compie alcun errore, né del primo né del secondo tipo, sarebbe invece caratterizzato da una curva blu che assume valore 1 per tutti i valori delle ascisse.

La curva ROC quindi può essere anche vista come una sorta di *trade-off* tra errori del primo tipo ed errori del secondo tipo. Infatti l'*hit rate* corrisponde al complemento a uno della frequenza di errori del primo tipo, mentre il *false rate* corrisponde alla frequenza di errori del secondo tipo. L'informazione contenuta nella ROC può anche essere sintetizzata in un'unica misura denominata Area Under Curve (AUC). Questo indicatore identifica la dimensione dell'area che giace al di sotto della curva ROC. Il valore assunto dall'AUC può essere interpretato come la potenza media del test su insolventi/solventi corrispondenti a tutti i possibili valori di *cut-off*  $k$ . L'area AUC vale 0.5 per un modello *random* senza potere di discriminazione e 1 per il modello perfetto. Il modello è tanto migliore quanto più il valore dell'AUC si avvicina a 1.

Una volta determinata l'area AUC è possibile individuare l'AR in quanto esiste una relazione che lega tali indicatori:  $AR=2*AUC-1$ <sup>1</sup>



**Figura 2:Receiver Operating Characteristic**

---

<sup>1</sup> Per una dimostrazione si veda Measuring the Discriminative Power of Rating Systems, Engelmann B., Hayden E., Tasche D., Quanteam Research Paper 2003.

---

## 1.4 - Backtesting

Le tecniche di *backtesting* sono la componente chiave per validare le procedure e gli indicatori fino ad ora utilizzati. Attraverso l'utilizzo dei dati raccolti e il loro campionamento sono state riprodotte delle stime delle variabili di interesse, quali l'Accuracy Ratio e l'Area Under Curve. La tecnica impiegata è quella *bootstrap* in quanto permette la convergenza in probabilità della distribuzione empirica a quella vera.

### 1.4.1 - La metodologia Bootstrap

La tecnica *bootstrap* è un metodo di ricampionamento *computer intensive* per l'inferenza statistica. E' utilizzato per stimare la distribuzione del campione di uno stimatore o lo stimatore stesso attraverso il ricampionamento con reimbussolamento dalla popolazione iniziale, molto spesso con l'obiettivo di ottenere anche lo *standard error* e gli intervalli di confidenza del parametro di interesse, quale l'Accuracy Ratio o l'Area Under Curve. Inoltre può essere usato per costruire i test di ipotesi ed è spesso utilizzato come un'alternativa all'inferenza statistica, quando si basa su dubbi assunti parametrici o quando l'inferenza parametrica è impossibile o richiede formule molto complicate per il calcolo dello *standard error*. Per una corretta stima del parametro di interesse il numero di campioni *bootstrap*, consigliati in letteratura, è crescente in quanto crescenti sono le potenzialità computazionali disponibili. Solitamente 1000 campioni sono sufficienti, ma il compromesso migliore è individuare il giusto *trade-off* tra numero di campioni e disponibilità computazionale-temporale.

Quando non è nota a priori la funzione di distribuzione della popolazione  $F$  è preferibile utilizzare un *bootstrap* non parametrico, cioè non vincolato ad alcuna funzione di distribuzione. Supposto che i dati osservati  $y_1, \dots, y_n$  siano realizzazioni di un campione di variabili casuali  $Y_1, \dots, Y_n$  indipendenti e identicamente distribuite da una legge di distribuzione  $F$ , non nota, e si sia interessati a un parametro  $\theta = t(F)$ , la stima di  $\theta$  è  $t(\hat{F})$ , dove  $\hat{F}$  è la stima di  $F$  e rappresenta la EDF (funzione di distribuzione empirica) ottenuta basandosi sui dati  $y_1, \dots, y_n$ . In altre parole si può utilizzare la EDF  $\hat{F}$  per stimare la funzione di distribuzione cumulata (CDF)  $F$ .

Per migliorare la tecnica *bootstrap* ci si può avvalere del ricampionamento **stratificato**. Tale metodo consiste nel dividere gli  $N$  individui della popolazione in

---

sottopopolazioni, o strati, sulla base di una caratteristica comune (ad esempio lo score), successivamente nell'estrarre un campione casuale semplice da ogni strato in modo indipendente e infine nel riunire insieme i risultati dei singoli campionamenti per formare un unico campione dell'ampiezza richiesta. Operando in tale modo si assicura che gli individui della popolazione siano rappresentati adeguatamente nel ricampionamento e quindi si garantisce una maggior precisione nelle stime dei parametri in quanto si riduce la varianza della stima, a seguito della riduzione della varianza in ciascun gruppo.

Per valutare l'accuratezza statistica delle stime ottenute si è deciso di calcolare anche la distorsione e l'errore standard sempre attraverso la metodologia *bootstrap*.

#### 1.4.1.1 - La distorsione

In statistica la differenza tra il valore atteso dello stimatore e il vero valore di un parametro da stimare è chiamata distorsione. Se tale differenza è nulla allora lo stimatore si dice non distorto o corretto.

Si supponga di aver estratto il campione  $(y_1, \dots, y_n)$ , tramite un campionamento casuale, da una popolazione della quale non si conosce la distribuzione di probabilità  $F$  e di voler stimare il parametro  $\theta = t(F)$  utilizzando una qualsiasi statistica  $\hat{\theta} = t(\hat{F})$ . Se  $t(F)$  è il vero valore del parametro  $\theta$ , la distorsione di  $\hat{\theta}$  è data da  $E_F(t(\hat{F})) - t(F) = E_F(\hat{\theta}) - \theta$ . In altre parole la distorsione è un termine che si riferisce a quanto la media statistica dista dal parametro da stimare, cioè è l'errore che scaturisce quando si stima una quantità. Tale errore in generale non è possibile eliminarlo ripetendo l'esperimento un numero elevato di volte e facendo successivamente una media dei risultati ottenuti, ma se ne può tenere conto a posteriori correggendo la percezione per diminuirne gli effetti.

Si può usare la metodologia *bootstrap* per stimare la distorsione di uno stimatore  $\hat{\theta}$ . La stima *bootstrap* della distorsione si ottiene sostituendo  $F$  con  $\hat{F}$  in tal caso la distorsione è:  $E_{\hat{F}}(t(y^*)) - t(\hat{F})$ . Per la maggior parte delle statistiche essa deve essere approssimata con simulazione di tipo Monte Carlo. Dopo aver generato  $B$  campioni *bootstrap* indipendenti e dopo aver calcolato tutte le repliche *bootstrap* dello stimatore  $\theta^{*B} = t(y^{*B})$ , si approssima  $E_{\hat{F}}(t(y^*))$  con la media delle  $\theta^{*B}$ , cioè

---

$\theta^{*B} = (1/B) \sum_{i=1}^B \theta^*$  . Ne segue che la stima *bootstrap* della distorsione basata sui  $B$

campioni estratti sarà:  $\theta^{*B}() - \hat{\theta}$ . In conclusione si può schematizzare in questo modo il diagramma *bootstrap* applicato a problemi di carattere generale:

Mondo reale		Mondo bootstrap	
Modello di probabilità sconosciuto $F$	dati osservati $= (y_1, \dots, y_n)$	Modello di probabilità stimato $\hat{\phantom{F}}$	Campione bootstrap $^* = (y^*_1, y^*_2, \dots, y^*_n)$
Parametro di interesse $\theta = t(F)$	Stima di $\theta$ $\hat{\theta} = t(y)$	Parametro stimato $\hat{\theta} = t(\hat{\phantom{F}})$	replica bootstrap $\theta^* = t(y^*)$

#### 1.4.1.2 - L'errore standard

Lo  $(SE)$  di un metodo di misurazione o di stima è la deviazione standard stimata dell'errore di tale metodo. Quindi indica la deviazione standard della differenza tra i valori stimati o misurati e i veri valori. In particolare lo *standard error* di una statistica, come può essere la media, è la deviazione standard della distribuzione campionaria della statistica.

$$= \sqrt{\sum_{i=1}^n \frac{(y_i - \bar{y})^2}{(n-1)}}$$

Lo *standard error* è quindi una misura importante perché riflette la fluttuazione campionaria intorno a una statistica, quindi è un passo fondamentale per verificare l'accuratezza di una stima. L'inferenza statistica si basa sullo *standard error* nella costruzione degli intervalli di confidenza e nei test di significatività. Quando è nota la distribuzione di probabilità può essere usato per calcolare l'intervallo di confidenza esatto. Quando non è nota la distribuzione di probabilità tramite relazioni, tipo Chebichev, può essere calcolato comunque un intervallo di confidenza.

IC al 95% = stima  $\pm$  2 standard error

IC al 99% = stima  $\pm$  3 standard error

Lo *standard error* di una statistica dipende dalla lunghezza del campione, in generale più ampio è il campione più piccolo è lo standard error e quindi tanto più è

attendibile il valore statistico calcolato. Se la numerosità del campione è elevata e tende all'infinito vale il teorema del limite centrale che garantisce che la distribuzione campionaria è asintoticamente normale.

Il metodo *bootstrap* ci può essere d'aiuto per il calcolo di tale indicatore in quanto a volte può non avere una formula semplice. Tale metodologia con la formazione di numerosi campioni estratti dalla distribuzione empirica  $\hat{\mu}$  e con il calcolo, per ogni campione, delle corrispondenti statistiche, ci permette di stimare l'errore standard di una statistica campionaria con l'errore standard empirico delle repliche *bootstrap*. Passando cioè da  $\bar{y}$  alla posizione generale e indicando con  $s(y^{*B})$  la statistica calcolata sul generico campione *bootstrap*  $b$ -esimo indipendente, la stima dell'errore standard sarà data dalla deviazione standard delle  $B$  repliche *bootstrap*:

$$SE = \left( \frac{1}{(B-1)} \sum_{b=1}^B (y^{*b} - \bar{y}^*)^2 \right)^{1/2}, \quad \text{dove } s^* = \left( \frac{1}{B} \sum_{b=1}^B (y^{*b})^2 \right)^{1/2}$$

Per quanto riguarda la scelta del numero dei campioni *bootstrap* da generare (che indicheremo come sopra con  $B$ ) si ha che spesso  $B = 50$  è sufficiente per una buona stima dell'errore standard e, comunque, raramente abbiamo bisogno di  $B > 200$ .

---

## **Parte II**

### **Approccio Empirico**

---

## 2.1 - Descrizione dei dati

I dati impiegati nell'elaborazione sono contenuti nel database Cerved Group e si riferiscono ad aziende italiane classificate per natura giuridica e per macrosettore nell'anno 2009. A ciascun individuo della popolazione sono associate 2 variabili: la probabilità di default e una variabile dicotomica indicante l'evento di default dell'azienda.

Le società non di capitale, dato che non esiste l'obbligo di redigere il bilancio, non hanno lo stesso numero di variabili economiche che hanno le società di capitale e questo porterà a esprimere delle considerazioni sulla validità del modello anche sulla base della natura giuridica.

## 2.2 - Modelli di valutazione (Testing Rating Accuracy)

Per la valutazione del modello di rating si individuano gli indicatori grafici precedentemente descritti: l'Accuracy Ratio e il Receiver Operating Characteristic che definisce di conseguenza l'area Under Curve.

<b>Accuracy Ratio</b>	0.64
<b>Area Under Curve</b>	0.82

**Tabella 2**

I risultati relativi all'intera popolazione, utilizzati come valori di riferimento, vengono comparati con quelli ottenuti suddividendo la popolazione originaria in macrosettori. Sulla base di tale confronto i macrosettori relativi alla produzione e all'ingrosso mostrano valori degli indicatori superiori a quelli relativi all'intera popolazione, mentre il macrosettore relativo al dettaglio mostra un AR e un AUC poco inferiori ai valori di riferimento (Tabella 3).

---

	<b>Intera Popolazione</b>	<b>Macrosettore Produzione</b>	<b>Macrosettore Ingrosso</b>	<b>Macrosettore Dettaglio</b>
<b>Accuracy Ratio</b>	0.64	0.634	0.68	0.722
<b>Area Under Curve</b>	0.82	0.817	0.84	0.861

**Tabella 3**

E' possibile inoltre suddividere la popolazione originaria in base alla natura giuridica (S.p.a., S.r.l., Società di Persone, Ditte Individuali). Applicando il nostro modello di rating singolarmente a ciascun ambito, si riscontra che gli indicatori sintetici assumono valori più elevati rispetto a quelli dell'intera popolazione; solo per le Società di persone vi è un valore dell'Accuracy Ratio inferiore. Tali risultati sono possibili in quanto l'AR e l'AUC dipendono dalla distribuzione delle insolvenze e non semplicemente dalla percentuale di insolvenze presente nel sottocampione.

	<b>Intera Popolazione</b>	<b>S.p.a</b>	<b>S.r.l.</b>	<b>Ditte Individuali</b>	<b>Società di Persone</b>
<b>Accuracy Ratio</b>	0.64	0.688	0.678	0.66	0.676
<b>Area Under Curve</b>	0.82	0.844	0.839	0.830	0.838

**Tabella 4**

---

### 2.3 - Backtesting

Le tecniche di *backtesting* sono utilizzate secondo le metodologie correnti per validare le procedure e gli indicatori calcolati. Nel caso specifico non è nota a priori la funzione di distribuzione della popolazione quindi si è utilizzato un *bootstrap* non parametrico; si è inoltre optato per un ricampionamento stratificato per ridurre la varianza della stima.

La stima dell'Accuracy Ratio è pari a 0.64. Inoltre viene valutata l'accuratezza di tale stima (distorsione ed errore standard). La distorsione, indicante l'errore che scaturisce quando si stima una quantità, è in valore assoluto. L'altra misura in grado di valutare l'accuratezza della stima è lo *standard error* che indica la deviazione standard della differenza tra i valori stimati e il valore reale. Tale misura, che esprime la fluttuazione campionaria intorno alla statistica, è contenuta e inferiore al 2%; questo indica che la distribuzione degli Accuracy Ratio è concentrata intorno alla media, come si nota dalla distribuzione nella figura successiva.

Se il valore della distorsione lo rapportiamo al valore dello *standard error* si può notare che l'errore a cui ci si riferiva in precedenza non può essere molto elevato perché le oscillazioni attorno alla media degli *accuracy ratio* sono comunque molto basse.

Pertanto, i suddetti valori dimostrano la validità e la replicabilità del modello di Rating Cerved Group.

Vero Valore	Stima Media	Distorsione	Standard Error
0.64	0.6091	-0.0309	0.01831

**Tabella 5**

---

**Bibliografia**

- Rischio e Valore nelle Banche, Risk Management e Capital Allocation – A.Sironi – Egea, Milano, 2005.
  - An Introduction to the Bootstrap with Applications in R – A.C. Davison, D. Kuonen – Statistical Computing & Statistical Graphics Newsletters Vol.13 n°1.
  - Validation Methodologies for Default Risk Models – J.Sobehart, S.Keenan, R.Stein – Default Risk, 2000.
  - Mr Risk: Getting to Know Him Better – A. Meucci – Egea, Milano, 2002.
  - Comunicazione della Commissione sulle Agenzie di Rating del Credito – Gazzetta Ufficiale dell’Unione Europea, 2006.
  - Measuring the Discriminative Power of Ratings Systems – B. Engelmann, E. Hayden, D.Tasche – Quanteam Research Paper, 2003.
-