



UNIVERSITA' DEGLI STUDI DI PADOVA

**DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI
"M.FANNO"**

CORSO DI LAUREA IN ECONOMIA E MANAGEMENT

PROVA FINALE

**LA PROBABILITÀ DI ESSERE "CRIMINALI":
ANALISI CON MODELLI A VARIABILE DIPENDENTE LIMITATA
DEL CASO PIEMONTESE**

**RELATORE:
CH.MO PROF. CAPPuccio NUNZIO**

**LAUREANDA: BALLIANA VALENTINA
MATRICOLA N. 1022092**

ANNO ACCADEMICO 2015 – 2016

INDICE

1. INTRODUZIONE.....	3
2. PRESENTAZIONE DEI DATI.....	6
2.1 DESCRIZIONE DEL DATABASE.....	6
2.2 ALCUNE SEMPLIFICAZIONI.....	9
3. STRUMENTI PER L'ANALISI.....	12
3.1 MODELLI ECONOMETRICI PER L'ANALISI DI VARIABILI DIPENDENTI LIMITATE.....	12
3.2 SCELTA DELLE VARIABILI INDIPENDENTI.....	17
3.3 PULIZIA DEI DATI.....	20
3.4 CREAZIONE DEL FOGLIO DATI TRAMITE EXCEL. IL SOFTWARE GRETL.....	21
4. COMMENTO DEI RISULTATI OTTENUTI DALL'ANALISI.....	23
4.1 ALCUNE CONSIDERAZIONI AI FINI DELLA COMPrensIONE DEI RISULTATI.....	23
4.2 SIGNIFICATIVITÀ DELLE VARIABILI INDIPENDENTI.....	25
4.3 DISCUSSIONE DEL SEGNO DELLE VARIABILI INDIPENDENTI SIGNIFICATIVE. UN RISCONTRO NELLA REALTÀ.....	29
5. CONCLUSIONI.....	32
6. TABELLE ALLEGATE.....	33
7. BIBLIOGRAFIA E SITOGRAFIA.....	43

1. INTRODUZIONE

Storicamente, la mafia italiana nasce nella Sicilia del XIX secolo, quando i grandi proprietari terrieri si avvalevano di gruppi di individui violenti per costringere i contadini a pagare il prezzo per poter coltivare i terreni di proprietà dei latifondisti stessi. In seguito, tali gruppi sono diventati autonomi, continuando il loro operato di violenza e soprusi, il tutto occultato da un regime di omertà. Sono queste, quindi, le fondamenta di uno dei problemi più urgenti e gravi dello Stato italiano, che nel corso del Novecento è diventato un fenomeno non solo di portata nazionale, ma anche internazionale (soprattutto negli Stati Uniti). Inoltre, la mafia ha sempre avuto un ruolo attivo nella storia italiana degli ultimi due secoli, infiltrandosi nel tessuto socio-politico ed economico della penisola, molto spesso influenzandone le vicende¹.

Che la mafia non sia più solamente circoscritta alle regioni del Mezzogiorno è ormai un dato di fatto, del tutto innegabile: l'immagine in seguito riportata, tratta dall'inchiesta di Repubblica *“Così la Mafia conquista il Nord d'Italia”* di Autieri Daniele, Baldessarro Giuseppe, Di Salvo Michele e Palazzolo Salvo, risalente a febbraio 2015, dimostra chiaramente come la criminalità organizzata si sia diffusa nelle regioni settentrionali, nessuna esclusa.

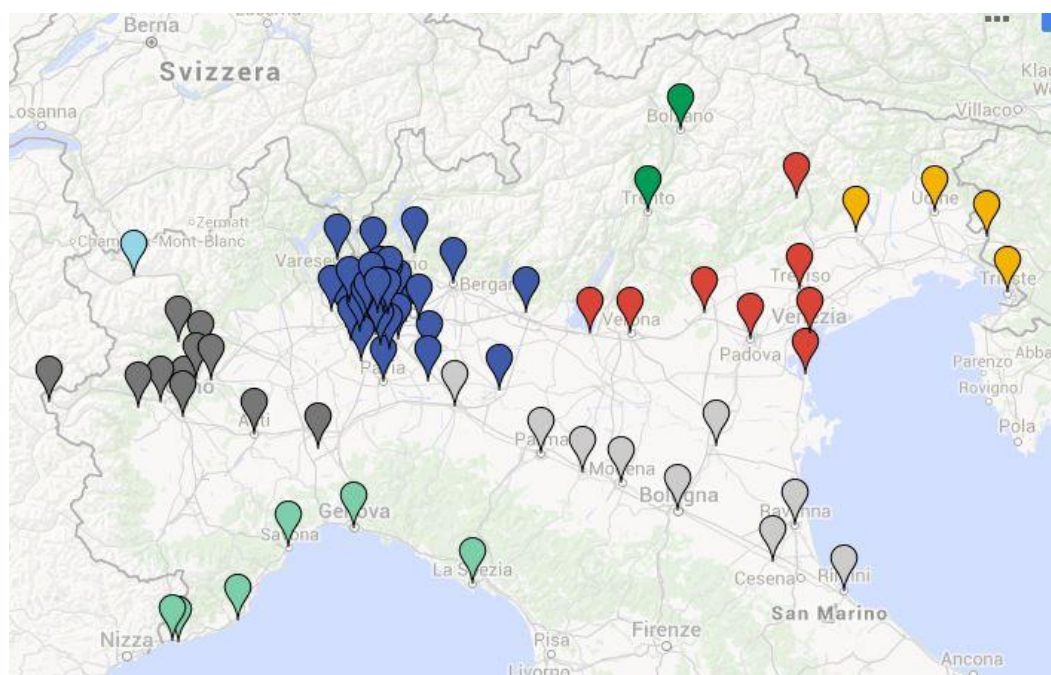


Figura 1: Criminalità organizzata nel Settentrione.

Leggendo il *“Primo Rapporto Trimestrale sulle aree settentrionali, per la Presidenza della Commissione Parlamentare di inchiesta sul fenomeno mafioso”*, a cura dell'Osservatorio sulla Criminalità Organizzata dell'Università degli Studi di Milano, del 29 settembre 2014, emerge, per prima cosa, come la criminalità organizzata nelle regioni settentrionali si diffonda prevalentemente nei piccoli comuni, dove è più facile gestire e manipolare le amministrazioni locali, infiltrare le imprese e creare rapporti di lealtà.

¹ www.treccani.it

Non solo piccole dimensioni: i comuni maggiormente colpiti presentano anche un'elevata densità demografica. Questo fatto può essere spiegato considerando che l'elevata densità può essere collegata a spostamenti di individui provenienti dalle zone originarie del clan, nonché dal fatto che questa favorisca la "mimetizzazione sociale" e, quindi, la possibilità di creare legami sociali ed economici senza essere troppo notati. Non mancano, tuttavia, casi in cui la criminalità organizzata trovi terreno fertile in zone scarsamente popolate: ciò è spiegabile considerando che tali aree, essendo isolate, sono potenzialmente più idonee ad essere controllate e, inoltre, rimangono più facilmente estranee alle correnti di pensiero che si sviluppano prevalentemente nelle grandi aree metropolitane.

Nel presente elaborato ci si sofferma, in particolare, sul **Piemonte**. In questa regione, il forte espansionismo urbano degli anni '60 ha certamente aiutato l'insediamento di organizzazioni mafiose, collegate soprattutto alla 'ndrangheta, come è emerso da due maxi-inchieste degli ultimi anni, l'operazione Minotauro (2011) e l'operazione Albachiara (2011).

Il secondo aspetto che si vuole analizzare non è tanto l'operato delle organizzazioni criminali di stampo mafioso nell'economia illegale (ad es. produzione di droga, sfruttamento della prostituzione, usura e corruzione ecc.) quanto più il modo in cui esse riescono ad infiltrare ed inquinare l'**economia legale**. I motivi che possono spingere la criminalità organizzata a investire nell'economia legale sono molteplici e di varia natura:

- Innanzitutto, sicuramente uno degli scopi principali è quello di realizzare ulteriori profitti, da impiegare per commettere illeciti e per finanziare l'impresa mafiosa;
- Investendo nell'economia legale è, inoltre, possibile occultare i proventi illeciti, in particolare attraverso operazioni di riciclaggio;
- La mafia, in tal modo, riesce anche a creare un consenso sociale, sia tra le fasce più povere della popolazione, offrendo nuovi posti di lavoro e garantendo così un reddito ai più bisognosi, sia tra gli altri imprenditori, arrivando perfino a stringere legami con le istituzioni politiche. In questo contesto il controllo del territorio sarà sicuramente più semplice;
- Anche inclinazioni personali e motivazioni di carattere culturale possono influenzare la decisione di infiltrare l'economia legale: i soggetti mafiosi possono, infatti, essere spinti a investire in settori in cui è tradizionalmente presente la mafia, o in settori ritenuti "più sicuri" dalla cultura italiana.

Sono queste, quindi, le principali motivazioni che portano le organizzazioni criminali a non concentrarsi esclusivamente su attività illegali; la scelta del settore in cui operare dipende essenzialmente dalle caratteristiche e dalle opportunità offerte dallo stesso in virtù di quale delle finalità appena descritte si preferisca realizzare. Quanto appena illustrato viene ben evidenziato dal Progetto PON Sicurezza 2007-2013 "*I beni sequestrati e confiscati alle organizzazioni criminali nelle regioni dell'Obiettivo Convergenza: dalle strategie di investimento della criminalità all'impiego di fondi comunitari nel riutilizzo dei beni già destinati*", realizzato da Transcrime, il Centro Universitario di Ricerca sulla Criminalità Transnazionale dell'Università Cattolica del Sacro Cuore di Milano.

Alla luce di quanto appena descritto, l'obiettivo del presente elaborato è quello di individuare quali siano le caratteristiche delle aziende che, variando, modificano la probabilità che le stesse vengano infiltrate da organizzazioni criminali. Per fare ciò, sono stati reperiti dei dati riportanti alcune informazioni relative a società di capitali piemontesi nonché i relativi bilanci aziendali. In particolare, è stata costruita la variabile CRIMINAL (che avrà il ruolo di variabile dipendente) che assume valore 1 se la società è mafiosa, zero altrimenti. Lo scopo prefissato viene realizzato scegliendo opportune variabili indipendenti (gli indici di bilancio) e grazie all'utilizzo di tre diversi modelli econometrici (il *modello di probabilità lineare*, il modello *logit* e quello *probit*).

Nel primo capitolo verrà presentato il data set fornito, descrivendo le informazioni che sono riportate al suo interno. Verranno anche espone le assunzioni effettuate, al fine di rendere più semplice l'analisi. Successivamente, nel secondo capitolo, saranno illustrati sia i modelli econometrici utilizzati per poter soddisfare l'obiettivo della tesi, sia le motivazioni che hanno spinto a scegliere gli indici di bilancio quali variabili indipendenti di questi modelli. Infine, nell'ultima parte, saranno riportati e discussi i risultati dell'analisi.

2. PRESENTAZIONE DEI DATI

2.1 DESCRIZIONE DEL DATABASE

Il data set iniziale è costituito dalle voci di stato patrimoniale e conto economico, riclassificate secondo lo schema civilistico (artt. 2423 ss.), riguardanti un totale di 64.525 aziende situate nella regione Piemonte. Tale riclassificazione, usata esclusivamente per le società di capitali, è di particolare importanza in quanto sarà la base per la determinazione degli indici di bilancio, che costituiscono le variabili indipendenti dei modelli econometrici usati per l'analisi.

In particolare, l'art. 2424 c.c. stabilisce il seguente schema di stato patrimoniale:

ATTIVO	PASSIVO
A. CREDITI VERSO SOCI PER VERSAMENTI ANCORA DOVUTI	A. PATRIMONIO NETTO I – IX : capitale sociale, riserve di varia natura, utili (perdite) portati a nuovo e d'esercizio
B. IMMOBILIZZAZIONI I. Immobilizzazioni immateriali II. Immobilizzazioni materiali III. Immobilizzazioni finanziarie*	B. FONDI PER RISCHI E ONERI
C. ATTIVO CIRCOLANTE I. Rimanenze II. Crediti* III. Attività finanziare (che non costituiscono immobilizzazioni)* IV. Disponibilità liquide	C. TRATTAMENTO DI FINE RAPPORTO DI LAVORO SUBORDINATO
D. RATEI E RISCOINTI ATTIVI	D. DEBITI
	E. RATEI E RISCOINTI PASSIVI
TOTALE ATTIVO (A+B+C+D)	TOTALE PASSIVO (A+B+C+D+E)

* per quanto riguarda i crediti e i debiti bisogna fare attenzione a distinguere tra quelli esigibili entro l'anno e tra quelli esigibili entro o oltre l'esercizio successivo.

Lo schema di conto economico (art. 2425 c.c.), invece, non presenta una struttura a sezioni contrapposte come nello stato patrimoniale, bensì una forma scalare, utile per mettere in risalto alcuni risultati intermedi nella formazione del risultato d'esercizio che altrimenti andrebbero persi ricorrendo alle sezioni divise:

A. VALORE DELLA PRODUZIONE
B. COSTI DELLA PRODUZIONE
DIFFERENZA TRA VALORE E COSTI DELLA PRODUZIONE (A-B)
C. PROVENTI E ONERI FINANZIARI
D. RETTIFICHE DI VALORE DI ATTIVITÀ FINANZIARIE
E. PROVENTI E ONERI STRAORDINARI
RISULTATO PRIMA DELLE IMPOSTE (A-B±C±D±E)
IMPOSTE SUL REDDITO
UTILE (PERDITE) D'ESERCIZIO

Tutte le informazioni sono riportate in Stata, un software statistico utilizzato per l'elaborazione e l'analisi di dati, per la creazione di grafici e la stima di numerosi modelli econometrici.

Per ciascuna azienda viene poi riportata la ragione sociale (sono prese in considerazione esclusivamente società di capitali), il codice fiscale, la provincia e il comune di appartenenza. Inoltre, ad ognuna viene attribuito un codice, fondamentale per poter aprire una nuova partita Iva, che permette di stabilire in modo specifico il tipo di attività svolta dagli agenti economici e, di conseguenza, quali debbano essere gli adempimenti fiscali a cui sono assoggettati. Le aziende, quindi, vengono ripartite in base alla classificazione delle attività economiche Ateco 2007, entrata in vigore a partire dal 1° gennaio 2008. Tale codice si presenta come una combinazione di lettere e numeri, i quali individuano una configurazione gerarchizzata a cinque livelli, in cui si parte da quello più generico fino ad arrivare a quello più dettagliato:

1. sezioni
2. divisioni
3. gruppi
4. classi
5. categorie
6. sottocategorie.

Un esempio di classificazione Ateco è il seguente²:

46.11.01 – AGENTI E RAPPRESENTANTI DI MATERIE PRIME AGRICOLE:

	CODICE	DEFINIZIONE
Sezione	G	Commercio all'ingrosso e al dettaglio; riparazione di autoveicoli e motocicli
Divisione	46	Commercio all'ingrosso (escluso quello di autoveicoli e motocicli)
Gruppo	46.1	Intermediari del commercio
Classe	46.11	Intermediari del commercio di materie prime agricole, di animali vivi, di materie prime tessili e di semilavorati
Categoria	46.11.0	Intermediari del commercio di materie prime agricole, di animali vivi, di materie prime tessili e di semilavorati; pelli grezze
Sottocategoria	46.11.01	Agenti e rappresentanti di materie prime agricole

Nel set di dati a disposizione viene riportato solo il secondo livello (divisioni), per cui le aziende vengono identificate da un numero compreso tra 1 e 99 (ad esempio “68” indica il settore “attività immobiliari”).

L'ultima informazione fornita dai dati riguarda la variabile CRIMINAL: ad ogni azienda viene attribuito il valore 0 se non viene considerata criminale (l'evento non si verifica) o il valore 1 qualora invece lo sia (l'evento si verifica). È importante specificare che una determinata azienda è ritenuta criminale se tra i suoi amministratori e azionisti ve ne sono alcuni che sono stati condannati per il reato di associazione a delinquere di stampo mafioso: per determinare ciò sono state recuperate e analizzate tutte le sentenze riguardanti tale crimine compiuto nella regione Piemonte, emesse dal 2005 al 2014. L'articolo 416-bis del codice penale afferma che “[...]L'associazione è di tipo mafioso quando coloro che ne fanno parte si avvalgono della forza di intimidazione del vincolo associativo e della condizione di assoggettamento e di omertà che ne deriva per commettere delitti, **per acquisire in modo diretto o indiretto la gestione o comunque il controllo di attività economiche, di concessioni, di autorizzazioni, appalti e servizi pubblici o per realizzare profitti o vantaggi ingiusti per sé o per altri.**[...]”. Una volta individuati i nomi dei condannati, si è potuto risalire alle relative aziende e accedere ai bilanci d'esercizio, tramite il ricorso alla Camera di

² Esempio tratto da www.istat.it.

Commercio e all'AIDA (Analisi Informatizzata delle Aziende Italiane), una banca dati contenenti informazioni di varia natura sulle società di capitali italiane.

Le aziende sono osservate nel corso del tempo; infatti, sono riportati conto economico e stato patrimoniale relativi a più annualità. Questo perché le aziende in questione sono considerate criminali dal momento in cui il soggetto colpevole è diventato socio o amministratore della stessa fino all'anno in cui l'azienda è stata indagata o all'anno in cui tale soggetto ha lasciato la propria posizione. Per quanto riguarda invece quelle non criminali, sono state prese in considerazione le informazioni relative a tutti gli anni resi disponibili dal database AIDA.

2.2 ALCUNE SEMPLIFICAZIONI

Il dataset appena descritto si trova però allo stato grezzo: sono presenti, infatti, alcune imprecisioni ed incongruenze, imputabili prevalentemente ad errori di inserimento delle informazioni o sviste, che devono essere corretti affinché non sia compromessa la successiva analisi. Altri errori sono dovuti al fatto che, per mancanza di informazioni di cui si parlerà dopo, gli indici vengono calcolati facendo riferimento agli schemi di stato patrimoniale e al conto economico forniti dal codice civile.

Prima di procedere con l'operazione di pulizia, e ancora prima di definire le variabili indipendenti, è necessario introdurre due semplificazioni che hanno lo scopo di facilitare suddetta analisi:

1. A causa dell'elevato numero di aziende rilevate, si è ritenuto necessario ridimensionare la portata dell'analisi per renderla più agevole e snella. Per tale motivo, quindi, si è deciso di concentrare l'attenzione solo su tre settori economici, in particolare su quelli in cui la variabile `CRIMINAL` assume più volte il valore 1. Per poterli individuare, è stato costruito un istogramma tramite Stata: sull'asse delle ascisse viene riportato il codice Ateco relativo a ciascun ambito mentre su quello delle ordinate è indicata la variabile `CRIMINAL`. Facendo riferimento a quest'ultimo asse è possibile stabilire quante aziende criminali ci sono per ogni settore.

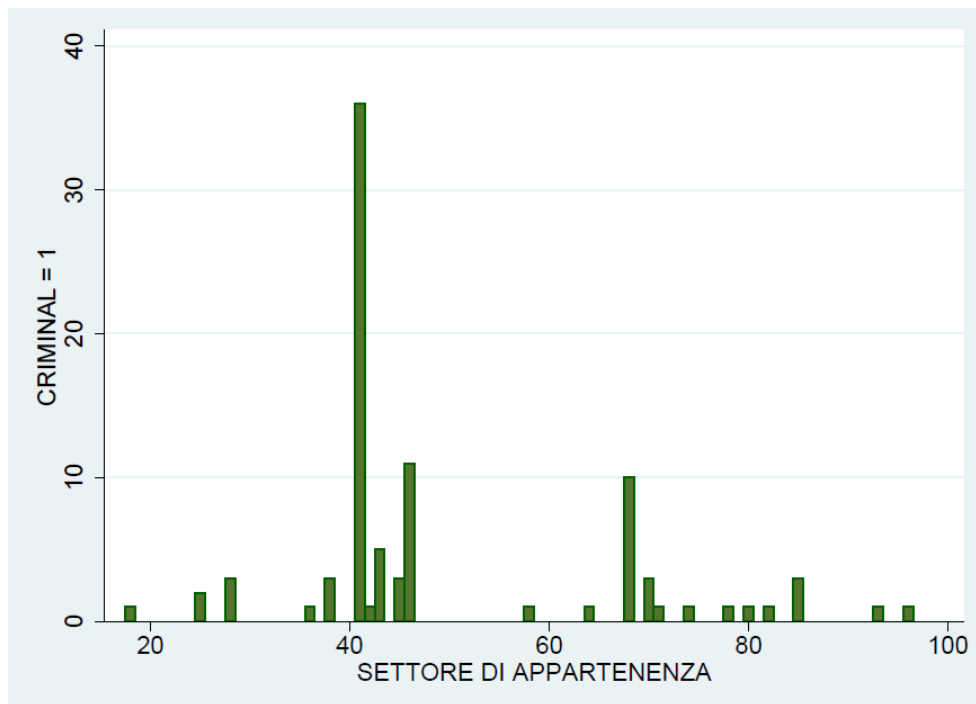


Figura 2: Numero di aziende criminali per ogni settore economico in Piemonte.

Osservando il grafico, si giunge quindi alla conclusione che i tre settori maggiormente colpiti dal fenomeno in esame sono:

- 1) costruzione di edifici (codice 41), 36 aziende connesse alla criminalità organizzata su un totale di 6.451;
- 2) commercio all'ingrosso, escluso quello di autoveicoli e motocicli (codice 46), 11 aziende criminali su un totale di 6.276;
- 3) attività immobiliari (codice 68), 10 aziende criminali su un totale di 8.483³.

Sono questi, quindi, i segmenti di mercato che verranno analizzati e in base ai quali si cercherà di determinare le caratteristiche delle aziende criminali al cui variare aumenta o diminuisce la probabilità che l'azienda sia criminale. Dal grafico si può anche notare che, in generale, il numero delle aziende criminali in Piemonte è molto basso. Alla luce di ciò, si è deciso sia di analizzare separatamente i tre settori individuati e di discutere i relativi risultati, sia di considerare, come verrà in seguito descritto, un unico flusso dati cumulativo dei tre: infatti vi è il rischio che, se presi singolarmente, la presenza di un numero limitato di CRIMINAL con valore pari a 1 porti ad avere dei risultati fuorvianti o inconcludenti.

³ L'individuazione dei settori avviene utilizzando i dati allo stato grezzo; ciò comporta che, dopo la pulizia, il numero di aziende, criminali e non, varierà inevitabilmente. La scelta è stata fatta in questo momento proprio con lo scopo di semplificare l'analisi.

2. Per quanto riguarda la dimensione temporale, essa viene trattata nel seguente modo: i dati non vengono considerati come serie storiche o panel, ma si è deciso di lavorare con valori medi: per ogni voce di conto economico e di stato patrimoniale di ciascuna azienda è stata calcolata la media aritmetica relativa al periodo di osservazione: in questo modo, in corrispondenza di ogni codice fiscale identificante l'azienda, si avrà un solo valore per i ricavi di vendita, uno per le disponibilità liquide, uno per il patrimonio netto ecc.

3. STRUMENTI PER L'ANALISI

In questo terzo capitolo verranno presentati, innanzitutto, i modelli utilizzati per individuare i caratteri che fanno aumentare la probabilità che le aziende prese in esame in questo elaborato possano diventare criminali. Essi sono, come già accennato, il modello di probabilità lineare, il modello logit e quello probit. Una volta descritti, si procederà invece ad illustrare quali siano gli indici di bilancio scelti come variabili indipendenti di tali modelli e le motivazioni che hanno spinto a considerare proprio tali quozienti.

3.1 MODELLI ECONOMETRICI PER L'ANALISI DI VARIABILI DIPENDENTI LIMITATE

In molte situazioni, per poter interpretare il comportamento degli agenti economici, è necessario introdurre nel modello di regressione variabili di carattere qualitativo, ovvero variabili le cui modalità non esprimono valori numerici bensì qualità (si pensi, ad esempio, al genere, all'età o al livello d'istruzione). Esse possono avere una funzione esplicativa oppure, come nel presente elaborato, è la stessa variabile dipendente ad essere di tipo qualitativo e, quindi, la regressione multipla viene usata per spiegare eventi non numerici. In particolare, ci si sofferma sul caso in cui la variabile dipendente possa assumere solo due valori:

CRIMINAL, infatti, è pari a uno se l'azienda presa in considerazione è mafiosa, è pari a zero altrimenti. Una variabile dipendente binaria non è altro che un esempio di **variabile dipendente limitata** (limitata in quanto il range di valori che può assumere è ristretto) e i modelli più frequentemente utilizzati per analizzare questo tipo di variabili di risposta sono appunto tre: il modello di probabilità lineare, il modello logit e il modello probit.

IL MODELLO DI PROBABILITÀ LINEARE

Il modello di probabilità lineare (in inglese: *Ordinary Least Squares – OLS*) è quello, tra i tre, di più facile interpretazione: consiste semplicemente nell'applicazione del modello di regressione multipla ad una variabile dipendente binaria. È utile quando la variabile Y_i è definita come:

$$Y_i = \begin{cases} 1, & \text{con probabilità } p, \text{ se l'evento di verifica} \\ 0, & \text{con probabilità } (1 - p), \text{ se l'evento non si verifica} \end{cases}$$

con $0 < p < 1$.

Dato il modello di regressione lineare multipla $Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_j X_{ji} + \varepsilon_i$,
 $i = 1, \dots, n$, assumendo che $E(\varepsilon_i | X_{1i}, \dots, X_{ji}) = 0$ si ottiene:

$$E(Y_i | X_{1i}, \dots, X_{ji}) = \beta_0 + \beta_1 X_{1i} + \dots + \beta_j X_{ji}, \quad i = 1, \dots, n,$$

e dal momento che Y_i si distribuisce come una variabile casuale di Bernoulli (potendo assumere solo due valori), si ha sempre che la probabilità di “successo” (ovvero la probabilità che $Y_i = 1$) è pari al valore atteso di Y_i :

$$\begin{aligned} E(Y_i | X_{1i}, \dots, X_{ji}) &= 1 \cdot P(Y_i = 1 | X_{1i}, \dots, X_{ji}) + 0 \cdot P(Y_i = 0 | X_{1i}, \dots, X_{ji}) \\ &= P(Y_i = 1 | X_{1i}, \dots, X_{ji}). \end{aligned}$$

Questa equazione dunque afferma che la probabilità condizionata di ottenere $Y_i = 1$ è una funzione lineare delle variabili esplicative X_{1i}, \dots, X_{ji} . Per questo motivo tale modello viene chiamato **modello di probabilità lineare** mentre $P(Y_i = 1 | X_{1i}, \dots, X_{ji})$ viene definita **probabilità di risposta**. Il coefficiente β_j misura il cambiamento nella probabilità di successo in seguito ad una variazione unitaria della variabile esplicativa X_j , tenendo fissi tutti gli altri fattori:

$$\Delta P(Y_i = 1 | X_{1i}, \dots, X_{ji}) = \beta_j \Delta x_j$$

Come nel modello di regressione lineare multipla, anche in questo caso i parametri possono essere stimati grazie al metodo dei minimi quadrati; $\widehat{\beta}_0, \widehat{\beta}_1, \dots, \widehat{\beta}_j$ si ottengono minimizzando la somma dei quadrati degli errori SQE:

$$SQE = \sum_{i=1}^n e_i^2$$

dove e_i rappresenta i residui, definiti da: $e_i = Y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 X_{1i} + \dots + \widehat{\beta}_j X_{ji})$, $i = 1, \dots, n$. $\widehat{\beta}_j$ rappresenta ora la variazione prevista nella probabilità di successo quando X_j aumenta di una unità.

Il modello di probabilità lineare presenta però alcuni problemi:

- il modello di probabilità lineare viola una delle assunzioni del teorema di Gauss-Markov: ora gli stimatori degli OLS sono corretti e consistenti, ma non sono più

efficienti. Ciò è dovuto al fatto che non viene rispettata l'assunzione di omoschedasticità, in quanto la varianza del termine d'errore non è più costante, ma dipende dal valore delle variabili indipendenti. In questo caso si parla di **eteroschedasticità**:

$$\text{Var}(\varepsilon_i|X = x) = \text{Var}(Y|X = x) = p(x)[1 - p(x)] = (\beta_1 + \beta_2 x) \times [1 - (\beta_1 + \beta_2 x)].$$

Questo problema viene trattato ricorrendo agli errori standard robusti all'eteroschedasticità.

- si possono ottenere valori stimati inferiori a zero o maggiori di uno; nella figura 3 la retta tratteggiata individua la funzione di regressione stimata, mentre le due linee continue indicano il luogo in cui giacciono le osservazioni. È possibile notare come la retta tratteggiata vada al di fuori dell'area delimitata da quelle orizzontali; questo dimostra chiaramente che si possono avere valori previsti al di fuori dell'intervallo $[0,1]$.

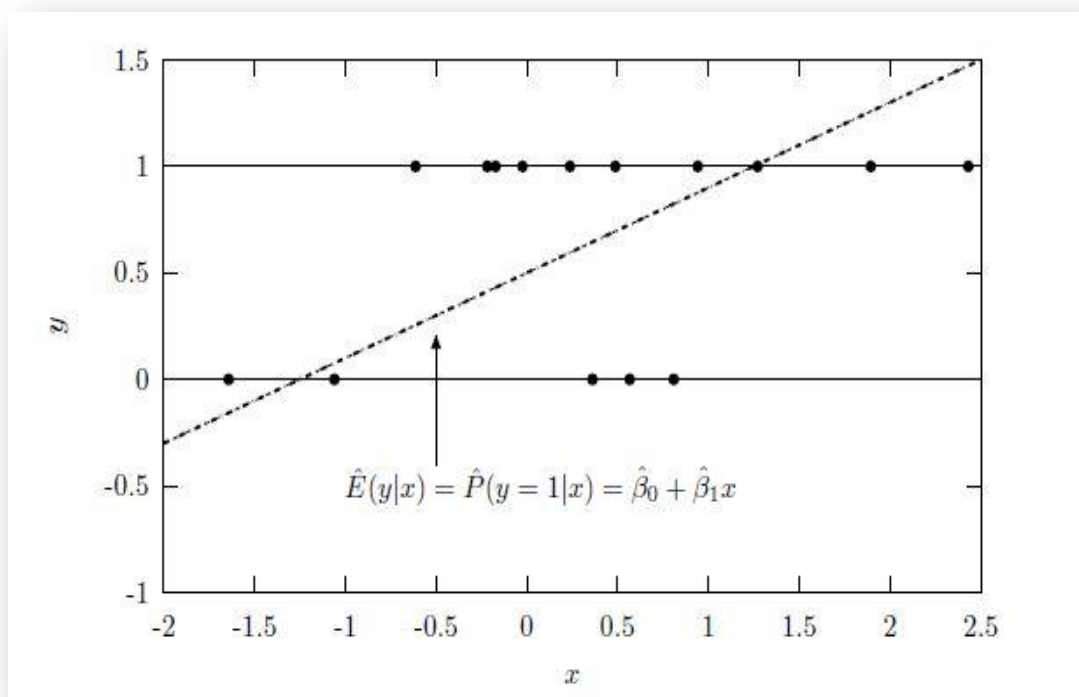


Figura 3: Il modello di probabilità nel caso di una regressione semplice⁴.

- gli errori, inoltre, non hanno distribuzione normale, dal momento che Y può assumere solo due valori.

Immagine tratta da Cappuccio N., Orsi R., 2005. *Econometria*. 2°ed. Bologna: Il Mulino, p.478.

I MODELLI LOGIT E PROBIT

Questi modelli sono più difficili da applicare e interpretare rispetto al precedente, ma permettono di superare alcuni dei limiti introdotti dal modello di probabilità lineare. Innanzitutto, per evitare di ottenere risultati che siano inferiori a zero o maggiori di uno, è necessario introdurre una forma funzionale per $F(\mathbf{x}'_i\beta)$ in modo che risulti che:

1. $F(\mathbf{x}'_i\beta) : \mathbb{R} \rightarrow [0,1]$
2. $\begin{cases} \lim_{\mathbf{x}'_i\beta \rightarrow +\infty} F(\mathbf{x}'_i\beta) = 1 \\ \lim_{\mathbf{x}'_i\beta \rightarrow -\infty} F(\mathbf{x}'_i\beta) = 0 \end{cases}$

In tal modo si è certi che $P_i = F(\mathbf{x}'_i\beta) \in [0,1]$.

In particolare, $F(\cdot)$ nei modelli logit e probit è rappresentata da due funzioni cumulate di probabilità, o funzioni di ripartizione: la **logistica** per il primo, la **normale** per il secondo.

Nel **modello logit**, si assume che il termine d'errore ε_i si distribuisca come una variabile casuale logistica, che ha densità pari a:

$$\lambda(\varepsilon_i) = \frac{\exp\{-\varepsilon_i\}}{(1 + \exp\{-\varepsilon_i\})^2}.$$

Una variabile con densità di tale tipo ha media pari a zero e varianza uguale a $\pi^2/3$.

La probabilità che l'evento si verifichi è data da:

$$P(Y_i = 1 | X_{1i}, \dots, X_{ji}) = \frac{e^{x'_i\beta}}{1 + e^{x'_i\beta}} = \frac{1}{1 + e^{-x'_i\beta}} = \Lambda(\mathbf{x}'_i\beta),$$

dove $\Lambda(\cdot)$ rappresenta la funzione di ripartizione della logistica.

Nel **modello probit**, ε_i si distribuisce come una normale con media nulla con varianza pari a σ^2 : $\varepsilon_i \sim N(0, \sigma^2)$.

La probabilità che l'evento si realizzi è data da:

$$P(Y_i = 1 | X_{1i}, \dots, X_{ji}) = \int_{-\infty}^{\frac{x'_i\beta}{\sigma}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\left(\frac{z_i^2}{2}\right)\right\} dz_i = \Phi\left(\frac{x'_i\beta}{\sigma}\right),$$

dove $\phi(\cdot)$ rappresenta la funzione di ripartizione della normale standardizzata.

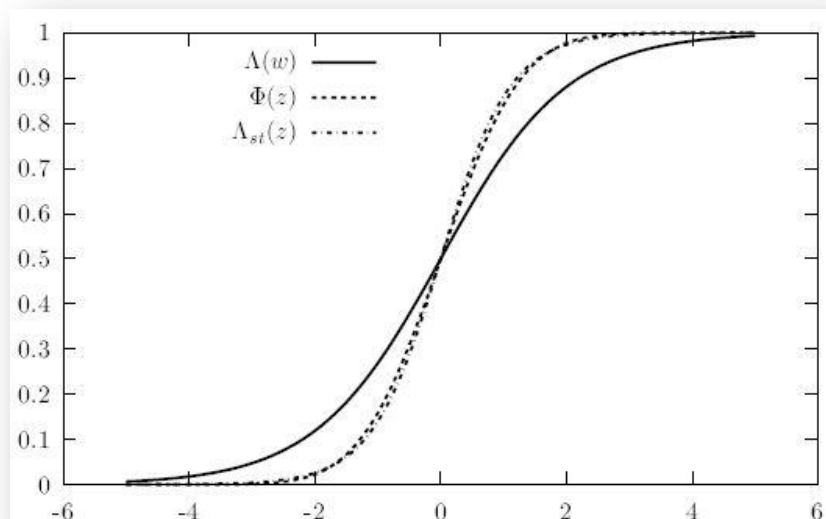


Figura 4: La funzione di ripartizione della logistica, $\Lambda(w)$, della logistica standardizzata, $\Lambda_{st}(w)$, e della normale standardizzata, $\Phi(z)$ ⁵.

In questo caso viene utilizzato il **modello della massima verosimiglianza** per la stima dei parametri.

Dal momento che la stima della massima verosimiglianza si basa sulla distribuzione di y data x , viene superato agevolmente il problema dell'eteroschedasticità, analizzato precedentemente per quanto riguarda il modello di probabilità lineare.

Dato un campione di n variabili indipendenti, per poter determinare lo stimatore di massima verosimiglianza è necessario utilizzare la densità di y_i data x_i . Si può, quindi, scrivere:

$$l_i(\beta) = [F(x_i'\beta)]^{y_i} [1 - F(x_i'\beta)]^{(1-y_i)}, \text{ con } y_i = 0 \text{ o } 1.$$

Quando $y = 1$, si ottiene $F(x_i'\beta)$, mentre quando $y = 0$ si ottiene $1 - F(x_i'\beta)$. Per un campione di n variabili indipendenti, la **funzione di log-verosimiglianza** è, dunque, data da:

$$L(\beta) = \sum_{i=1}^n y_i \log(F(x_i'\beta)) + \sum_{i=1}^n (1 - y_i) \log(1 - F(x_i'\beta)).$$

⁵ Immagine tratta da Cappuccio N., Orsi R., 2005. *Econometria*. 2°ed. Bologna: Il Mulino, p. 487.

Lo stimatore di massima verosimiglianza di β , indicato con $\hat{\beta}$, massimizza questa log-verosimiglianza. Se $F(\cdot)$ è la funzione di ripartizione della logistica, allora $\hat{\beta}$ è lo stimatore di massima verosimiglianza nel modello logit; nel caso in cui $F(\cdot)$ sia la funzione di ripartizione della normale standardizzata, $\hat{\beta}$ è lo stimatore di massima verosimiglianza nel modello probit.

3.2 SCELTA DELLE VARIABILI INDIPENDENTI

Rimane da descrivere quali siano le variabili esplicative di questi modelli appena illustrati. Le caratteristiche che, in seguito a un incremento o diminuzione, aumentano la probabilità che l'azienda sia mafiosa possono essere ricavate analizzando il bilancio d'esercizio e valutando se in esso è possibile trovare dei "segnali" che possano far pensare che, dietro un'apparente facciata di legalità, si nasconda invece una gestione criminale.

Questi "segnali" sono rappresentati dagli indici di bilancio: sono degli strumenti che rappresentano rapporti tra voci di bilancio e che consentono di porre in evidenza gli aspetti essenziali della gestione aziendale.

Normalmente, gli indici di bilancio vengono costruiti facendo riferimento allo stato patrimoniale classificato secondo il criterio finanziario e al conto economico a ricavi e costo del venduto. Il database analizzato fornisce però, come detto precedentemente, i bilanci civilistici delle varie aziende; sarebbe, quindi, necessario passare dall'una all'altra riclassificazione. Ciò non è possibile, in quanto, tra le informazioni a disposizione riguardo lo stato patrimoniale, non vengono indicate né le quote dei crediti esigibili entro o oltre l'esercizio né il disagio o l'aggio su prestiti. Non si conosce nemmeno la durata dei crediti verso soci, né la quota a breve e a lungo termine dei ratei e dei risconti attivi e passivi, dei fondi per rischi e oneri e del trattamento di fine rapporto. Di conseguenza, si è optato per calcolare gli indici facendo riferimento agli schemi dettati dal codice civile, apportando le opportune modifiche. Per semplicità verrà mantenuta la stessa nomenclatura.

Gli indici di bilancio ritenuti più significativi ai fini dell'analisi sono sette. Essi vengono costruiti facendo riferimento alle tabelle di stato patrimoniale e conto economico introdotte nel paragrafo 2.1.

Di seguito si fornisce un elenco di tali indici con le relative formule, e una breve descrizione delle motivazioni che hanno spinto a prenderli in considerazione.

- **INDICE DI LIQUIDITÀ:** qui viene calcolato come il rapporto tra le disponibilità liquide e il totale attivo (entrambe ricavate dallo stato patrimoniale):

$$\frac{\text{DISPONIBILITÀ LIQUIDE}}{\text{TOTALE ATTIVO}}$$

È stato scelto tale quoziente in quanto la liquidità, a seconda che sia più o meno elevata, potrebbe essere segnale di un'attività mafiosa, in quanto garantisce la possibilità di avere immediatamente a disposizione risorse per commettere illeciti;

- **ELASTICITÀ DEGLI INVESTIMENTI:** viene ottenuta dividendo il totale delle immobilizzazioni (dato dalla somma di immobilizzazioni materiali, immateriali e finanziarie) per il totale dell'attivo (i valori necessari per il calcolo vengono ricavati dallo stato patrimoniale):

$$\frac{\text{TOTALE DELLE IMMOBILIZZAZIONI}}{\text{TOTALE ATTIVO}}$$

Viene presa in considerazione l'elasticità degli investimenti in quanto consente di comprendere se investimenti più o meno elevati in fattori produttivi possono influenzare la probabilità che l'azienda sia mafiosa.

- **ROA (Redditività dell'attivo netto):** è dato dal rapporto tra il risultato operativo e il totale dell'attivo:

$$\frac{\text{RISULTATO OPERATIVO}}{\text{TOTALE ATTIVO}}$$

Il numeratore è ottenuto dal conto economico, il denominatore dallo stato patrimoniale.

La redditività di un'azienda è uno degli aspetti fondamentali della gestione; non poteva, quindi, mancare un indice indicativo di tale grandezza: il ROA infatti misura il rendimento di tutte le attività della gestione caratteristica e accessoria-patrimoniale, indipendentemente dai metodi con cui sono state raccolte le risorse per il finanziamento e dai costi sostenuti. Non sono compresi in questo quoziente i fattori straordinari e le imposte.

Consente di capire se è più probabile che una buona o una cattiva gestione crei terreno fertile per l'infiltrazione di organizzazioni criminali;

- Il ROE, nonostante sia uno degli indici di bilancio più importanti e significativi, utile sempre per l'analisi della redditività dell'azienda, non viene incluso nella trattazione in quanto, con il tipo di riclassificazione a cui si fa riferimento in questa sede, potrebbe risultare negativo. Questo sarebbe un errore grave, in quanto il ROE è positivo per definizione. Per tale motivo si è optato per utilizzare il ROA.
- **TASSO DI INDEBITAMENTO:** rappresenta in che misura l'azienda ricorre ai mezzi di terzi; in questo caso viene rapportata la differenza tra il totale attivo e il patrimonio netto (ottenendo così proprio l'ammontare dei debiti) al totale attivo stesso:

$$\frac{(\text{TOTALE ATTIVO} - \text{PATRIMONIO NETTO})}{\text{TOTALE ATTIVO}}$$

Sia il totale attivo sia il patrimonio netto provengono dallo stato patrimoniale.

Viene preso in considerazione anche il tasso di indebitamento per avere un'idea di come tale aspetto, che rappresenta un punto critico per chiunque eserciti un'attività d'impresa (in quanto incrementa il rischio di insolvenza) possa influenzare il fatto che un'azienda diventi mafiosa.

- **LN(TOTALE ATTIVO):** grazie al calcolo del logaritmo naturale del totale dell'attivo, ricavato dallo stato patrimoniale, è possibile misurare la dimensione di un'azienda; consente di determinare se è più probabile che le aziende colpite dal fenomeno siano piccole o grandi.
- Gli ultimi due indici a cui si fa ricorso sono rappresentativi della **STRUTTURA DEI COSTI**; si ha quindi:

1) il rapporto tra i costi per il personale e il costo totale della

$$\text{produzione: } \frac{\text{COSTI PER IL PERSONALE}}{\text{COSTO DELLA PRODUZIONE}};$$

2) il rapporto tra i costi sostenuti per i servizi e il costo totale della

$$\text{produzione: } \frac{\text{COSTI PER SERVIZI}}{\text{COSTO DELLA PRODUZIONE}}.$$

I valori necessari sono ricavati dal conto economico.

In conclusione sono stati scelti due indici di tale tipo per valutare se la distribuzione dei costi all'interno di un'azienda possa influenzare la probabilità che un'azienda diventi o no criminale.

Gli indici di bilancio sono stati costruiti attraverso l'utilizzo di Excel: sono state esportate da Stata le colonne relative alle voci di bilancio necessarie per la determinazione dei quozienti e, appunto, grazie alle funzioni di Excel è stato possibile calcolare il valore per ciascuno di essi. Una volta definiti, si è proceduto a determinare le medie.

Infine, è opportuno specificare che la scelta delle variabili indipendenti è avvenuta prima dell'operazione di correzione dei dati, in quanto molti degli errori sono emersi proprio nella fase di costruzione degli stessi.

3.3 PULIZIA DEI DATI

Si tratta, ora, l'ultimo passaggio necessario prima di poter procedere con l'analisi. Come già precedentemente accennato, si è reso necessario correggere alcune incongruenze, alcune già originariamente presenti nel database iniziale, altre, invece, sorte dopo la costruzione degli indici di bilancio.

Nel caso in cui il problema riguardi anche una sola delle voci di bilancio, si è deciso di procedere eliminando l'azienda completamente, e non, quindi, solo il bilancio all'interno del quale si trova l'incongruenza (si ricorda che le società di capitali in questione vengono osservate nel corso degli anni e, quindi, è presente un bilancio d'esercizio relativo ad ogni annualità). Anche nel caso in cui l'errore si trovi in uno solo dei quozienti determinati, si è optato per escludere l'intera azienda dall'analisi. La conseguenza di questa operazione è che il numero di aziende osservate, e in particolare l'ammontare di quelle con la variabile CRIMINAL pari a 1, **si ridurrà** inevitabilmente.

Le correzioni che si è dovuto apportare sono le seguenti:

- innanzitutto, vi sono delle aziende che in alcuni anni presentano la variabile CRIMINAL pari a zero, mentre in altri è pari a uno. In tal caso si è proceduto in questo modo: sono state eliminate le aziende che presentavano un numero uguale o quasi di anni in cui sono considerate mafiose e in cui non lo sono; è stato attribuito il valore 1 nel caso in cui una prevalenza di CRIMINAL = 1 diventava all'improvviso pari a zero; viceversa, è stato assegnato il valore 0 nel caso in cui una prevalenza di CRIMINAL = 0 diventava pari a 1;
- Prima del calcolo degli indici, è stato necessario eliminare tutte le aziende che presentavano disponibilità liquide, totale delle immobilizzazioni, costi per servizi e per il personale e i costi complessivi di produzione negativi;

- Dopo aver determinato gli indici, è emerso che alcuni di essi presentavano valori privi di significato: sono state eliminate quindi tutte le società che presentavano un rapporto disponibilità liquide/totale attivo maggiore di uno, tassi di indebitamento negativi e rapporti costi per i servizi/costo della produzione e costo per il personale/costo della produzione maggiori di uno;
- Per evitare che valori troppo alti o troppo bassi potessero alterare i risultati dell'analisi, tramite Excel sono stati calcolati i percentili, in particolare il primo e il novantanovesimo: tutti i valori dei quozienti di bilancio inferiori al primo percentile o superiori al novantanovesimo sono stati sostituiti con i valori stessi dei percentili. Per fare ciò è però stato necessario eliminare quelle aziende alle quali o mancavano degli indici (in quanto le relative voci di bilancio necessarie per la costruzione non erano state osservate e riportate nel dataset) o risultavano dei quozienti con valore indeterminato (in quanto venivano rapportate grandezze entrambe pari a zero).

In seguito a questa operazione di pulizia dei dati sono rimaste:

- ✓ 5.415 aziende osservate nel settore della costruzione degli edifici (codice Ateco 41), di cui 28 criminali;
- ✓ 5.849 aziende osservate nel settore del commercio all'ingrosso (escluso quello di autoveicoli e motocicli – codice Ateco 46), di cui 9 criminali;
- ✓ 7.498 aziende osservate nel settore delle attività immobiliari (codice Ateco 68), di cui 10 criminali.

3.4 CREAZIONE DEL FOGLIO DATI TRAMITE EXCEL. IL SOFTWARE GRETL

Come si può facilmente notare, il numero delle aziende criminali rispetto al totale è molto basso; si è quindi deciso di compiere un'ulteriore analisi considerando i tre settori assieme, per un totale di 47 società di capitali mafiose su un totale di 18.762 aziende osservate. Pertanto, conclusa la fase di correzione, si è creato un unico foglio Excel in cui nella prima colonna viene indicato il valore della variabile CRIMINAL per ciascuna azienda, mentre nelle successive sono riportati i valori dei quozienti di bilancio. Questo file si aggiunge a quelli realizzati per i singoli segmenti mercato considerati, e anch'essi presentano la stessa struttura appena descritta.

In seguito alla decisione di unire i tre settori, è stata introdotta un'ulteriore variabile indipendente, che assume il nome di COSTRUZIONI: come si può facilmente intuire, essa è pari a uno quando l'azienda svolge un'attività rientrante in tale ambito mentre è pari a zero

altrimenti. Tale variabile fornisce, così, un'ulteriore informazione, ossia se l'appartenenza al settore della costruzione di edifici possa influenzare la probabilità che l'azienda venga infiltrata.

Il foglio Excel si presenta come segue (non sono riportate tutte le colonne per ragioni di spazio):

1	CRIMINAL	INDICE DI LIQUIDITÀ	ELASTICITÀ DEGLI INVESTIMENTI	ROA	TASSO DI INDEBITAMENTO	LN TOTALE ATTIVO
2	1	0,499472481	0,074367276	-0,1943512	0,83178057	11,7551068
3	0	0,021208041	0,572277836	0,03369561	0,628333092	15,05358359
4	0	0,033103851	0,16215646	-0,0265387	0,427741196	14,09254941
5	0	0,20270546	0,183332226	0,21270518	0,35919022	14,4588233
6	0	0,005812696	0,474270003	-0,1038419	1,456677347	15,80025895
7	0	0,007669789	0,411853599	-0,0032679	0,776754534	16,0566267
8	0	0,01321615	0,332137496	-0,0089132	0,810093772	14,9639042
9	0	0,027522462	0	-0,0116066	0,516812773	11,44512514
10	0	0,005658097	0,341905902	0,3206906	0,586864465	14,58933234
11	0	0,008937828	0,334879581	0,0367242	0,145880149	13,85147298
12	0	0,005581342	0,069640861	0,02730876	0,92562235	16,79938453
13	0	0,123526248	0,074971493	0,05395434	0,828868951	13,84909003
14	0	0,007401099	0,029579785	0,01561782	0,9474951	15,1650251

Figura 5: Tabella finale da inserire nel software GRETL relativa al settore del commercio all'ingrosso (codice Ateco 46).

Sia i fogli Excel relativi ai singoli settori, sia quello cumulativo dei tre vengono caricati nel software **Gretl**. Gretl è un software open-source finalizzato a facilitare l'analisi econometrica. Grazie a tale programma è possibile applicare i tre modelli descritti precedentemente e ottenere così i caratteri delle aziende il cui variare aumenta la probabilità che le stesse diventino mafiose.

Si ricorda che il modello di probabilità lineare viene stimato con l'opzione robusta per gli errori standard, a causa dell'eteroschedasticità del termine d'errore.

I risultati dell'analisi verranno analizzati nel prossimo capitolo.

4. COMMENTO DEI RISULTATI OTTENUTI DALL'ANALISI

4.1 ALCUNE CONSIDERAZIONI AI FINI DELLA COMPrensIONE DEI RISULTATI

Come già precedentemente annunciato, in ogni settore economico il numero di aziende la cui variabile CRIMINAL assume valore pari a 1 è assai limitato, pertanto un'analisi che si basasse su così pochi “black swans” rispetto al totale delle osservazioni potrebbe risultare inconcludente. A prova di ciò, vengono ora presentate delle tabelle, dove vengono messi a confronto i risultati ottenuti applicando i tre modelli econometrici ad ogni settore considerato singolarmente. L'ultima tabella riportata mostra invece i risultati che si ottengono analizzando i tre settori insieme. In questo modo sarà più agevole per il lettore poter interpretare e comprendere le conclusioni a cui si perviene.

Una volta ottenuti da Gretl gli output relativi a ciascun codice Ateco, è stato utilizzato lo stesso software statistico per poter costruire tali tabelle. Gli output completi verranno riprodotti nell'appendice alla fine dell'elaborato.

Nella prima colonna sono riportate le variabili indipendenti dei modelli (gli indici di bilancio), mentre nelle altre colonne sono riportati i coefficienti e i p-values (valori tra parentesi) che si ottengono applicando ciascun modello alle osservazioni. Il **p-value**, o livello di significatività osservato, indica la probabilità di ottenere un risultato maggiore o al più uguale del valore osservato, considerando vera l'ipotesi (nel nostro caso l'ipotesi nulla H_0 è che l'indice di bilancio analizzato *non comporti necessariamente un'attività criminale*). I livelli di significatività, rappresentati dal simbolo α , sono stati fissati pari a 10% , 5% , 1%. Se il p-value (che dipende dai dati osservati) assume valori $> \alpha$, l'evidenza empirica non è sufficiente a smentire l'ipotesi nulla, pertanto *viene accettata l'ipotesi nulla*. Se invece il p-value assume valori $< \alpha$, l'evidenza empirica è contraria all'ipotesi nulla, pertanto *viene rifiutata l'ipotesi nulla*.

Particolare attenzione deve essere posta alla terza colonna: infatti, i valori ottenuti dai modelli logit e probit non sono direttamente confrontabili senza un'opportuna decodifica, essendo modelli con varianza (e quindi distribuzione) diversa. Pertanto, i valori ottenuti con il modello probit (che hanno distribuzione normale standardizzata) vengono moltiplicati per $\sqrt{\frac{\pi^2}{3}} \approx 1,81$. Una volta effettuata la trasformazione, possono essere confrontati i risultati. Il file di output del software Gretl fornisce, oltre ai valori di stima dei vari modelli, anche il loro livello di significatività: viene infatti associato a ciascun risultato un asterisco * per $\alpha = 0,10$,

due asterischi ** per $\alpha = 0,05$ e tre asterischi *** per $\alpha = 0,01$. Minore il valore di α , maggiore la significatività, ragion per cui va prestata particolare attenzione ai risultati con ***. Viene ricordato come il modello OLS risulti poco pratico al fine dell'analisi, in quanto presenta problemi legati alla presenza di eteroschedasticità, risulta inefficiente e può portare a risultati non accettabili (valori di probabilità maggiori di 1 o minori di 0), pertanto viene ridimensionata la sua importanza rispetto agli altri due modelli applicati. Per completezza viene anche riportato il coefficiente di determinazione, R^2 , il quale offre una misura della "bontà" del modello (nei casi di logit e probit viene riportato lo **Pseudo- R^2 di McFadden**). È compreso tra 0 e 1: nel primo caso, il modello non è in grado di dire nulla riguardo i dati; nel secondo, invece, il modello li spiega perfettamente. Ne segue, quindi, che tanto più elevato è il valore del coefficiente di determinazione, tanto più il modello si adatta bene alle osservazioni. "lnL" rappresenta invece la log- verosimiglianza.

Successivamente verranno presentati i risultati in termini di significatività dei vari settori, lasciando le conclusioni e discussione dei risultati all'ultimo paragrafo di tale capitolo.

4.2 SIGNIFICATIVITÀ DELLE VARIABILI INDIPENDENTI

TABELLA COMPARATIVA ATECO 41

Analizzando la tabella di output ottenuta dall'analisi dei dati delle aziende con codice ATECO 41 (settore COSTRUZIONE DI EDIFICI), si nota come gli indici maggiormente significativi siano il ROA e il costo del personale. In particolare, considerando principalmente i modelli logit e probit, il primo risulta significativo al 10% , mentre il secondo addirittura all'1%. Scarsa significatività sembra avere la dimensione aziendale, mentre le altre variabili vengono considerate non sufficienti a smentire l'ipotesi nulla.

Variabile dipendente: CRIMINAL				
	Modello 1 OLS	Modello 2 Probit	BProbit x 1,81	Modello 3 Logit
INDICE_DI_LIQUI	0,01921 (0,19531)	0,6133 (0,16160)	1,1101	1,595 (0,16445)
ELASTICITA_DEGL	-7,941e-06 (0,99761)	-0,05151 (0,87963)	-0,0932	-0,1986 (0,83867)
ROA	-0,001666*** (0,00079)	-0,1348* (0,07509)	-0,2440	-0,3252* (0,08012)
TASSO_DI_INDEBI	0,001678 (0,21297)	0,05266 (0,62418)	0,0953	0,04809 (0,85978)
LN_TOTALE_ATTIV	-0,001368 (0,12018)	-0,08387* (0,09140)	-0,1518	-0,2194 (0,11487)
COSTO_PERSONALE	0,01976** (0,02636)	1,211*** (0,00868)	2,1920	3,387*** (0,00663)
COSTO_SERVIZI_S	0,007001 (0,13924)	0,5837 (0,11142)	1,0565	1,796* (0,08899)
n	5415	5415		5415
R-quadro	0,005277	0,06799		0,06635
lnL	6599,05	-163,419		-163,706

P-values tra parentesi
 * indica significatività al livello del 10 per cento
 ** indica significatività al livello del 5 per cento
 *** indica significatività al livello del 1 per cento
 Per logit e probit, R-quadro è lo pseudo-R-quadro di McFadden

TABELLA COMPARATIVA ATECO 46

Per il settore 46 (COMMERCIO ALL'INGROSSO), si può notare come questa volta la dimensione aziendale (LN_TOTALE_ATTIVO) e i costi per la manodopera (COSTO_PERSONALE) siano variabili significative al 10% , mentre le rimanenti risultano trascurabili.

Variabile dipendente: CRIMINAL				
	Modello 1 OLS	Modello 2 Probit	βProbit x 1,81	Modello 3 Logit
INDICE_DI_LIQUI	0,01152 (0,14053)	1,062* (0,09285)	1,9222	3,000 (0,10722)
ELASTICITA_DEGL	0,0002654 (0,91238)	0,1413 (0,82438)	0,2558	0,1057 (0,95722)
ROA	-0,002639 (0,15535)	-0,08384 (0,83760)	-1,5175	-0,1513 (0,88585)
TASSO_DI_INDEBI	-0,0009153 (0,27671)	-0,06771 (0,71574)	-0,1226	-0,2376 (0,66925)
LN_TOTALE_ATTIV	-0,0007357* (0,05559)	-0,1752* (0,05770)	-0,3171	-0,5558** (0,04513)
COSTO_PERSONALE	0,01199 (0,28060)	1,375* (0,08926)	2.4888	4,673** (0,02926)
COSTO_SERVIZI_S	0,001053 (0,76818)	0,1793 (0,73143)	0,3245	0,8019 (0,59555)
n	5849	5849		5849
R-quadro	0,004454	0,1313		0,1375
lnL	10659,6	-58,4513		-58,0339

P-values tra parentesi

* indica significatività al livello del 10 per cento

** indica significatività al livello del 5 per cento

*** indica significatività al livello del 1 per cento

Per logit e probit, R-quadro è lo pseudo-R-quadro di McFadden

TABELLA COMPARATIVA ATECO 68

La tabella successiva, che riporta i risultati dell'analisi dei dati del settore 68 (ATTIVITÀ IMMOBILIARI), risulta apparentemente inconcludente: nessuna variabile viene considerata significativa. Questo risultato può essere in parte dovuto al numero di aziende CRIMINAL

estremamente ridotto rispetto al totale (10/7498). E' possibile tuttavia considerare la variabile COSTO_PERSONALE appena più significativa tra le altre, avendo il p-value minore.

Variabile dipendente: CRIMINAL

	Modello 1 OLS	Modello 2 Probit	βProbit x 1,81	Modello 3 Logit
INDICE_DI_LIQUI	-0,0005549 (0,88648)	-0,1295 (0,88479)	-0,2344	-0,4093 (0,88848)
ELASTICITA_DEGL	-0,0004312 (0,70327)	-0,1039 (0,71341)	-0,1881	-0,3218 (0,72579)
ROA	9,265e-05 (0,95177)	0,02195 (0,97645)	0,0397	0,06290 (0,97855)
TASSO_DI_INDEBI	0,0001805 (0,83954)	0,04153 (0,85379)	0,0752	0,1207 (0,86748)
LN_TOTALE_ATTIV	-9,946e-06 (0,98058)	-0,001768 (0,97963)	-0,0032	-0,003284 (0,98872)
COSTO_PERSONALE	0,004264 (0,41787)	0,7865 (0,43013)	1,4236	2,414 (0,43264)
COSTO_SERVIZI_S	0,0007737 (0,62472)	0,1700 (0,72144)	0,3077	0,5857 (0,70205)
n	7498	7498		7498
R-quadro	0,0001364	0,006192		0,005932
lnL	14184,0	-75,7196		-75,7394

P-values tra parentesi
 * indica significatività al livello del 10 per cento
 ** indica significatività al livello del 5 per cento
 *** indica significatività al livello del 1 per cento
 Per logit e probit, R-quadro è lo pseudo-R-quadro di McFadden

TABELLA COMPLESSIVA

Si passa ora a considerare i tre settori, COSTRUZIONE DI EDIFICI, COMMERCIO ALL'INGROSSO e ATTIVITÀ IMMOBILIARI, in un unico flusso di caricamento.

L'obiettivo di questa ulteriore analisi è duplice:

- Si vuole innanzitutto capire se i risultati ottenuti in precedenza (in particolare per gli ultimi due settori), siano falsati o meno dalla scarsa presenza di CRIMINAL = 1 sul totale: è stato pertanto creato il flusso con lo scopo di aumentarne la percentuale;

- Partendo dall'ipotesi che, tradizionalmente, il settore delle costruzioni (cod. 41) è uno tra quelli tipicamente interessati dall'infiltrazione mafiosa (attraverso gare per ottenere appalti, tangenti, riciclaggio ecc.), ci si è chiesti se l'appartenenza a questo possa influire sulla probabilità che la società sia criminale. Per cercare riscontro a tale questione, si è introdotta un'ulteriore variabile, denominata "COSTRUZIONI". Essa assume valore pari a 1 se la società opera nel settore in questione, 0 altrimenti.

	Modello 1 OLS	Modello 2 Probit	β Probit x 1,81	Modello 3 Logit
INDICE_DI_LIQUI	0,009664* (0,06572)	0,6463** (0,04199)	1,1698	1,750** (0,04842)
ELASTICITA_DEGL	0,001062 (0,24101)	0,06072 (0,74790)	0,1099	0,06351 (0,91373)
ROA	-0,001177*** (0,00026)	-0,1356* (0,07746)	-0,2454	-0,3287* (0,07426)
TASSO_DI_INDEBI	0,0002661 (0,60773)	0,01106 (0,89033)	0,0200	-0,02187 (0,92165)
LN_TOTALE_ATTIV	-0,0006186* (0,05629)	-0,07172** (0,03945)	-0,1298	-0,2139** (0,04110)
COSTO_PERSONALE	0,01341** (0,01206)	1,107*** (0,00169)	2,0043	3,258*** (0,00083)
COSTO_SERVIZI_S	0,002909 (0,12217)	0,4354* (0,07186)	0,7880	1,433** (0,04989)
COSTRUZIONI	0,003515*** (0,00136)	0,3625*** (0,00099)	0,6561	1,057*** (0,00129)
n	18762	18762		18762
R-quadro	0,003705	0,08113		0,08143
lnL	29623,2	-301,797		-301,700

P-values tra parentesi
* indica significatività al livello del 10 per cento
** indica significatività al livello del 5 per cento
*** indica significatività al livello del 1 per cento
Per logit e probit, R-quadro è lo pseudo-R-quadro di McFadden

Dalla tabella riassuntiva, si nota come le variabili considerate significative siano aumentate notevolmente. In particolare, la nuova variabile inserita risulta significativa all' 1% in tutti e tre i modelli, assieme al costo del personale (che è significativo al 5% secondo il modello

OLS ma, come già detto prima, quest'ultimo è meno rilevante degli altri due modelli a cui si fa ricorso). Risultano significative al 5% le variabili relative all'indice di liquidità, alla dimensione aziendale (LN_TOTALE_ATTIVO) e al rapporto tra costi dei servizi su totale. Il fatto che l'indice di liquidità risulti piuttosto significativo solo nel caso in cui i tre settori siano studiati assieme può essere un'ulteriore conferma del fatto che, quando i settori sono studiati singolarmente, la variabile dipendente assuma troppe poche volte valore pari a 1 per poter ottenere dei risultati concludenti e, soprattutto, non fuorvianti. Il ROA risulta invece significativo al 10% (anche in tal caso si prediligono i risultati forniti dal probit e dal logit, per i motivi di cui sopra); non risultano, invece, sufficienti a confutare l'ipotesi nulla i valori dell'elasticità degli investimenti e il tasso di indebitamento dell'azienda, pertanto queste due ultime variabili non possono dire nulla riguardo al fenomeno in questione.

4.3 DISCUSSIONE DEL SEGNO DELLE VARIABILI INDIPENDENTI SIGNIFICATIVE. UN RISCONTRO NELLA REALTÀ

Rimane da valutare quale sia il segno delle variabili che, aumentando o diminuendo, possono far variare la probabilità che l'azienda osservata sia criminale, nonché se è possibile trovare, nella realtà concreta, una spiegazione sul perché tali fattori possano maggiormente attirare l'interesse delle organizzazioni di carattere mafioso.

Partendo innanzitutto da quelle maggiormente significative, è possibile notare come la variabile COSTRUZIONI abbia segno positivo (oltre ad essere estremamente influenzante il fenomeno): ciò sta a indicare che è più facile, avviando un'attività in questo settore nella regione Piemonte, venire infiltrati dalla mafia, rispetto ad operare nell'ambito del commercio o delle attività immobiliari.

L'altra variabile assai significativa è il COSTO DEL PERSONALE sul totale dei costi (significatività all'1% per il settore contrassegnato dal codice Ateco 41 e nel caso dello studio complessivo dei tre settori; significatività al 10% per il codice Ateco 46). In tutti i casi in cui essa risulta determinante nell'influenzare il fenomeno in esame, il relativo segno è positivo: ciò sta a indicare che, se il costo del personale aumenta, allora aumenta anche la probabilità che l'azienda divenga mafiosa. Ciò può essere giustificato considerando che una società che è già disposta a pagare salari e stipendi elevati ai propri dipendenti consente di mascherare un'eventuale ricorso a tangenti, utili al fine di corrompere il personale a perseguire gli scopi criminali dell'organizzazione. Un costo notevole del personale può indicare anche il fatto che la società sia propensa alle assunzioni, e ciò è favorevole per l'associazione mafiosa, in

quanto, come già spiegato nell'introduzione, creare nuovi posti di lavoro è un modo per generare consenso tra le fasce meno abbienti della popolazione.

Per quanto riguarda il resto degli indici di bilancio presi in esame, l'INDICE DI LIQUIDITÀ ha segno positivo (risulta significativo al 5% solo nel caso in cui vengono analizzati i tre settori contemporaneamente). Ne deriva che, anche in questo caso, un aumento della liquidità di una società, aumenta la probabilità che la stessa attiri organizzazioni criminali. Questa conclusione può essere suffragata dal fatto che una società che già dispone di un'elevata liquidità può rivelarsi utile alla mafia in quanto quest'ultima avrebbe a disposizione fin dall'inizio risorse per commettere illeciti.

Anche la variabile COSTO DEI SERVIZI sul totale dei costi è positiva in tutti i casi in cui risulta, anche se non molto, rilevante (significatività al 10% sia nel caso del settore delle costruzioni di edifici sia nel caso dello studio complessivo). Come negli altri casi in cui la variabile ha segno positivo, anche in questo caso un aumento del costo dei servizi fa aumentare la probabilità che l'azienda sia criminale. Nella realtà, si trova riscontro di ciò nel fatto che acquistate servizi può essere un modo per riciclare il denaro proveniente dalle attività illecite compiute, quindi potrebbe risultare meno sospetta una società che già è caratterizzata da un elevato costo per i servizi.

Hanno segno negativo, invece, le ultime due variabili che risultano significative: da ciò deriva, innanzitutto, che più la società è piccola, più è probabile che venga infiltrata (LN TOTALE ATTIVO, variabile significativa al 10% per il settore del commercio all'ingrosso e al 5% nel caso dei tre settori uniti). La mafia risulta quindi, sulla base dei risultati ottenuti, maggiormente attirata da società di piccole dimensioni: in queste realtà è, infatti, più facile nascondere l'infiltrazione ed evitare controlli, in quanto le aziende sono meno "sotto i riflettori". Inoltre è più semplice avvicinare e manipolare l'imprenditore e quanti sono predisposti alla gestione di tali entità.

L'altra variabile dal segno negativo è il ROA, significativo al 10% in entrambi i casi nei quali tale variabile risulta influente (ovvero, nel caso del settore con codice Ateco 41 e nel caso dei settori analizzati assieme). Il segno negativo comporta che, più il ROA è negativo, e quindi tanto peggiore è la gestione dell'impresa, tanto più è facile che l'azienda divenga criminale (si deve comunque tenere presente il fatto che questa variabile non è particolarmente significativa). Questo può essere spiegato dal fatto che è più probabile che, per risanare la situazione della propria società, l'imprenditore ceda alla lusinghe delle organizzazioni criminali.

Infine, come già accennato in precedenza, le variabili ELASTICITÀ DEGLI INVESTIMENTI e TASSO DI INDEBITAMENTO non concorrono a influenzare la probabilità che l'azienda venga infiltrata. Nulla, pertanto, si può dire riguardo a questi due indici di bilancio.

5. CONCLUSIONI

Il lungo procedimento descritto nei capitoli precedenti ha portato a individuare quali siano le caratteristiche che, variando, possono potenzialmente influire sulla probabilità che le società di capitali relative al Piemonte vengano infiltrate da organizzazioni di carattere mafioso. Ciò è stato possibile applicando tre diversi modelli econometrici, il modello di probabilità lineare, il modello probit e quello logit, e scegliendo opportune variabili indipendenti, ovvero gli indici di bilancio.

Risulta, quindi, che i fattori maggiormente rilevanti sono le variabili denominate **COSTRUZIONI** e **COSTI DEL PERSONALE** sul totale dei costi. Entrambe presentano, infatti, una significatività pari all'1%. Per gli altri quozienti di bilancio, invece, la significatività è pari o al 5% o al 10%, eccetto per le variabili **ELASTICITÀ DEGLI INVESTIMENTI** e **TASSO DI INDEBITAMENTO**, che nulla possono dire riguardo il fenomeno in esame.

Questi risultati contribuiscono a sottolineare il fatto che la mafia non è più unicamente un problema delle regioni meridionali italiane, ma è ormai un fenomeno di livello nazionale, che può raggiungere e inquinare qualsiasi attività che viene posta in essere. Uno degli aspetti, infatti, che si premeva sottolineare con questo elaborato è che non si deve nemmeno più credere che le organizzazioni criminali si limitino ad operare nell'economia illegale. Oggi tali organizzazioni possiedono tutti i mezzi e le conoscenze necessarie per poter infiltrare anche quella legale. Le motivazioni sono molteplici e ben spiegate nell'introduzione della tesi. Ciò significa che l'attacco alla mafia deve essere duplice e riguardare, come si è ben dimostrato, entrambi gli ambiti di attività. Ancora maggiore deve essere, quindi, lo sforzo sostenuto dalla collettività per poter combattere questo fenomeno, che da ormai troppo tempo costituisce un notevole problema per l'Italia.

6. TABELLE ALLEGATE

Come conclusione di questo elaborato, per completezza si è deciso di riportare, in allegato, tutte le tabelle di output generate da GRETL, quindi sia quelle relative ai tre settori, singolarmente analizzati, sia quelle riportanti i risultati derivati dallo studio congiunto degli stessi.

CODICE ATECO 41: COSTRUZIONE DI EDIFICI

- MODELLO DI PROBABILITÀ LINEARE

Modello 1: Stime OLS usando le 5415 osservazioni 1-5415

Variabile dipendente: CRIMINAL

Errori standard robusti rispetto all'eteroschedasticità, variante HC1

VARIABILE	COEFFICIENTE	ERRORE STD	STAT T	P-VALUE
const	0,0159339	0,0134516	1,185	0,23625
INDICE_DI_LIQUI	0,0192071	0,0148295	1,295	0,19531
ELASTICITA_DEGL	-7,94054E-06	0,00264637	-0,003	0,99761
ROA	-0,00166586	0,000496225	-3,357	0,00079 ***
TASSO_DI_INDEBI	0,00167767	0,00134690	1,246	0,21297
LN_TOTALE_ATTIV	-0,00136796	0,000880139	-1,554	0,12018
COSTO_PERSONALE	0,0197583	0,00889409	2,222	0,02636 **
COSTO_SERVIZI_S	0,00700096	0,00473404	1,479	0,13924

Media della variabile dipendente = 0,00517082

Scarto quadratico medio della variabile dipendente = 0,0717289

Somma dei quadrati dei residui = 27,7082

Errore standard dei residui = 0,0715857

R-quadro = 0,00527651

R-quadro corretto = 0,00398873

Statistica F (7, 5407) = 2,18337 (p-value = 0,0327)

Log-verosimiglianza = 6599,05

Criterio di informazione di Akaike (AIC) = -13182,1

Criterio bayesiano di Schwarz (BIC) = -13129,3

Criterio di Hannan-Quinn (HQc) = -13163,7

Escludendo la costante, il p-value è massimo per la variabile 3 (ELASTICITA_DEGL)

- MODELLO PROBIT

```
Iterazione 0: log verosimiglianza = -3775,36016655
Iterazione 1: log verosimiglianza = -674,054656876
Iterazione 2: log verosimiglianza = -290,811725449
Iterazione 3: log verosimiglianza = -192,122451028
Iterazione 4: log verosimiglianza = -168,274450502
Iterazione 5: log verosimiglianza = -163,830585137
Iterazione 6: log verosimiglianza = -163,428587513
Iterazione 7: log verosimiglianza = -163,419298249
Iterazione 8: log verosimiglianza = -163,419204731
Iterazione 9: log verosimiglianza = -163,419203892
Iterazione 10: log verosimiglianza = -163,419203885
```

Convergenza raggiunta dopo 11 iterazioni

Modello 2: Stime Probit usando le 5415 osservazioni 1-5415

Variabile dipendente: CRIMINAL

VARIABILE	COEFFICIENTE	ERRORE STD	STAT T	P-VALUE
const	-1,97903	0,737955	-2,682	0,00732 ***
INDICE_DI_LIQUI	0,613266	0,438140	1,400	0,16160
ELASTICITA_DEGL	-0,0515102	0,340148	-0,151	0,87963
ROA	-0,134848	0,0757600	-1,780	0,07509 *
TASSO_DI_INDEBI	0,0526556	0,107473	0,490	0,62418
LN_TOTALE_ATTIV	-0,0838745	0,0496867	-1,688	0,09140 *
COSTO_PERSONALE	1,21058	0,461253	2,625	0,00868 ***
COSTO_SERVIZI_S	0,583658	0,366660	1,592	0,11142

Media di CRIMINAL = 0,005

Numero dei casi 'previsti correttamente' = 5387 (99,5%)

f(beta'x) nella media delle variabili indipendenti = 0,011

Pseudo-R-quadro di McFadden = 0,0679854

Log-verosimiglianza = -163,419

Test del rapporto di verosimiglianza: Chi-quadro(7) = 23,8411 (p-value 0,001215)

Criterio di informazione di Akaike (AIC) = 342,838

Criterio bayesiano di Schwarz (BIC) = 395,614

Criterio di Hannan-Quinn (HQc) = 361,261

- MODELLO LOGIT

```
Iterazione 0: log verosimiglianza = -3767,13669113
Iterazione 1: log verosimiglianza = -755,731141352
Iterazione 2: log verosimiglianza = -330,368722960
Iterazione 3: log verosimiglianza = -206,845639666
Iterazione 4: log verosimiglianza = -172,440274727
Iterazione 5: log verosimiglianza = -164,682432328
Iterazione 6: log verosimiglianza = -163,733346525
Iterazione 7: log verosimiglianza = -163,706421101
Iterazione 8: log verosimiglianza = -163,706387085
```

Convergenza raggiunta dopo 9 iterazioni

Modello 3: Stime Logit usando le 5415 osservazioni 1-5415

Variabile dipendente: CRIMINAL

VARIABILE	COEFFICIENTE	ERRORE STD	STAT T	P-VALUE
const	-3,81599	2,08618	-1,829	0,06737 *
INDICE_DI_LIQUI	1,59502	1,14729	1,390	0,16445
ELASTICITA_DEGL	-0,198583	0,975399	-0,204	0,83867
ROA	-0,325215	0,185836	-1,750	0,08012 *
TASSO_DI_INDEBI	0,0480871	0,272207	0,177	0,85978
LN_TOTALE_ATTIV	-0,219377	0,139141	-1,577	0,11487
COSTO_PERSONALE	3,38689	1,24746	2,715	0,00663 ***
COSTO_SERVIZI_S	1,79608	1,05605	1,701	0,08899 *

Media di CRIMINAL = 0,005

Numero dei casi 'previsti correttamente' = 5387 (99,5%)

f(beta'x) nella media delle variabili indipendenti = 0,004

Pseudo-R-quadro di McFadden = 0,0663475

Log-verosimiglianza = -163,706

Test del rapporto di verosimiglianza: Chi-quadro(7) = 23,2667 (p-value 0,001532)

Criterio di informazione di Akaike (AIC) = 343,413

Criterio bayesiano di Schwarz (BIC) = 396,188

Criterio di Hannan-Quinn (HQc) = 361,835

CODICE ATECO 46: COMMERCIO ALL'INGROSSO (ESCLUSO QUELLO DI AUTOVEICOLI E MOTOCICLI)

- MODELLO DI PROBABILITÀ LINEARE

Modello 1: Stime OLS usando le 5849 osservazioni 1-5849

Variabile dipendente: CRIMINAL

Errori standard robusti rispetto all'eteroschedasticità, variante HC1

VARIABILE	COEFFICIENTE	ERRORE STD	STAT T	P-VALUE
const	0,00930363	0,00525489	1,770	0,07670 *
INDICE_DI_LIQUI	0,0115178	0,00781379	1,474	0,14053
ELASTICITA_DEGL	0,000265358	0,00241132	0,110	0,91238
ROA	-0,00263870	0,00185684	-1,421	0,15535
TASSO_DI_INDEBI	-0,000915321	0,000841408	-1,088	0,27671
LN_TOTALE_ATTIV	-0,000735714	0,000384255	-1,915	0,05559 *
COSTO_PERSONALE	0,0119853	0,0111072	1,079	0,28060
COSTO_SERVIZI_S	0,00105269	0,00357127	0,295	0,76818

Media della variabile dipendente = 0,00153872
 Scarto quadratico medio della variabile dipendente = 0,0391997
 Somma dei quadrati dei residui = 8,94613
 Errore standard dei residui = 0,0391358
 R-quadro = 0,00445362
 R-quadro corretto = 0,00326053
 Statistica F (7, 5841) = 1,14545 (p-value = 0,331)
 Log-verosimiglianza = 10659,6
 Criterio di informazione di Akaike (AIC) = -21303,2
 Criterio bayesiano di Schwarz (BIC) = -21249,8
 Criterio di Hannan-Quinn (HQC) = -21284,6

Escludendo la costante, il p-value è massimo per la variabile 3 (ELASTICITA_DEGL)

- MODELLO PROBIT

Iterazione 1: log verosimiglianza = -672,787255746
 Iterazione 2: log verosimiglianza = -234,258267160
 Iterazione 3: log verosimiglianza = -109,740778998
 Iterazione 4: log verosimiglianza = -72,6056498492
 Iterazione 5: log verosimiglianza = -61,7513470762
 Iterazione 6: log verosimiglianza = -58,9308886823
 Iterazione 7: log verosimiglianza = -58,4742975674
 Iterazione 8: log verosimiglianza = -58,4514877156
 Iterazione 9: log verosimiglianza = -58,4512955929
 Iterazione 10: log verosimiglianza = -58,4512944511
 Iterazione 11: log verosimiglianza = -58,4512944426

Convergenza raggiunta dopo 12 iterazioni

Modello 2: Stime Probit usando le 5849 osservazioni 1-5849
 Variabile dipendente: CRIMINAL

VARIABILE	COEFFICIENTE	ERRORE STD	STAT T	P-VALUE
const	-1,16648	1,22941	-0,949	0,34272
INDICE_DI_LIQUI	1,06150	0,631634	1,681	0,09285 *
ELASTICITA_DEGL	0,141281	0,636654	0,222	0,82438
ROA	-0,0838431	0,409071	-0,205	0,83760
TASSO_DI_INDEBI	-0,0677053	0,185925	-0,364	0,71574
LN_TOTALE_ATTIV	-0,175248	0,0923340	-1,898	0,05770 *
COSTO_PERSONALE	1,37477	0,809022	1,699	0,08926 *
COSTO_SERVIZI_S	0,179273	0,522328	0,343	0,73143

Media di CRIMINAL = 0,002

Numero dei casi 'previsti correttamente' = 5840 (99,8%)

f(beta'x) nella media delle variabili indipendenti = 0,002

Pseudo-R-quadro di McFadden = 0,131279
 Log-verosimiglianza = -58,4513
 Test del rapporto di verosimiglianza: Chi-quadro(7) = 17,666 (p-value 0,013571)
 Criterio di informazione di Akaike (AIC) = 132,903
 Criterio bayesiano di Schwarz (BIC) = 186,295
 Criterio di Hannan-Quinn (HQC) = 151,468

- MODELLO LOGIT

Iterazione 1: log verosimiglianza = -764,520207511
 Iterazione 2: log verosimiglianza = -281,480657011
 Iterazione 3: log verosimiglianza = -129,975492026
 Iterazione 4: log verosimiglianza = -80,1950100486
 Iterazione 5: log verosimiglianza = -64,1051375905
 Iterazione 6: log verosimiglianza = -59,1922660019
 Iterazione 7: log verosimiglianza = -58,1261267554
 Iterazione 8: log verosimiglianza = -58,0348927906
 Iterazione 9: log verosimiglianza = -58,0338769429
 Iterazione 10: log verosimiglianza = -58,0338767875

Convergenza raggiunta dopo 11 iterazioni

Modello 3: Stime Logit usando le 5849 osservazioni 1-5849

Variabile dipendente: CRIMINAL

VARIABILE	COEFFICIENTE	ERRORE STD	STAT T	P-VALUE
const	-0,825361	3,68146	-0,224	0,82261
INDICE_DI_LIQUI	2,99954	1,86214	1,611	0,10722
ELASTICITA_DEGL	0,105717	1,97067	0,054	0,95722
ROA	-0,151350	1,05427	-0,144	0,88585
TASSO_DI_INDEBI	-0,237587	0,556184	-0,427	0,66925
LN_TOTALE_ATTIV	-0,555829	0,277436	-2,003	0,04513 **
COSTO_PERSONALE	4,67336	2,14380	2,180	0,02926 **
COSTO_SERVIZI_S	0,801888	1,51067	0,531	0,59555

Media di CRIMINAL = 0,002

Numero dei casi 'previsti correttamente' = 5840 (99,8%)

f(beta'x) nella media delle variabili indipendenti = 0,001

Pseudo-R-quadro di McFadden = 0,137482

Log-verosimiglianza = -58,0339

Test del rapporto di verosimiglianza: Chi-quadro(7) = 18,5008 (p-value 0,009904)

Criterio di informazione di Akaike (AIC) = 132,068

Criterio bayesiano di Schwarz (BIC) = 185,46

Criterio di Hannan-Quinn (HQC) = 150,633

CODICE ATECO 68: ATTIVITÀ IMMOBILIARI

- MODELLO DI PROBABILITÀ LINEARE

Modello 1: Stime OLS usando le 7498 osservazioni 1-7498

Variabile dipendente: CRIMINAL

Errori standard robusti rispetto all'eteroschedasticità, variante HC1

VARIABILE	COEFFICIENTE	ERRORE STD	STAT T	P-VALUE
const	0,00120867	0,00637285	0,190	0,84958
INDICE_DI_LIQUI	-0,000554887	0,00388663	-0,143	0,88648
ELASTICITA_DEGL	-0,000431189	0,00113194	-0,381	0,70327
ROA	9,26508E-05	0,00153176	0,060	0,95177
TASSO_DI_INDEBI	0,000180541	0,000891603	0,202	0,83954
LN_TOTALE_ATTIV	-9,94636E-06	0,000408545	-0,024	0,98058
COSTO_PERSONALE	0,00426403	0,00526317	0,810	0,41787
COSTO_SERVIZI_S	0,000773727	0,00158164	0,489	0,62472

Media della variabile dipendente = 0,00133369

Scarto quadratico medio della variabile dipendente = 0,0364978

Somma dei quadrati dei residui = 9,9853

Errore standard dei residui = 0,0365123

R-quadro = 0,000136351

R-quadro corretto = -0,000798101

Statistica F (7, 7490) = 0,632492 (p-value = 0,729)

Log-verosimiglianza = 14184

Criterio di informazione di Akaike (AIC) = -28351,9

Criterio bayesiano di Schwarz (BIC) = -28296,6

Criterio di Hannan-Quinn (HQC) = -28332,9

Escludendo la costante, il p-value è massimo per la variabile 6 (LN_TOTALE_ATTIV)

- MODELLO PROBIT

Iterazione 0: log verosimiglianza = -5205,17762949

Iterazione 1: log verosimiglianza = -858,606637062

Iterazione 2: log verosimiglianza = -295,017532166

Iterazione 3: log verosimiglianza = -135,048386717

Iterazione 4: log verosimiglianza = -88,7503256470

Iterazione 5: log verosimiglianza = -77,4795893876

Iterazione 6: log verosimiglianza = -75,8037616440

Iterazione 7: log verosimiglianza = -75,7201578124

Iterazione 8: log verosimiglianza = -75,7196417619

Iterazione 9: log verosimiglianza = -75,7196405513

Iterazione 10: log verosimiglianza = -75,7196405433

Convergenza raggiunta dopo 11 iterazioni

Modello 2: Stime Probit usando le 7498 osservazioni 1-7498

Variabile dipendente: CRIMINAL

VARIABILE	COEFFICIENTE	ERRORE STD	STAT T	P-VALUE
const	-3,04036	1,00753	-3,018	0,00255 ***
INDICE_DI_LIQUI	-0,129470	0,893493	-0,145	0,88479
ELASTICITA_DEGL	-0,103943	0,283008	-0,367	0,71341
ROA	0,0219480	0,743612	0,030	0,97645
TASSO_DI_INDEBI	0,0415308	0,225369	0,184	0,85379
LN_TOTALE_ATTIV	-0,00176786	0,0692332	-0,026	0,97963
COSTO_PERSONALE	0,786469	0,996824	0,789	0,43013
COSTO_SERVIZI_S	0,170002	0,476823	0,357	0,72144

Media di CRIMINAL = 0,001

Numero dei casi 'previsti correttamente' = 7488 (99,9%)

f(beta'x) nella media delle variabili indipendenti = 0,004

Pseudo-R-quadro di McFadden = 0,00619168

Log-verosimiglianza = -75,7196

Test del rapporto di verosimiglianza: Chi-quadro(7) = 0,943506 (p-value 0,995689)

Criterio di informazione di Akaike (AIC) = 167,439

Criterio bayesiano di Schwarz (BIC) = 222,818

Criterio di Hannan-Quinn (HQ) = 186,456

- MODELLO LOGIT

Iterazione 0: log verosimiglianza = -5202,20469858

Iterazione 1: log verosimiglianza = -976,427049299

Iterazione 2: log verosimiglianza = -355,766103726

Iterazione 3: log verosimiglianza = -160,948652298

Iterazione 4: log verosimiglianza = -97,9227339061

Iterazione 5: log verosimiglianza = -79,7863664932

Iterazione 6: log verosimiglianza = -76,0862054254

Iterazione 7: log verosimiglianza = -75,7451944420

Iterazione 8: log verosimiglianza = -75,7394300273

Iterazione 9: log verosimiglianza = -75,7394274512

Convergenza raggiunta dopo 10 iterazioni

Modello 3: Stime Logit usando le 7498 osservazioni 1-7498

Variabile dipendente: CRIMINAL

VARIABILE	COEFFICIENTE	ERRORE STD	STAT T	P-VALUE
const	-6,77904	3,41537	-1,985	0,04716 **
INDICE_DI_LIQUI	-0,409338	2,91902	-0,140	0,88848
ELASTICITA_DEGL	-0,321841	0,917626	-0,351	0,72579
ROA	0,0628960	2,33956	0,027	0,97855
TASSO_DI_INDEBI	0,120668	0,723144	0,167	0,86748
LN_TOTALE_ATTIV	-0,00328352	0,232166	-0,014	0,98872
COSTO_PERSONALE	2,41448	3,07704	0,785	0,43264
COSTO_SERVIZI_S	0,585732	1,53112	0,383	0,70205

Media di CRIMINAL = 0,001
 Numero dei casi 'previsti correttamente' = 7488 (99,9%)
 f(beta'x) nella media delle variabili indipendenti = 0,001
 Pseudo-R-quadro di McFadden = 0,00593198
 Log-verosimiglianza = -75,7394
 Test del rapporto di verosimiglianza: Chi-quadro(7) = 0,903932 (p-value 0,996233)
 Criterio di informazione di Akaike (AIC) = 167,479
 Criterio bayesiano di Schwarz (BIC) = 222,858
 Criterio di Hannan-Quinn (HQC) = 186,496

OUTPUT RELATIVI ALLO STUDIO DEI TRE SETTORI CONGIUNTI

- MODELLO DI PROBABILITÀ LINEARE

Modello 1: Stime OLS usando le 18762 osservazioni 1-18762
 Variabile dipendente: CRIMINAL
 Errori standard robusti rispetto all'eteroschedasticità, variante HC1

VARIABILE	COEFFICIENTE	ERRORE STD	STAT T	P-VALUE
const	0,00651757	0,00486159	1,341	0,18006
INDICE_DI_LIQUI	0,00966440	0,00525106	1,840	0,06572 *
ELASTICITA_DEGL	0,00106221	0,000905934	1,173	0,24101
ROA	-0,00117673	0,000321730	-3,658	0,00026 ***
TASSO_DI_INDEBI	0,000266107	0,000518404	0,513	0,60773
LN_TOTALE_ATTIV	-0,000618564	0,000324044	-1,909	0,05629 *
COSTO_PERSONALE	0,0134052	0,00533909	2,511	0,01206 **
COSTO_SERVIZI_S	0,00290885	0,00188178	1,546	0,12217
COSTRUZIONI	0,00351476	0,00109750	3,203	0,00136 ***

Media della variabile dipendente = 0,00250506
 Scarto quadratico medio della variabile dipendente = 0,0499892
 Somma dei quadrati dei residui = 46,7086
 Errore standard dei residui = 0,0499072
 R-quadro = 0,00370475
 R-quadro corretto = 0,00327973
 Statistica F (8, 18753) = 2,89875 (p-value = 0,00314)
 Log-verosimiglianza = 29623,2
 Criterio di informazione di Akaike (AIC) = -59228,3
 Criterio bayesiano di Schwarz (BIC) = -59157,8
 Criterio di Hannan-Quinn (HQC) = -59205,2

Escludendo la costante, il p-value è massimo per la variabile 5 (TASSO_DI_INDEBI)

- MODELLO PROBIT

```
Iterazione 0: log verosimiglianza = -13041,9557672
Iterazione 1: log verosimiglianza = -2205,72067497
Iterazione 2: log verosimiglianza = -820,961553984
Iterazione 3: log verosimiglianza = -438,977690850
Iterazione 4: log verosimiglianza = -332,858256915
Iterazione 5: log verosimiglianza = -306,508952649
Iterazione 6: log verosimiglianza = -302,045311638
Iterazione 7: log verosimiglianza = -301,798209733
Iterazione 8: log verosimiglianza = -301,797044980
Iterazione 9: log verosimiglianza = -301,797043630
Iterazione 10: log verosimiglianza = -301,797043621
```

Convergenza raggiunta dopo 11 iterazioni

Modello 2: Stime Probit usando le 18762 osservazioni 1-18762
Variabile dipendente: CRIMINAL

VARIABILE	COEFFICIENTE	ERRORE STD	STAT T	P-VALUE
const	-2,40004	0,506183	-4,741	<0,00001 ***
INDICE_DI_LIQUI	0,646335	0,317823	2,034	0,04199 **
ELASTICITA_DEGL	0,0607247	0,188929	0,321	0,74790
ROA	-0,135586	0,0767929	-1,766	0,07746 *
TASSO_DI_INDEBI	0,0110595	0,0802073	0,138	0,89033
LN_TOTALE_ATTIV	-0,0717165	0,0348230	-2,059	0,03945 **
COSTO_PERSONALE	1,10735	0,352681	3,140	0,00169 ***
COSTO_SERVIZI_S	0,435402	0,241886	1,800	0,07186 *
COSTRUZIONI	0,362541	0,110101	3,293	0,00099 ***

Media di CRIMINAL = 0,003

Numero dei casi 'previsti correttamente' = 18715 (99,7%)

f(beta'x) nella media delle variabili indipendenti = 0,005

Pseudo-R-quadro di McFadden = 0,0811332

Log-verosimiglianza = -301,797

Test del rapporto di verosimiglianza: Chi-quadro(8) = 53,2956 (p-value 0,000000)

Criterio di informazione di Akaike (AIC) = 621,594

Criterio bayesiano di Schwarz (BIC) = 692,15

Criterio di Hannan-Quinn (HQC) = 644,75

- MODELLO LOGIT

```
Iterazione 0: log verosimiglianza = -13028,0724048
Iterazione 1: log verosimiglianza = -2496,89782238
Iterazione 2: log verosimiglianza = -968,307864114
Iterazione 3: log verosimiglianza = -499,445768776
Iterazione 4: log verosimiglianza = -353,433894162
Iterazione 5: log verosimiglianza = -311,753103890
Iterazione 6: log verosimiglianza = -302,625032263
Iterazione 7: log verosimiglianza = -301,716051895
Iterazione 8: log verosimiglianza = -301,700096201
```

Iterazione 9: log verosimiglianza = -301,700089694

Convergenza raggiunta dopo 10 iterazioni

Modello 3: Stime Logit usando le 18762 osservazioni 1-18762

Variabile dipendente: CRIMINAL

VARIABILE	COEFFICIENTE	ERRORE STD	STAT T	P-VALUE
const	-4,75961	1,52792	-3,115	0,00184 ***
INDICE_DI_LIQUI	1,75050	0,886910	1,974	0,04842 **
ELASTICITA_DEGL	0,0635065	0,586205	0,108	0,91373
ROA	-0,328688	0,184141	-1,785	0,07426 *
TASSO_DI_INDEBI	-0,0218728	0,222385	-0,098	0,92165
LN_TOTALE_ATTIV	-0,213893	0,104721	-2,043	0,04110 **
COSTO_PERSONALE	3,25792	0,974190	3,344	0,00083 ***
COSTO_SERVIZI_S	1,43331	0,730933	1,961	0,04989 **
COSTRUZIONI	1,05698	0,328367	3,219	0,00129 ***

Media di CRIMINAL = 0,003

Numero dei casi 'previsti correttamente' = 18715 (99,7%)

f(beta'x) nella media delle variabili indipendenti = 0,002

Pseudo-R-quadro di McFadden = 0,0814284

Log-verosimiglianza = -301,7

Test del rapporto di verosimiglianza: Chi-quadro(8) = 53,4895 (p-value 0,000000)

Criterio di informazione di Akaike (AIC) = 621,4

Criterio bayesiano di Schwarz (BIC) = 691,956

Criterio di Hannan-Quinn (HQC) = 644,556

7. BIBLIOGRAFIA

Autieri, D., Baldessarro, G., Di Salvo, M., Palazzolo, S., 2015. Così la Mafia conquista il Nord d'Italia. *Repubblica.it* [online].

Disponibile su < http://inchieste.repubblica.it/it/repubblica/rep-it/2015/03/02/news/mafia_al_nord_la_verit_negata-102063713/>. [Data di accesso: 25/05/2016].

Cappuccio, N., Orsi, R., 2005. *Econometria*. 2° ed. Bologna: Il Mulino.

Cerbioni, F., Cinquini, L., Sòstero, U., 2015. *Contabilità e bilancio*. 4°ed. Milano: McGraw – Hill Education.

De Nova, G., 2015. *Codice civile e leggi collegate*. Bologna: Zanichelli.

Ferrarese, P., Sòstero, U., 2000. *Analisi di bilancio. Strutture formali, indicatori e rendiconto finanziario*. 1°ed. Milano: Giuffrè.

Johnston, J. (Autore), Costa, M. (a cura di), Paruolo, P. (a cura di), 2003. *Econometrica*. 3°ed. Milano: Franco Angeli.

Osservatorio sulla Criminalità Organizzata dell'Università degli Studi di Milano, 2014. *Primo rapporto trimestrale sulle aree settentrionali, per la Presidenza della Commissione Parlamentare di inchiesta sul fenomeno mafioso* [online]. Milano: Università degli Studi di Milano. Disponibile su <<http://www.cross.unimi.it/primo-rapporto-trimestrale-aree-settentrionali/>> . [Data di accesso: 20/05/2016].

Palomba, G., 2004. *Elementi di statistica per l'econometria*. 3°ed. Ancona: Edizioni CLUA.

Transcrime, 2013. *Progetto PON Sicurezza 2007-2013 “ I beni sequestrati e confiscati alle organizzazioni criminali nelle regioni dell'Obiettivo Convergenza: dalle strategie di investimento della criminalità all'impiego di fondi comunitari nel riutilizzo dei beni già destinati”* [online]. Milano: Università Cattolica del Sacro Cuore. Disponibile su <<http://www.transcrime.it/pubblicazioni/progetto-pon-sicurezza-2007-2013/>>. [Data di accesso: 18/05/2016]

Wooldridge, J. M., 2012. *Econometrics: A Modern Approach*. 5°ed. Mason, Ohio: South-Western, Cengage Learning.

SITOGRAFIA

www.agenziaentrate.gov.it

www.brocardi.it

www.cross.unimi.it

www.istat.it

www.transcrime.it

www.treccani.it

NUMERO TOTALE DI PAROLE (ESCLUDENDO INTESTAZIONE, BIBLIOGRAFIA E SITOGRAFIA): 7613