

Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Triennale in
Statistica Economia e Finanza



RELAZIONE FINALE

**Probabilità di possesso e utilizzo di alcuni strumenti di
pagamento: relazione di uno stage svolto presso la
Banca San Giorgio Quinto Valle Agno**

Relatore Prof. Guglielmo Weber
Dipartimento di Scienze Economiche

Laureando: Nicola Manzardo
Matricola N 1006362
Anno Accademico 2013/2014

Indice generale

Introduzione.....	1
1. Lo stage presso la Banca San Giorgio Quinto Valle Agno.....	3
1.1 La Banca San Giorgio Quinto Valle Agno.....	3
1.2 L'esperienza dello stage.....	6
2. I modelli con variabile dipendente discreta.....	9
2.1 Il modello di probabilità lineare.....	11
2.2 I modelli probit e logit.....	12
2.3 Stima di massima verosimiglianza.....	15
2.4 Interpretazione dei parametri stimati ed effetti marginali.....	17
3. Il dataset e i risultati ottenuti.....	19
3.1 Analisi descrittiva delle variabili.....	19
3.2 Approccio econometrico ai dati.....	31
Conclusioni.....	46

Introduzione

La presente relazione ha lo scopo di illustrare il lavoro svolto durante il periodo di stage presso la Banca San Giorgio Quinto Valle Agno con sede a Fara(VI). Il suddetto stage si è svolto da Ottobre 2013 a Gennaio 2014 per un totale di circa 400 ore.

Oltre ad aver imparato il funzionamento di una filiale di banca sotto i molteplici aspetti che la caratterizzano, i quali verranno descritti successivamente nell'apposito paragrafo, ho avuto modo, grazie alla collaborazione della sede centrale e in particolare dell'Ufficio Tecnologia, di ottenere alcuni dati riguardanti le operazioni effettuate presso la filiale di Caldogeno(VI) durante il mese di Ottobre 2013.

Dato il crescente utilizzo di strumenti informatici per l'esecuzione di operazioni bancarie, si è voluto analizzare quali variabili influiscono sul possesso e sull'utilizzo di questi strumenti (home banking, carta di credito, bancomat e "carta ricarica"). Nel caso della Banca San Giorgio Quinto Valle Agno, la funzione telematica di interazione con i conti correnti è fornita da Phoenix Informatica Bancaria S.p.a. e prevede diverse tipologie di prodotto in base alle esigenze del cliente. I dataset che mi sono stati forniti inizialmente erano tre, divisi in "Operazioni fatte allo sportello", "Operazioni tramite home banking" e "Operazioni eseguite sui conti correnti".

Ogni dataset conteneva all'incirca una quindicina di variabili riguardanti il cliente referente della singola operazione; durante lo stage ho potuto riorganizzare e completare le informazioni relative ai clienti, grazie all'utilizzo del programma di gestione anagrafica della banca, ottenendo un singolo file con tutte le variabili di interesse necessarie, senza riferimento all'identità dei singoli clienti (dati anonimizzati). Tramite un approccio econometrico e alla stima di massima verosimiglianza costruiremo un modello che sia in grado di calcolare la probabilità di possedere uno tra bancomat, carta di credito e "carta ricarica";

avere accesso all'home banking e uno dei precedenti strumenti; accedere al sopraccitato servizio da parte di tutti i clienti e da quelli effettivamente in possesso, il tutto in funzione delle variabili esplicative selezionate. Dato che le variabili dipendenti di interesse sono dicotomiche, ovvero possono assumere valore 0 o 1, non si utilizzerà la regressione lineare semplice ma i modelli probit e logit. Questa tipologia di modelli, come vedremo nei paragrafi dedicati, è la più indicata per il trattamento di variabili discrete (e in particolare dicotomiche), come nel nostro caso la probabilità di possesso di uno strumento bancario.

1. Lo stage presso la Banca San Giorgio Quinto Valle Agno

1.1 La Banca San Giorgio Quinto Valle Agno

In questo paragrafo riprendo la storia, lo sviluppo e i valori della banca San Giorgio Quinto Valle Agno direttamente dal sito "www.bancasangiorgio.it".

"Il 27 settembre 1896, nella sacrestia della Chiesa Arcipretale di San Giorgio di Perlina venne fondata dal Parroco Don Gaetano Plebs, insieme a trentuno agricoltori, la "Cassa Rurale di prestiti San Giorgio di Perlina". Essa si prefiggeva come scopo statutario "il miglioramento morale ed economico dei soci, escluso qualunque fine politico". Di fatto l'obiettivo era aiutare concretamente una compatta comunità di piccoli proprietari terrieri, fittavoli e altri agricoltori che operavano, con ridotte possibilità economiche, in un territorio geograficamente circoscritto. Nel 1938, a seguito di nuove disposizioni legislative, la denominazione sociale fu modificata in "Cassa Rurale e Artigiana di S. Giorgio di Fara". Nel 1976 fu assunto un giovane direttore che diede inizio ad un nuovo ciclo della Cassa Rurale. Significativa fu, in particolare, la nuova politica di presidio ed espansione del territorio di competenza, perseguita con l'apertura della prima filiale di Molvena nel 1986, per raggiungere la ventesima resa operativa a fine 2008. Nel gennaio del 2000, a seguito dell'approvazione di un progetto di Sede Staccata nella Valle dell'Agno elaborato in collaborazione con soci, clienti e

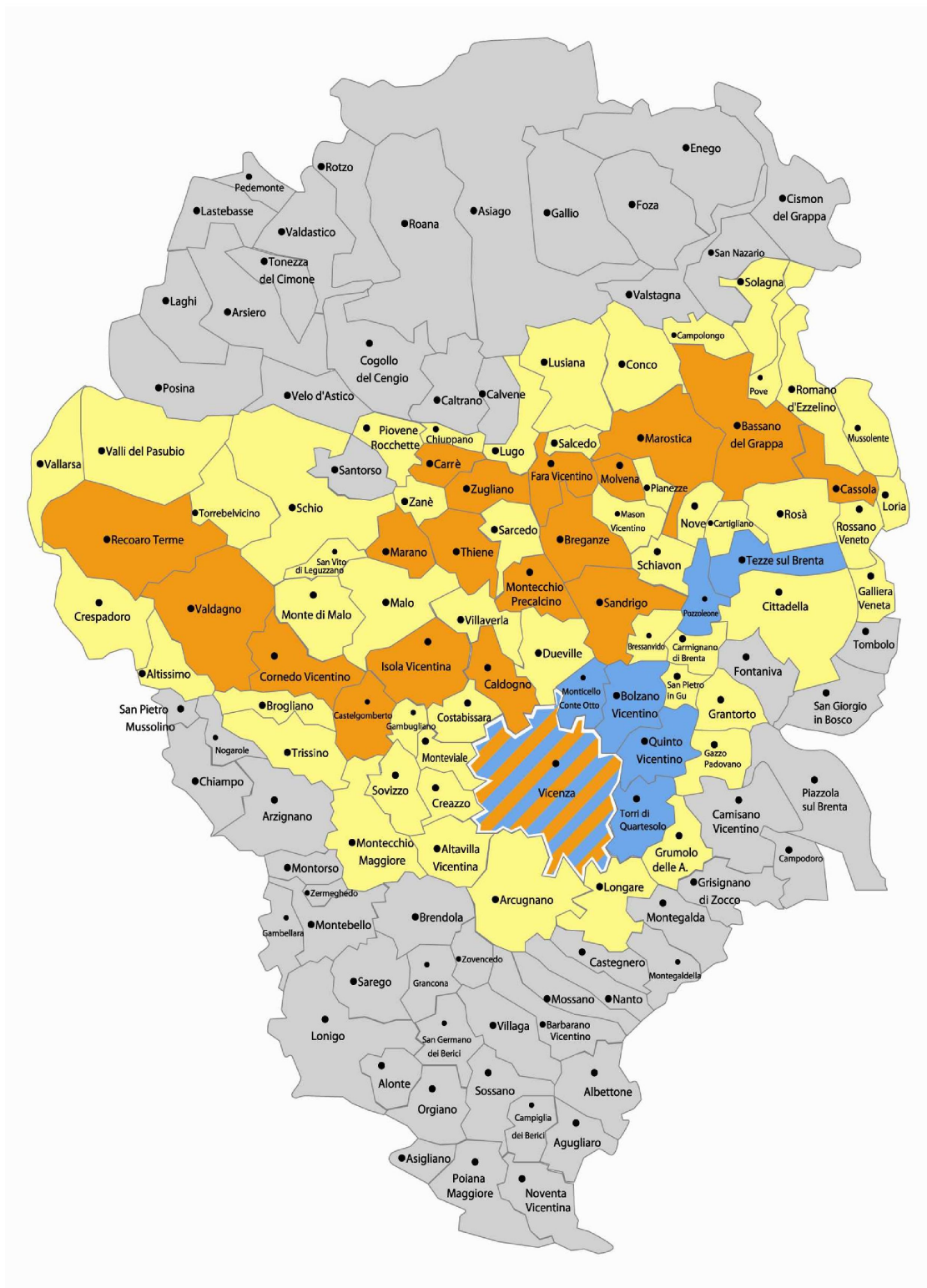


Figura 1: in giallo i comuni di competenza, in arancione le filiali Banca San Giorgio Quinto Valle Agno e in azzurro le filiali ex BCC Quinto Vicentino ora Banca San Giorgio Quinto Valle Agno

imprenditori di quella vallata e finalizzato allo sviluppo, anche in quell'area, degli ideali di cooperazione e mutualità, oltre che per colmare l'esigenza di una banca locale, l'Assemblea Straordinaria delibera la variazione della denominazione sociale in "Banca San Giorgio e Valle Agno - Credito Cooperativo di Fara Vic". Detto progetto trova attuazione con l'apertura di una Sede Staccata a Cornedo Vicentino nell'aprile 2000 e delle filiali di Valdagno - Piazzetta San Giorgio e Via 7 Martiri -, Castelgomberto, Recoaro Terme. Nel Luglio 2011 Banca San Giorgio e Valle Agno, dopo l'approvazione di Banca d'Italia e le deliberazioni favorevoli delle assemblee straordinarie dei soci, conclude l'iter di fusione per l'incorporazione della BCC di Quinto Vicentino. La denominazione sociale risultante è "Banca San Giorgio Quinto Valle Agno - Credito Cooperativo" con efficacia dal 1° Gennaio 2012. Ad oggi la banca conta 29 filiali, che operano in un territorio che copre tutta l'area pedemontana della provincia di Vicenza, come illustrato in *figura 1*.

L'attività di Banca San Giorgio Quinto Valle Agno "si ispira ai principi cooperativi della mutualità senza fini di speculazioni privata" e persegue "lo scopo di favorire i soci e gli appartenenti alle comunità locali nelle operazioni e nei servizi di banca". Particolarmente rilevante è l'importanza del localismo che si manifesta con l'attenzione al territorio dove opera. Si cerca infatti di favorirne "il miglioramento delle condizioni morali, culturali ed economiche" e di promuovere "lo sviluppo della cooperazione e l'educazione al risparmio e alla previdenza". La Banca continua a perseguire la politica di agevolazione ai soci e di riversamento nel territorio di parte degli utili conseguiti al fine di sostenere iniziative mutualistiche e territoriali, di carattere solidale, culturale, sportivo e promozionale. Banca San Giorgio Quinto Valle Agno è una Banca di Credito Cooperativo - BCC: è una banca particolare, differente, originale a partire dalla sua missione e dai suoi valori: è una società cooperativa senza finalità di lucro, dove si vive la rara esperienza della democrazia economica in una logica di imprenditorialità. Non si limita cioè a cercare la massimizzazione dei profitti (o la

minimizzazione dei costi) come le banche commerciali, ma trova la sua ragion d'essere nell'assicurazione di un vantaggio ai soci e alle comunità locali. Banca San Giorgio Quinto Valle Agno è una banca mutualistica, in quanto eroga il credito principalmente ai propri soci. Il suo obiettivo è quello di favorire la loro partecipazione alla vita economica e sociale, di porre ciascuno di essi nelle condizioni di essere, almeno in parte, autore del proprio sviluppo come persona o come impresa. Nel contempo Banca San Giorgio Quinto Valle Agno è anche banca radicata sul territorio (i soci sono espressione del contesto in cui l'azienda opera); per il territorio (il risparmio raccolto finanzia lo sviluppo dell'economia reale); nel territorio (sostiene il contesto sociale, culturale e morale della realtà in cui opera). Questo vuol dire che le sue radici e la sua vitalità nascono dal territorio e al territorio ritornano. La specifica vocazione ad operare nei mercati locali, unita ad un assetto istituzionale che implica i principi cooperativistici e il vincolo della mutualità, caratterizza la sua l'essenza. Le BCC italiane, per non perdere i vantaggi legati alla piccola dimensione, si sono strutturate in un sistema nazionale che prende il nome di Credito Cooperativo: assicurare l'integrazione, le sinergie e le economie di scala tra e per le singole aziende. Le BCC formano una grande rete di banche locali e costituiscono un sistema creditizio capillarmente diffuso sull'intero territorio nazionale: ogni BCC conosce da vicino la realtà del territorio in cui è presente, ma ha la forza di appartenere ad un sistema. Questa struttura ruota intorno alle 438 aziende complessivamente presenti sul territorio nazionale con 4.044 sportelli (l'11,8% del panorama bancario italiano), attraverso una presenza diretta in 2.576 Comuni (operatività in circa 4 mila) e in 98 Province; in 542 comuni rappresentano l'unica realtà bancaria. Le BCC sono supportate da un lato dal sistema associativo (Federazioni Regionali e Federazione Nazionale che svolgono funzioni di coordinamento, assistenza tecnica, consulenza ed erogazione di servizi a favore delle banche associate) e dall'altro dal sistema imprenditoriale (Gruppo Iccrea Banca, Cassa Centrale e organismi consortili che riguardano la fornitura di prodotti e servizi).”

1.2 L'esperienza dello stage

Mi è stata data la possibilità di svolgere lo stage presso la filiale della Banca San Giorgio Quinto Valle Agno di Caldogeno(VI). Quest'ultima era composta da 5 figure professionali: un direttore, un vicedirettore, un consulente, un operatore di “back office” (“retrospostello”) e un cassiere.

Il direttore ha la funzione di supervisore della struttura filiale e ha il compito di monitorare gli indici necessari alla realizzazione del budget, condividendo periodicamente i risultati ottenuti con la sede centrale. Detiene anche il potere di delega del credito ai clienti, nei limiti consentiti dal Consiglio di Amministrazione.

In materia di sviluppo ha il compito di raccogliere indicazioni sulla potenzialità delle aree riguardanti la propria zona di competenza per indirizzare le eventuali iniziative connesse. Gestisce i rapporti con la clientela e cura l'apertura di conti correnti, il perfezionamento della concessione dei fidi e delle garanzie ricevute, gli incarichi ricevuti dalla clientela per la prestazione di servizi di intermediazione mobiliare e raccolta ordini.

Il vicedirettore, oltre a fare le veci del direttore in caso di assenza di quest'ultimo, nella filiale in oggetto aveva la funzione di consulenza per quanto concerne l'acquisto/vendita di titoli, conti deposito, stipula di polizze assicurative e informazioni riguardanti la posizione finanziaria del cliente.

Il consulente segue tutto ciò che riguarda l'area creditizia: cura la raccolta delle domande di fido corredandole di informazioni e gestisce le pratiche per la richiesta di mutui ipotecari/chirografari, il tutto a stretto contatto con l'Area Crediti della sede centrale e in caso di necessità con il Polo Fidi.

L'operatore di “back office”, oltre a eseguire operazioni bancarie non riguardanti l'utilizzo del contante, in quanto in questa filiale sprovvisto di cassa, ha principalmente il compito di seguire le aziende nei rapporti che intercorrono tra queste ultime e la banca stessa, come anticipo fatture e pagamenti vari. Lavora inoltre a stretto contatto con il direttore per quanto concerne il controllo giornaliero dei pagamenti in arrivo sui conti correnti, accertando la disponibilità

di pagamento.

Il cassiere infine provvede alla gestione del contante ed esegue tutte quelle operazioni bancarie richieste dalla clientela, tra le quali il prelievo/versamento di contanti, il versamento/cambio assegni della propria banca o di altre banche, l'emissione di assegni circolari, il pagamento deleghe, bonifici, MAV, RAV.

Durante i primi giorni dello stage ho avuto modo di conoscere il funzionamento della filiale nella sua totalità, affiancando in particolar modo l'operatore di "back office", il quale mi ha illustrato le principali operazioni che vengono svolte in banca.

Successivamente ho svolto un lavoro di raccolta firme concernenti il "questionario di adeguata verifica", un documento nel quale il cliente conferma di essere il titolare del conto corrente a lui intestato e di non utilizzarlo per scopi di riciclaggio o altre azioni illegali. Data una lista di clienti privi di questionario firmato, ho avuto il compito di contattarli telefonicamente e assicurarmi che venissero a firmare, data l'imminente scadenza dei termini di firma.

Finito questo lavoro, ho collocato i vari questionari nelle apposite cartelline contenenti tutti i documenti relativi al cliente in questione.

Per l'archiviazione dei documenti, in generale, la clientela è divisa in "affidata" o "non affidata": gli "affidati" sono clienti ai quali è stato concesso un fido dalla banca, mentre i "non affidati" sono correntisti che non ne hanno mai fatto richiesta.

Dato che i due tipi di clientela occupano archivi separati l'uno dall'altro, per ogni correntista occorre un precedente controllo di appartenenza tramite anagrafe informatica.

Dopo qualche settimana mi sono occupato della scansione e archiviazione, sia in archivio che nel sistema anagrafico informatico, di alcuni documenti riguardanti la clientela, tra cui documenti d'identità, rapporti generali tra banca/cliente e moduli per la privacy.

Durante lo svolgimento di queste mansioni ho avuto il compito di rispondere alle chiamate da parte dei clienti: inizialmente le giravo ai colleghi mentre, col passare del tempo, sono riuscito a gestire autonomamente alcune richieste, tra le quali informazioni riguardanti la conferma di codici IBAN o la verifica di movimenti sui conti correnti. Durante l'ultimo periodo ho avuto modo di elaborare i file forniti dalla sede centrale grazie al contributo del sistema informativo SIB2000, con il quale ho potuto sviluppare e implementare i dati.

2. I modelli con variabile dipendente discreta

Per poter affrontare modelli con variabili dicotomiche dobbiamo prima partire dalla definizione di econometria e di regressione lineare.

Come scritto da *Maddala* in *“Introduction to Econometrics”*, *“L'econometria è l'applicazione di metodi matematici e statistici all'analisi di dati economici, con l'obiettivo di dare contenuto empirico alle teorie economiche e verificarle o rifiutarle”*. L'econometria ha quindi due obiettivi principali: la previsione del futuro e la comprensione delle correlazioni presenti tra le variabili in esame.

Per fare ciò, lo strumento maggiormente utilizzato è la regressione lineare, che mette in relazione la variabile dipendente con le variabili indipendenti, secondo il modello riportato, valido per una singola osservazione:

$$y = \beta_0 + x_1 \beta_1 + \dots + x_n \beta_n + \varepsilon$$

dove:

- y rappresenta la variabile dipendente;
- x_i rappresenta l' i -esima variabile indipendente o esplicativa;
- β_i è l' i -esimo coefficiente che verrà successivamente stimato;
- ε è l'errore non misurabile che si commette durante il processo.

Come indicato in *“Econometria”* di *Cappuccio e Orsi*, le ipotesi che stanno alla base della regressione lineare sono:

a) Linearità. La funzione di regressione è lineare:

$$E(y_i | x_1, \dots, x_i, \dots, x_n) = \beta_0 + \beta_1 x_i, \quad i = 1, \dots, n$$

b) Omoschedasticità. La varianza condizionale è costante al variare delle osservazioni:

$$Var(y_i | x_1, \dots, x_i, \dots, x_n) = \sigma^2, \quad i = 1, \dots, n$$

c) Incorrelazione. Le osservazioni y_i sono tra loro incorrelate in relazione alle x_i :

$$\text{Cov}(y_i, y_j | x_1, \dots, x_i, \dots, x_n) = 0, \quad \text{per } i \neq j$$

d) Rango pieno. Nessuna variabile x_i può essere scritta come funzione lineare delle altre variabili indipendenti del modello.

Durante lo studio di modelli che descrivono comportamenti socio-economici, è frequente imbattersi in variabili dipendenti qualitative, nel nostro caso dicotomiche, ovvero che possono assumere solo due modalità. Ad esempio il possesso o meno di titoli azionari, l'utilizzo o meno di mezzi pubblici, l'avvenimento di un dato evento prima o dopo una certa data etc...

Data la particolare natura della variabile dipendente, nessuno di questi casi si presta direttamente ad analisi di regressione lineare. Essi possono essere tuttavia analizzati costruendo modelli che, al variare di un insieme di fattori, ritornano una risposta o una scelta. L'approccio più diffuso è quello che lavora con modelli di probabilità. Se consideriamo una variabile con risposta binaria, per la quale viene dato valore 1 se l'evento si verifica e 0 se l'evento non si verifica, la nostra attenzione ricade sulle seguenti risposte di probabilità:

$$P(y_i=1|x_i) = F(x_i, \beta)$$

$$P(y_i=0|x_i) = 1 - F(x_i, \beta)$$

dove x_i è il vettore delle variabili esplicative mentre β è il vettore dei parametri.

Nel prossimo paragrafo vedremo come questo problema viene affrontato attraverso il modello di probabilità lineare, che pur essendo di semplice applicazione, presenta diverse problematiche dal punto di vista statistico ed econometrico.

2.1 Il modello di probabilità lineare

Riprendendo quanto scritto in *“Econometria”* di Cappuccio e Orsi, descriviamo la costruzione di questo modello iniziando col definire una funzione di regressione lineare. Specifichiamo la variabile dicotomica y_i data da:

$$y_i = \begin{cases} 1 & \text{se si realizza l'evento analizzato} \\ 0 & \text{in caso contrario} \end{cases}$$

Si suppone quindi che il valore atteso subordinato a un vettore $K \times 1$ di variabili esplicative x_i , corrisponda a una funzione lineare del tipo:

$$E(y_i|x_i) = \beta' x_i = \sum_{j=1}^n \beta_j x_{ij}$$

dove β corrisponde al vettore dei coefficienti di regressione.

La variabile casuale y_i si distribuisce dunque come una bernoulliana, con valore atteso subordinato alla probabilità che si realizzi l'evento $\{y_i\}$; si ha infatti:

$$E(y_i|x_i) = 1 \cdot P(y_i=1|x_i) + 0 \cdot P(y_i=0|x_i) = P(y_i=1|x_i)$$

Essendo la probabilità di risposta lineare nei parametri si ha che:

$$P(y_i=0|x_i) = 1 - P(y_i=1|x_i)$$

La stima dei parametri si può calcolare con il metodo dei Minimi Quadrati Ordinari. Pur risultando gli errori in media indipendenti dalle variabili esplicative, data la natura delle variabili di errore i primi risultano affetti da eteroschedasticità condizionale. Quest'ultima porta lo stimatore a essere non efficiente, avendo violato una delle ipotesi principali della regressione lineare.

Se questo problema può essere risolto tramite l'utilizzo di standard error robusti o la stima dei minimi quadrati ponderati (WLS), esiste un problema di difficile risoluzione a meno di non perdere di flessibilità. Per alcuni valori delle variabili esplicative può succedere che i valori della corrispondente variabile dipendente

giacciono al di fuori dell'intervallo $[0,1]$. Tali risultati porterebbero a misure di probabilità senza senso, cioè negative o maggiori di 1.

Nonostante quindi la facilità di stima tramite i Minimi Quadrati Ordinari, la facile interpretazione dei parametri e il risolvibile problema di eteroschedasticità, la mancanza di vincolo della probabilità porta a preferire altri tipi di modelli per i casi di variabile indipendente dicotomica.

2.2 I modelli probit e logit

In questo paragrafo prenderemo spunto da *Modelli a risposta discreta: "Lezioni dottorato-Econometria III"* di Marianna Belloc per approfondire i modelli probit e logit. Per poter risolvere il problema affrontato nel paragrafo precedente faremo uso di un particolare modello, chiamato generalmente **modello indice**, che produce la seguente risposta di probabilità binaria:

$$P(y_i=1|x_i)=G(x_i'\beta)\equiv p(x_i)$$

dove si assume che $0 < G(z) < 1$ per ogni $z \in \mathbb{R}$ e nella maggior parte delle applicazioni è considerata una funzione cumulata di probabilità.

Il modello indice viene ricavato dal seguente **modello a variabile latente**:

$$y_i = x_i'\beta + \varepsilon_i \quad y_i = 1 \quad [y_i^* > 0]$$

dove ε_i è una variabile indipendente da x_i e la sua distribuzione viene molto frequentemente assunta simmetrica allo zero.

La funzione $y_i = 1 \quad [y_i^* > 0]$ viene chiamata **funzione indicatore** perché ci dice che y_i assume valore 1 ogni volta che la condizione tra parentesi viene soddisfatta (in questo caso che sia positiva, ma all'interno delle quadre può esserci qualsiasi altro valore soglia posto come condizione). Da questo risulta che:

$$y_i = \begin{cases} 1 & \text{se } y_i^* > 0 \\ 0 & \text{se } y_i^* \leq 0 \end{cases}$$

La y_i^* viene dunque chiamata **variabile latente**, la quale non viene osservata direttamente, bensì viene osservato un indicatore del suo segno, ovvero y_i .

Il problema da risolvere rimane quello riguardante la distribuzione della variabile ε_i ; i due approcci più utilizzati, come vedremo, sono l'utilizzo dei modelli **probit** e **logit**.

Nel **modello probit** si ipotizza che la variabile ε_i si distribuisca come una variabile normale standard, con $E[\varepsilon_i|x_i]=0$ e $i=1,\dots,n$. Si ha quindi:

$$E[y_i|x_i]=F(x_i'\beta)=\Phi(x_i'\Sigma)\int_{-\infty}^{(x_i'\beta)}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2}\right\}dx$$

Nel **modello logit** invece la variabile ε_i si ipotizza distribuita come una logistica standard, con $E[\varepsilon_i|x_i]=0$ e $V(\varepsilon_i|x_i)=\pi^2/3$. Si ha quindi:

$$E[y_i|x_i]=F(x_i'\beta)=\Lambda(x_i'\beta)=\frac{\exp\{x_i'\beta\}}{1+\exp\{x_i'\beta\}}$$

I due modelli in generale sono molto simili e tendono a produrre valori di probabilità analoghi rendendo difficile esprimere una preferenza su base empirica. La distribuzione logistica, diffusa solitamente in Medicina e Demografia, è di poco preferibile in presenza di molti 0 o 1. Inoltre quest'ultima presenta code più pesanti, come si può notare dalla *figura 2* e *figura 3*. Il modello probit è invece più utilizzato in Economia, dove l'ipotesi di distribuzione normale è più diffusa. Per quanto riguarda l'interpretazione dei coefficienti, in entrambi i modelli i parametri non sono direttamente decifrabili. Tuttavia, per ricavare informazioni utili al fine dell'inferenza sul modello, è possibile utilizzare il segno dei coefficienti, che influenza positivamente o negativamente la probabilità di successo o insuccesso e, come vedremo nei prossimi paragrafi, l'utilizzo degli effetti marginali.

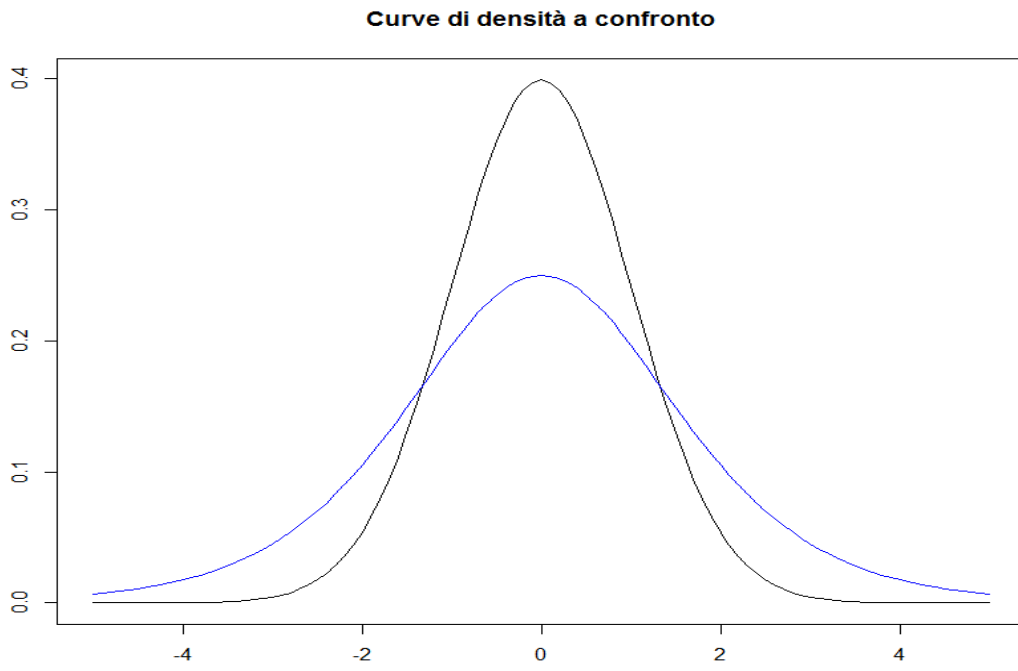


Figura 2: in blu la curva logistica e in nero la curva normale.

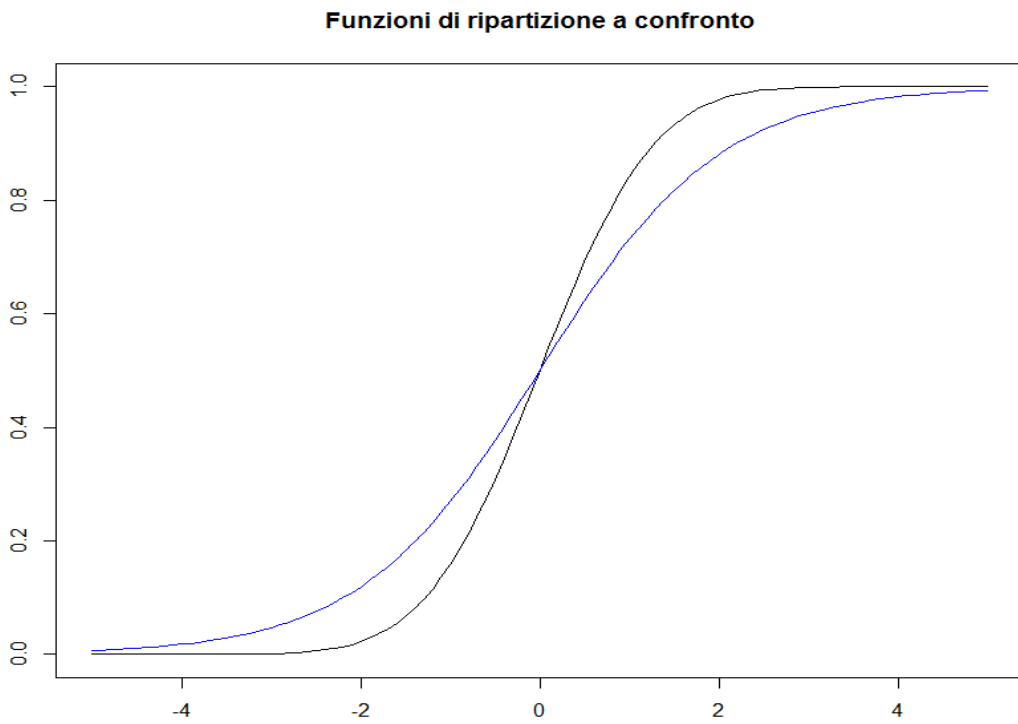


Figura 3: in blu la funzione logistica e in nero la funzione normale.

2.3 Stima di massima verosimiglianza

In questo paragrafo spiegheremo, con l'aiuto degli appunti delle "Lezioni di introduzione all'econometria: Modelli per variabili dipendenti limitate" di Alessandro Buccioli, il metodo di stima di massima verosimiglianza per i modelli probit e logit. A differenza del modello di probabilità lineare, dove la stima dei parametri veniva effettuata tramite i Minimi Quadrati Ordinari, nei modelli probit e logit viene utilizzato il metodo della massima verosimiglianza. La funzione di densità di y_i dato x_i può essere scritta nel seguente modo:

$$f(y_i|x_i;\beta) = [G(x_i'\beta)]^{y_i} [1 - G(x_i'\beta)]^{1-y_i}$$

Per un campione di n osservazioni la funzione di verosimiglianza è quindi data da:

$$L(\beta) = \prod_{i=1}^n [G(x_i'\beta)]^{y_i} [1 - G(x_i'\beta)]^{1-y_i}$$

dalla quale si ottiene facilmente la log-verosimiglianza che risulta:

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

Per trovare la stima di massima verosimiglianza di β è necessario massimizzare la funzione di massima verosimiglianza. Per fare ciò si deve ottenere la derivata prima di quest'ultima, che viene chiamata in ambito statistico **funzione score**, e successivamente porla uguale a 0, risolvendo il sistema di condizioni del primo ordine, come illustrato di seguito:

$$\max_{\beta} l(\beta)$$

$$\frac{\partial l(\beta)}{\partial \beta} = \sum_{i=1}^n \left[\frac{y_i - G(x_i'\beta)}{G(x_i'\beta)(1 - G(x_i'\beta))} g(x_i'\beta) \right] x_i = 0$$

dove $g(z)$ indica la funzione di densità di probabilità ottenuta dalla $G(z)$, ovvero $\partial G(z)/\partial z$.

Il termine tra parentesi quadra viene chiamato **residuo generalizzato** ed è pari a:

$$\tilde{\epsilon}_i = \frac{g(x_i' \beta)}{G(x_i' \beta)} \quad \text{per } y_i = 1$$

$$\tilde{\epsilon}_i = -\frac{g(x_i' \beta)}{1 - G(x_i' \beta)} \quad \text{per } y_i = 0$$

Sia nel modello probit che in quello logit il risultato non è disponibile in forma chiusa; è necessario quindi ricorrere all'utilizzo di calcoli numerici, come ad esempio il metodo di Newton-Raphson.

Per il modello probit, dopo alcuni passaggi, si arriva alla formulazione della condizione del primo ordine:

$$\frac{\partial l(\beta)}{\partial \beta} = \sum_{y_i=0} \frac{-\phi(x_i' \beta)}{1 - \Phi(x_i' \beta)} x_i + \sum_{y_i=1} \frac{-\phi(x_i' \beta)}{\Phi(x_i' \beta)} x_i = 0$$

nel caso logit si ottiene invece:

$$\frac{\partial l(\beta)}{\partial \beta} = \sum_{i=1}^n [y_i - \Lambda(x_i' \beta)] x_i = 0$$

Per quanto riguarda la derivata seconda, nel modello probit il calcolo risulta un po' laborioso; per aiutarci scriveremo le seguenti semplificazioni:

$$\psi_i = \psi_{0i} = \frac{-\phi(x_i' \beta)}{1 - \Phi(x_i' \beta)} \quad \text{se } y_i = 0$$

$$\psi_i = \psi_{1i} = \frac{\phi(x_i' \beta)}{\Phi(x_i' \beta)} \quad \text{se } y_i = 1$$

grazie a queste facilitazioni possiamo riscrivere la funzione score come:

$$\frac{\partial l(\beta)}{\partial \beta} = \sum_{y_i=0} \psi_{0i} x_i + \sum_{y_i=1} \psi_{1i} x_i = \sum_i \psi_i x_i = 0$$

la derivata seconda risulta quindi:

$$\frac{\partial^2 l(\beta)}{\partial \beta \partial \beta'} = -\sum_i \psi_i (\psi_i + x_i' \beta) x_i x_i' < 0$$

che risulta negativa per tutti i valori di β .

Passando al modello logit, il calcolo della derivata seconda si dimostra molto più semplice; ne consegue il seguente risultato:

$$\frac{\partial^2 l(\beta)}{\partial \beta \partial \beta'} = - \sum_i \Lambda(x_i' \beta) [1 - \Lambda(x_i' \beta)] x_i x_i' < 0$$

come è facile controllare la derivata seconda è sempre negativa per tutti i valori di β .

2.4 Interpretazione dei parametri stimati ed effetti marginali

Prendendo spunto da *Modelli a risposta discreta: "Lezioni dottorato-Econometria III"* di Marianna Belloc illustreremo come vengono interpretati gli effetti marginali nei modelli logit e probit.

Come anticipato nei paragrafi precedenti, il generico parametro β_k non può essere interpretato come contributo relativo della variabile k-esima alla definizione dell'evento certo.

Oltre al segno di β_k , che ci indica se la variabile esplicativa influisce positivamente o negativamente sull'esito dell'evento, è possibile calcolare l'effetto parziale della variazione della k-esima variabile esplicativa.

Nell'affrontare questo problema analizzeremo due casi in particolare; il caso nel quale la variabile esplicativa d'interesse è continua e il caso nel quale la variabile esplicativa d'interesse è discreta.

Se la variabile x_k è continua si ha genericamente:

$$\frac{\partial p}{\partial x_{ki}} = g(x_i' \beta) \cdot \beta_k \quad \text{dove} \quad g(z) \equiv \frac{\partial G(z)}{\partial z}$$

Per i modelli probit e logit, gli effetti marginali risultano rispettivamente:

$$\frac{\partial p}{\partial x_{ki}} = \phi(x_i' \beta) \cdot \beta_k$$

$$\frac{\partial p}{\partial x_{ki}} = \Lambda(x_i' \beta) \cdot [1 - \Lambda(x_i' \beta)] \cdot \beta_k$$

Poiché gli effetti marginali non sono costanti, essi vengono valutati utilizzando la media campionaria delle osservazioni sulle variabili:

$$\frac{1}{N} \cdot \sum_{i=1}^N g(x_i' \beta) \cdot \beta_k$$

L'effetto marginale non dipende quindi da x_i : per variabili continue x_{ki} e x_{hi} , il rapporto degli effetti marginali è costante ed è dato da:

$$\frac{\partial p(x_i) / \partial x_{ki}}{\partial p(x_i) / \partial x_{hi}} = \frac{\beta_k}{\beta_h}$$

Se stiamo invece trattando una variabile discreta, nel nostro caso binaria, si ricorre al calcolo delle differenze tra le probabilità per valutare l'effetto marginale:

$$F(\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \dots + \beta_K x_{Ki}) - F(\beta_0 + \beta_1 x_{1i} + \dots + \beta_K x_{Ki})$$

L'effetto marginale in questo caso dipende da tutte le variabili che compongono il vettore x .

Nel caso infine si tratti di una variabile discreta non binaria, cioè che può assumere più di due valori, si fa ricorso ad una generalizzazione dell'espressione utilizzata per il calcolo dell'effetto marginale di una variabile binaria. Se x_k è una variabile discreta che assume i valori $c_1, c_2, \dots, c_k, \dots, c_K$ e si vuole calcolare l'effetto di una variazione di x_k da c_k a c_k+1 , avremo:

$$F(\beta_0 + \beta_1 x_{1i} + \dots + \beta_k (c_k + 1) \dots + \beta_K x_{Ki}) - F(\beta_0 + \beta_1 x_{1i} + \dots + \beta_k c_k \dots + \beta_K x_{Ki})$$

3. Il dataset e i risultati ottenuti

Il dataset con il quale lavoreremo, che è il risultato di un lavoro di selezione e implementazione dei dati iniziali, è composto da 588 osservazioni, le quali sono state estratte casualmente dal campione totale, e riguardano 12 variabili di interesse. I dati sono riconducibili a persone fisiche e aziende che hanno eseguito una qualsiasi operazione sul proprio conto corrente dal 1° Ottobre 2013 al 31 Ottobre 2013. Le variabili che andremo ad analizzare, le quali verranno descritte più accuratamente nel prossimo paragrafo, sono: *tipo*, *inbank*, *bancomat*, *credito*, *ricarica*, *sex*, *età_fasce*, *socio*, *gruppo*, *sotto*, *prof*, *uso_ib*. Per le analisi descrittive e le stime econometriche delle variabili sono stati utilizzati i programmi *Excel* e *Stata*[®].

3.1 Analisi descrittiva delle variabili

La prima variabile che andremo ad analizzare è la variabile *tipo*; è una variabile discreta che prende valore 0 se il cliente è un privato, 1 se è un'impresa e 2 se è un ente diverso o un'associazione senza scopo di lucro.

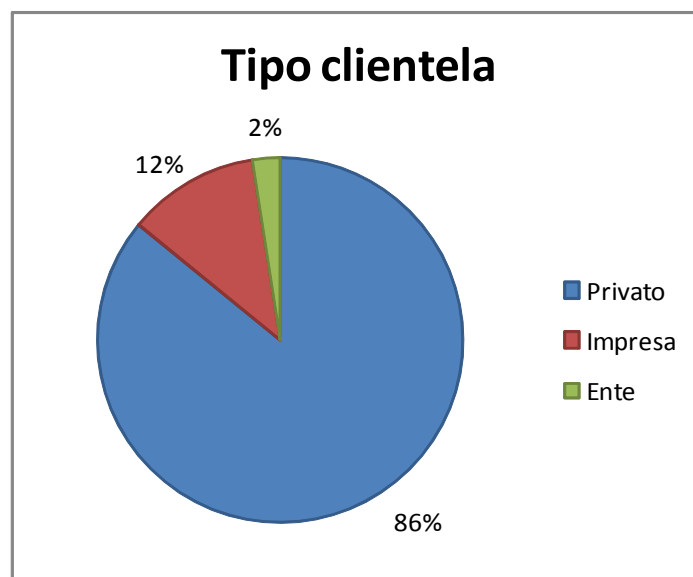


Grafico 1: suddivisione della clientela per tipologia.

Come si evince dal *grafico 1*, la maggioranza della clientela appartiene al gruppo privato.

La variabile *inbank* è una variabile binaria che assume valore 1 se il cliente ha l'accesso all'home banking e 0 in caso contrario.

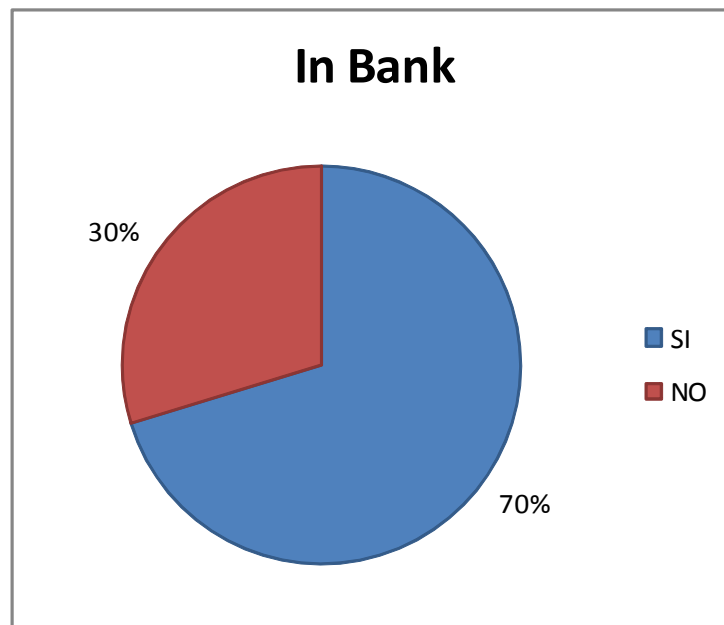


Grafico 2: diffusione dei clienti con accesso all'home banking.

Come si nota dal *grafico 2*, buona parte dei clienti possiede l'accesso a questa tipologia di pagamento.

Altre tre variabili, che descriveremo insieme data la somiglianza della tipologia di dati analizzati, sono *bancomat*, *credito* e *ricarica*: sono variabili binarie che assumono valore 1 in caso di possesso rispettivamente di bancomat, carta di credito o "carta ricarica" e 0 in caso contrario. Le tre variabili si possono osservare nel *grafico 3*, *grafico 4* e *grafico 5*.

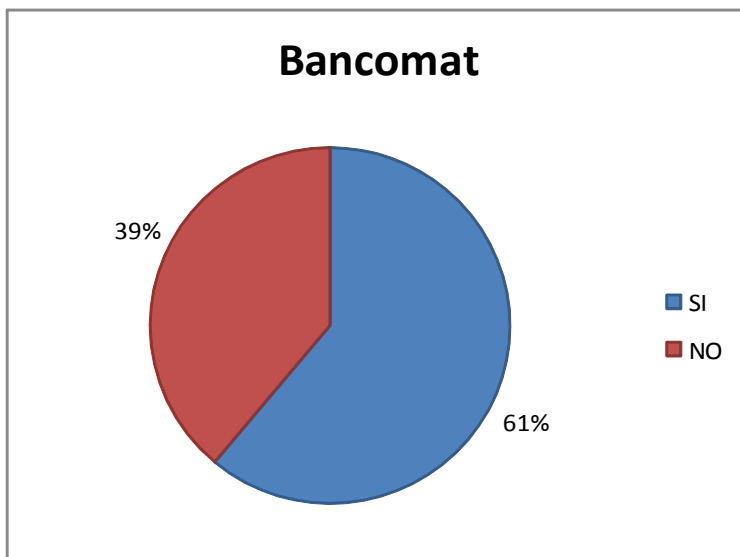


Grafico 3: possessori bancomat.

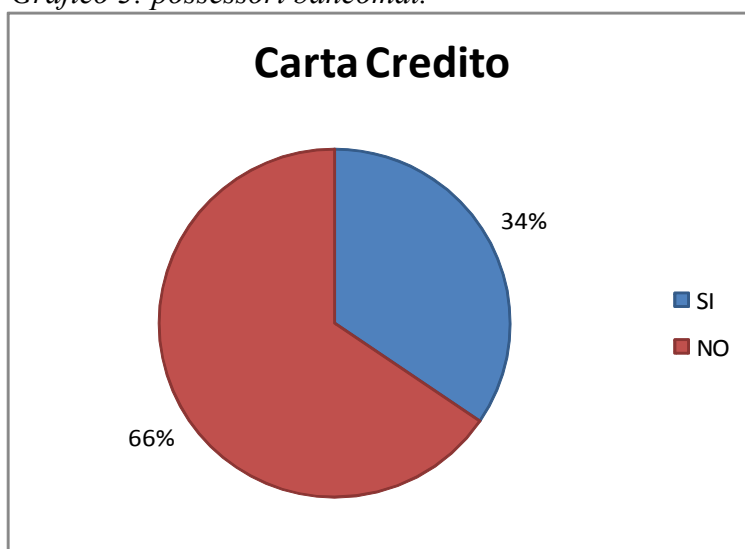


Grafico 4: possessori carta di credito.

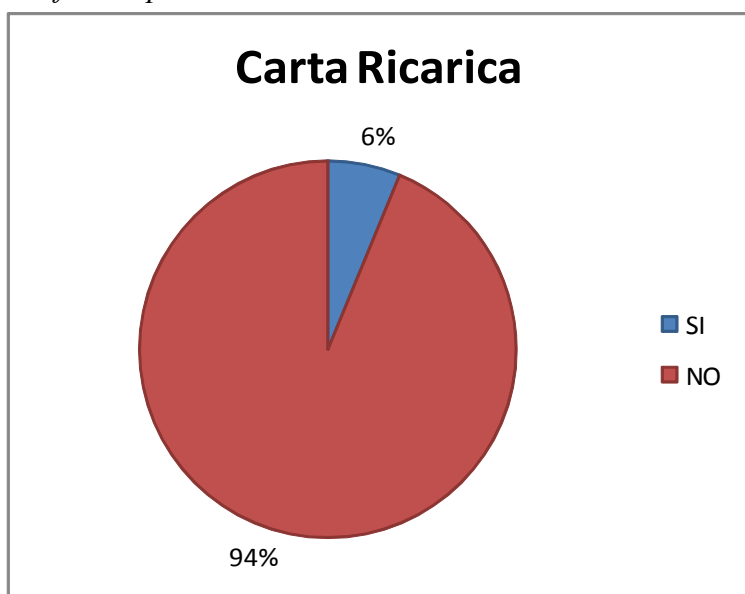


Grafico 5: possessori "carta ricarica".

Tra le tre variabili potrebbe stupire in particolare la bassa diffusione di bancomat e "carta ricarica". Mentre per quest'ultima si tratta effettivamente di uno scarso possesso del mezzo di pagamento in oggetto, per il bancomat evidenziamo il fatto che le imprese e gli enti diversi o associazioni, generalmente non possiedono questa tipologia di mezzo di pagamento, come si può notare dalla tabella a doppia entrata tra le variabili *tipo* e *bancomat*.

Possesso bancomat	Tipologia cliente			Total
	0	1	2	
0	29.04	99.00	100.00	38.93
1	70.96	1.00	0.00	61.07
Total	100.00	100.00	100.00	100.00

La percentuale di possessori di bancomat tra privati, escludendo le altre due tipologie di clientela, salirebbe così a circa il 71%.

La variabile *sex* è una variabile binaria che identifica con 0 il fatto di essere femmina e con 1 il fatto di essere maschio, come evidenziato dal *grafico 6*.

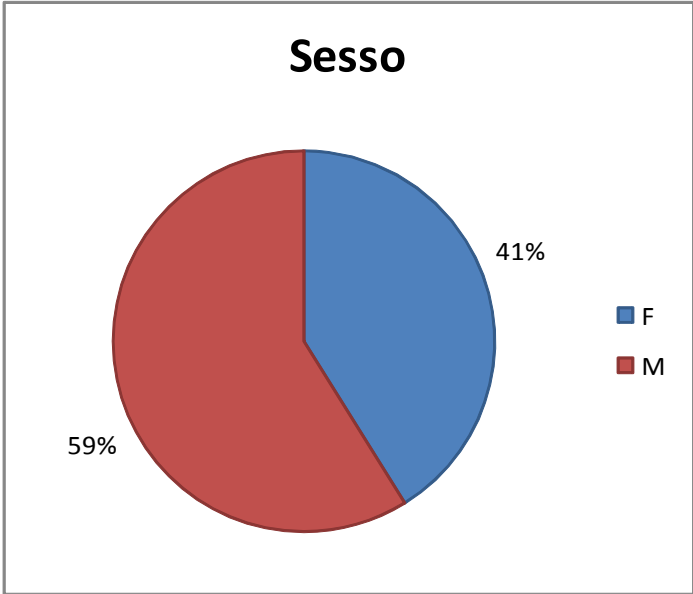


Grafico 6: genere della clientela.

La variabile *età_fasce*, descritta dal *grafico 7*, che inizialmente era rappresentata dall'età in anni della clientela, è stata trasformata in una variabile discreta che assume 5 valori in base alla fascia d'età dentro la quale rientra il cliente in oggetto. Le fasce d'età sono "Fino a 29", "da 30 a 44", "da 45 a 64", da "65 a 74" e "oltre 75".

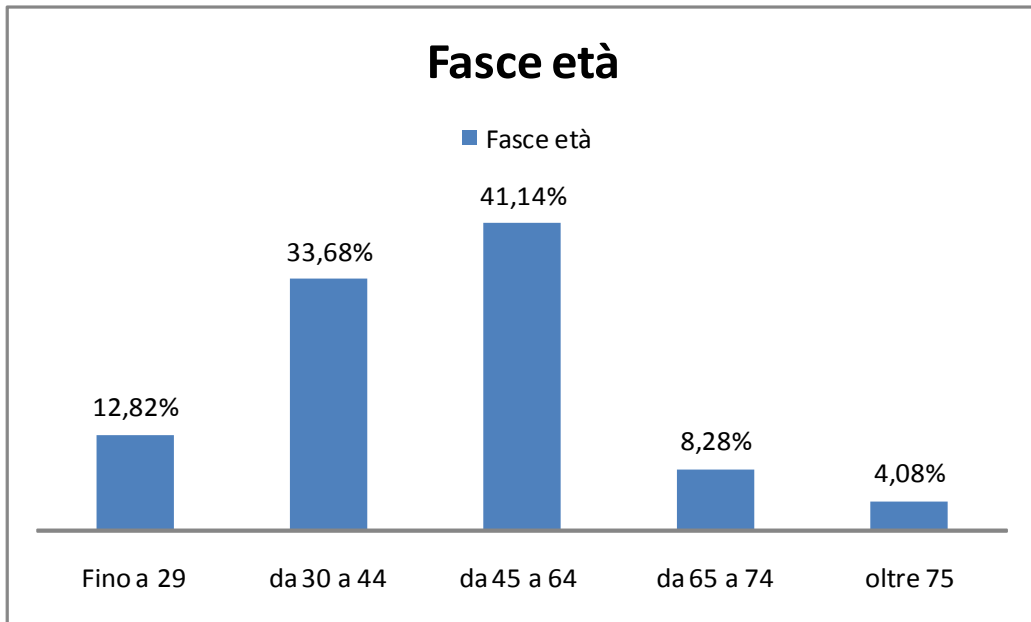


Grafico 7: distribuzione dei clienti per fasce d'età.

Per una migliore analisi della variabile possiamo esaminare l'età con i singoli valori appartenenti ad ogni cliente.

Percentiles	Smallest		
1%	20	18	
5%	23	19	
10%	27	19	Obs 858
25%	36	19	Sum of Wgt. 858
50%	45		Mean 46.67133
		Largest	Std. Dev. 14.64505
75%	56	85	
90%	67	86	Variance 214.4776
95%	74	87	Skewness .3204094
99%	82	89	Kurtosis 2.631836

Dalla tabella di frequenza e dal *grafico 8* si può notare che l'età media dei correntisti è 46,7 anni e la mediana si posiziona sui 45 anni.

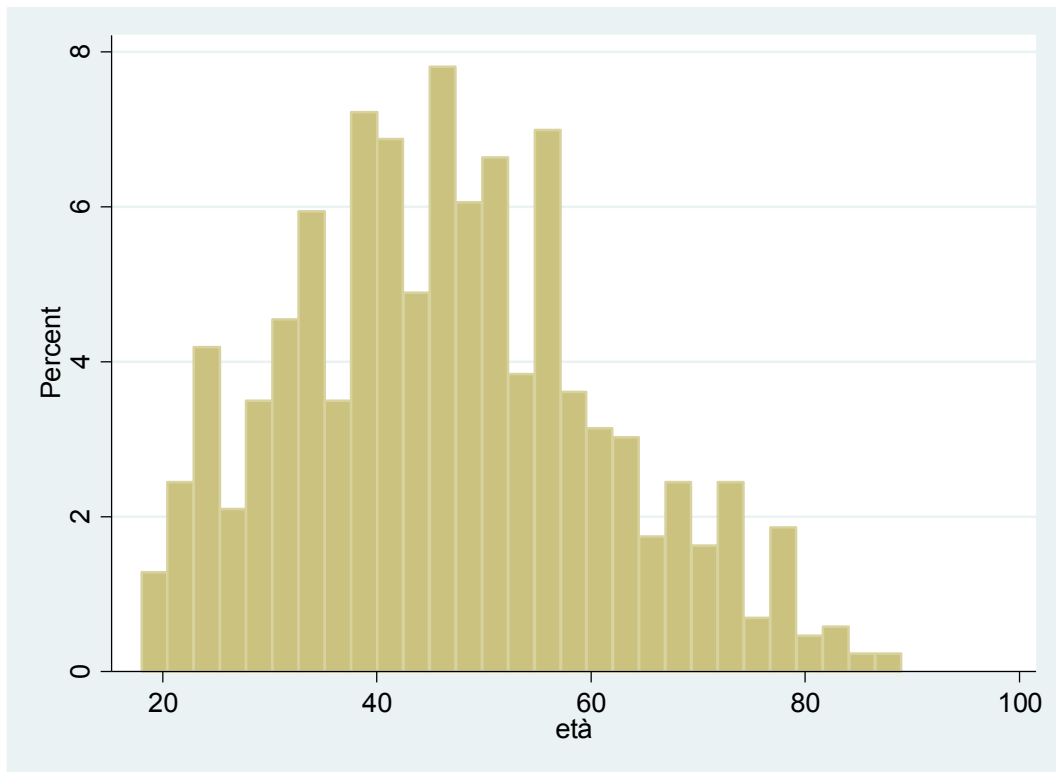


Grafico 8: distribuzione di frequenza età clienti.

La variabile *socio* assume valore 1 se il cliente ha lo status di socio, ovvero gode di particolari agevolazioni in cambio del pagamento di una quota, mentre assume il valore 0 se è un cliente normale (*figura 9*).

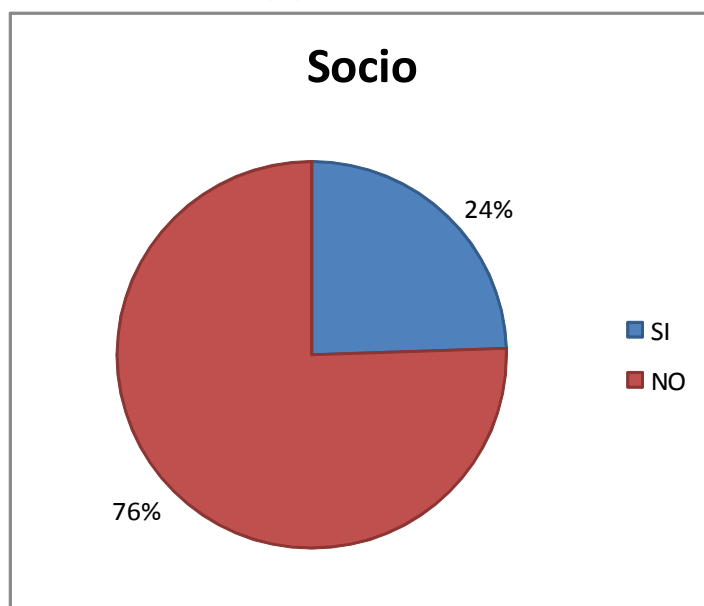


Grafico 9: clienti soci della banca.

La variabile *gruppo* descrive il gruppo economico al quale appartiene il singolo cliente; non essendo presente questo attributo per ogni osservazione, i valori della variabile sono stati così distribuiti: 0 se non è presente alcun valore, 1 se il cliente appartiene al gruppo economico “Produzione/distribuzione generi alimentari”, 2 se appartiene a “Produzione/lavorazione metalli, macchinari elettromeccanici”, 3 se appartiene a “Servizi alla persona vari”, 4 se è del gruppo “Abitazioni/fabbricati e relativi materiali” e 5 se appartiene al gruppo “Articoli non alimentari”.

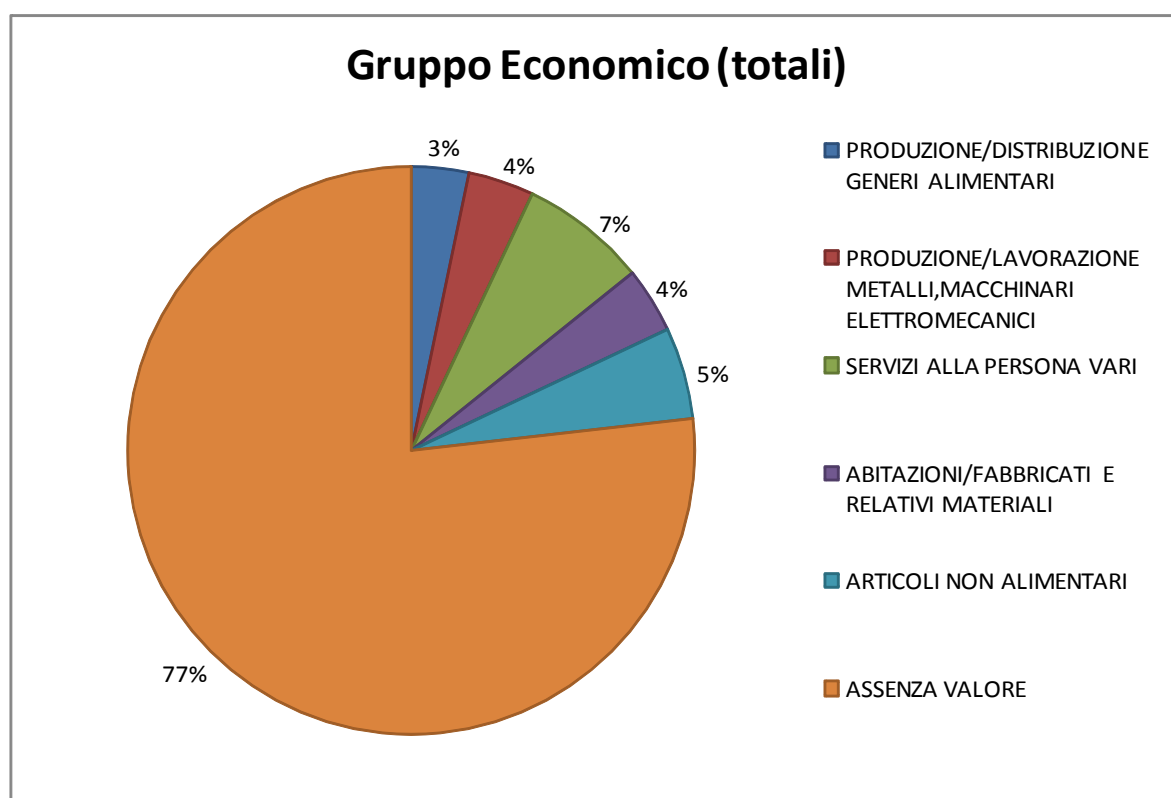


Grafico 10: distribuzione gruppo economico per ogni cliente.

Come si può notare dal *grafico 10* a più del 75% dei clienti risulta mancare il valore in oggetto, questo perché l'attributo viene dato principalmente ad imprese o enti/associazioni. Togliendo la modalità “Assenza valore” possiamo vedere come si distribuisce la caratteristica tra i possessori di quest'ultima.

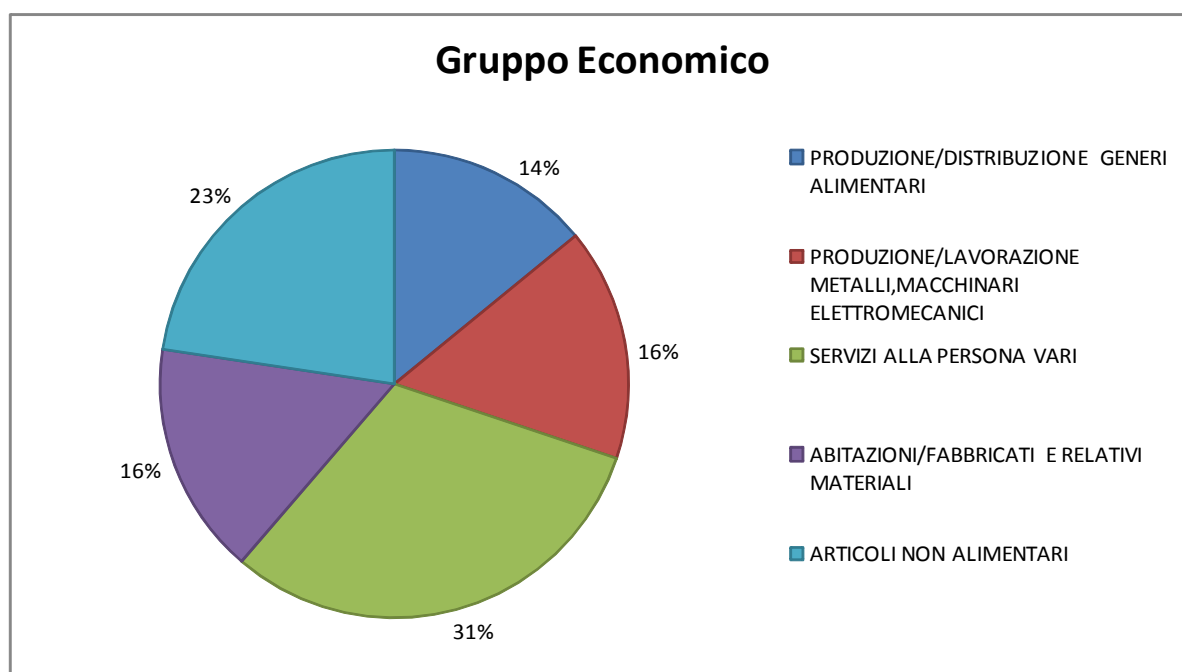


Grafico 11: distribuzione per gruppi economici tra clienti con valore presente.

Come si può vedere dal *grafico 11*, c'è una prevalenza del gruppo "Servizi alla persona", seguito da "Articoli alimentari". I restanti gruppi seguono distaccati attestandosi su di un 15% circa.

La variabile *sotto* è molto simile alla variabile *tipo* e come vedremo durante la stima dei modelli le due variabili entreranno in conflitto creando problemi di collinearità. La variabile descrive il sottogruppo economico di appartenenza: prende valore 1 nel caso di "Imprese produttive", valore 2 nel caso di "Istituti senza scopo di lucro" e valore 3 nel caso di "Famiglie consumatrici".

Dal *grafico 12* si nota come la modalità "Famiglie consumatrici" sia la più diffusa con il 74%, seguita da "Imprese produttive" con il 23% e infine "Istituzioni senza scopo di lucro" che comprendono una minima parte del totale con il 3%.

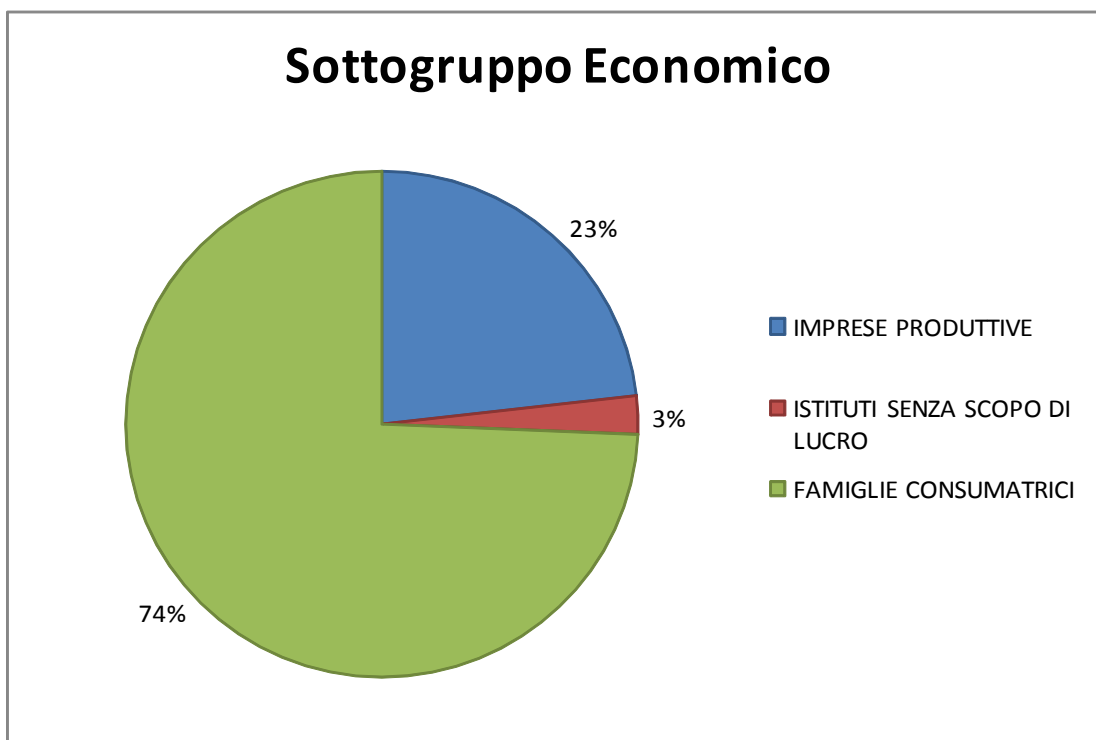


Grafico 12: distribuzione per sottogruppi economici.

La variabile *prof* indica la professione svolta dal cliente in oggetto; si divide in “Non occupato/pensionato/studente” che prende valore 1, “Lavoratore autonomo/indipendente” che assume valore 2 e infine “Lavoratore dipendente” che prende valore 3.

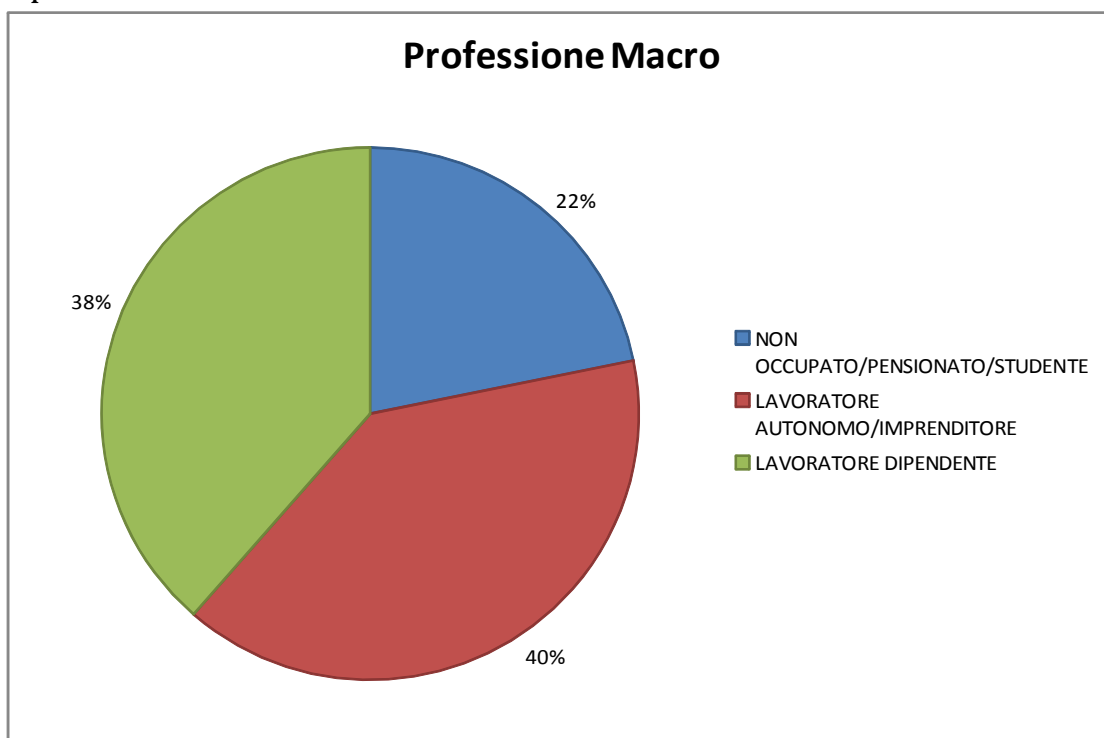


Grafico 13: distribuzione delle professioni macro.

Dal *grafico 13* si può vedere che le categorie “Lavoratore dipendente” e “Lavoratore autonomo/imprenditore” sono presenti in percentuali molto simili e rappresentano quasi l'80% del totale, con il rimanente completato dalla categoria “Non occupato/pensionato/studente”.

L'ultima variabile presa in considerazione è *uso_ib*, la quale riguarda solo coloro i quali hanno accesso all'home banking; assume valore 1 in caso di utilizzo del servizio nel periodo considerato e 0 in caso di non utilizzo.

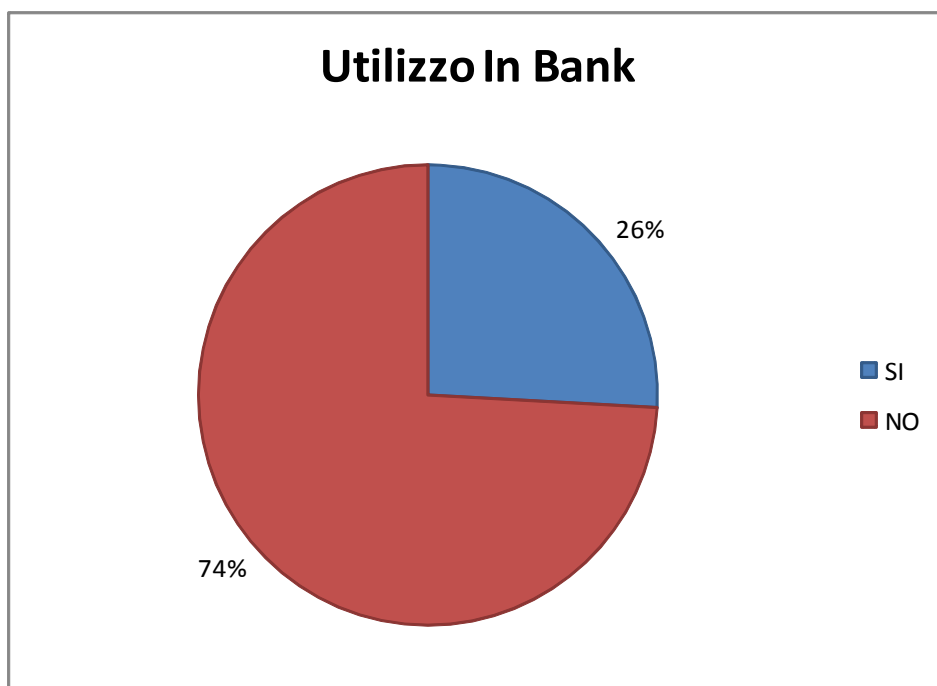


Grafico 14: utilizzo home banking durante il mese di Ottobre 2013.

Dal *grafico 14* si nota come solo il 26% dei possessori di home banking abbia avuto occasione di utilizzarlo nel periodo considerato.

Sono state create inoltre altre due variabili, *almeno* e *ib_and*, per poter ottenere i risultati richiesti. La prima è una variabile binaria che assume valore 1 se il cliente in oggetto possiede uno tra bancomat, carta di credito e "carta ricarica" e 0 se non ne possiede nessuna delle tre. La seconda è anch'essa una variabile binaria che assume valore 1 se viene soddisfatta la condizione della variabile *almeno* e contemporaneamente il cliente ha accesso all'home banking e 0 altrimenti.

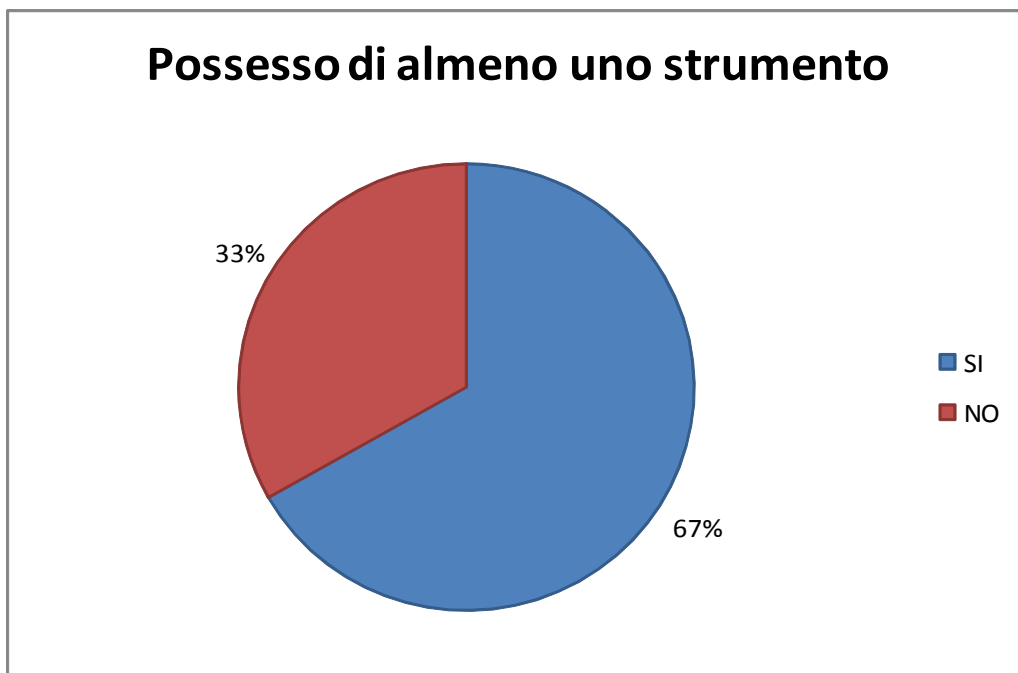


Grafico 15: percentuale di possesso di uno tra bancomat, carta di credito e "carta ricarica".

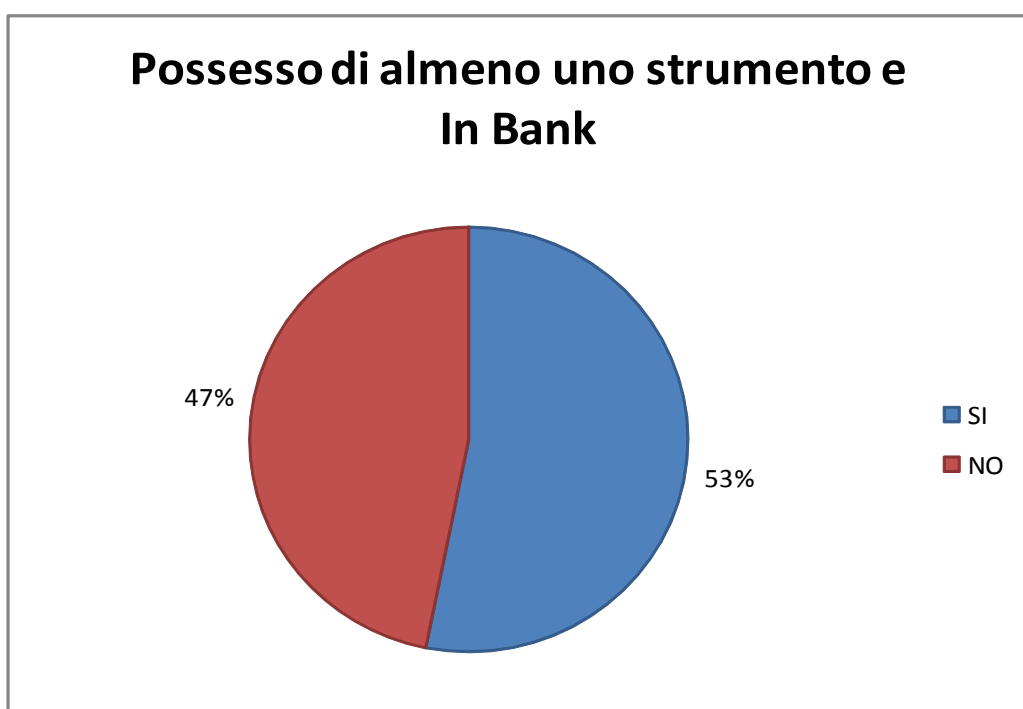


Grafico 16: percentuale di clienti con accesso all'home banking e possesso di uno tra bancomat, carta di credito e "carta ricarica".

Come si può vedere dal *grafico 15* e *grafico 16* circa 2 clienti su 3 possiedono uno tra bancomat, carta di credito e "carta ricarica" e più della metà hanno accesso all'home banking e una delle tre carte sopracitate.

Analizzeremo ora brevemente, attraverso l'utilizzo di alcuni grafici a barre, come si distribuiscono le variabili *inbank* e *almeno* tra le variabili *età_fasce* e *sex*. Incrociamo le variabili *inbank* e *almeno* con *età_fasce*.

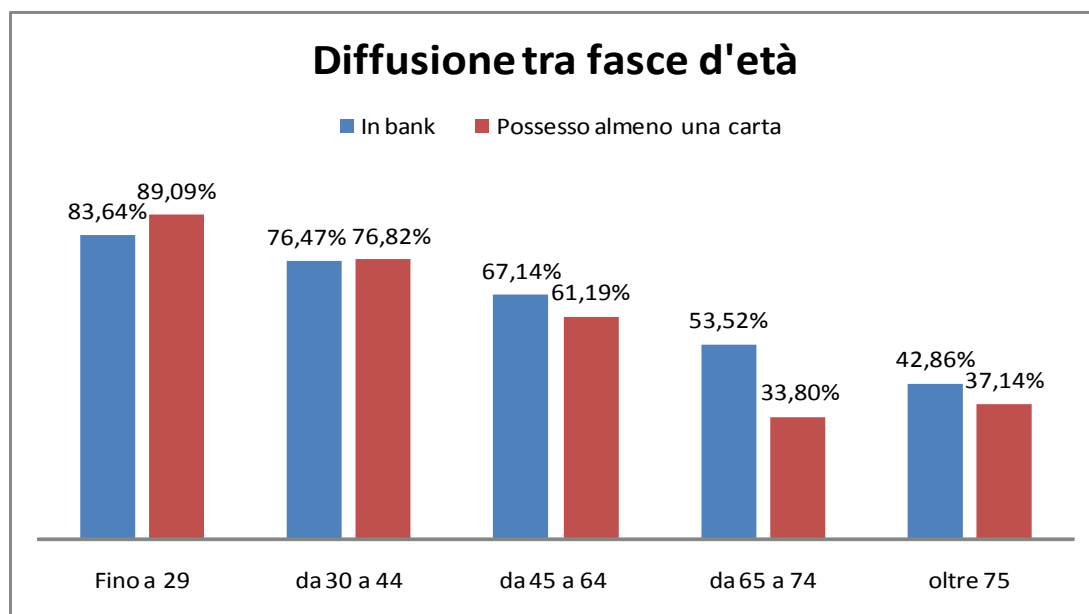


Grafico 17: diffusione accesso home banking e possesso di uno tra bancomat, carta di credito e "carta ricarica" tra le fasce d'età.

Possiamo vedere dal *grafico 17* come il possesso di home banking parta da un 84% nella popolazione "Fino a 29" per arrivare gradatamente ad un 43% nella classe "oltre 75". Questo dimostra come questo strumento sia più diffuso tra i giovani rispetto alle categorie con un'età maggiore.

Sempre dal *grafico 17* notiamo come quasi 9 persone su 10 nella fascia "Fino a 29" posseda almeno una delle tre carte, per poi scendere del 15% circa fino alla fascia "da 45 a 64". Da quest'ultima fascia alla seguente, ovvero "da 65 a 74" c'è una riduzione del 30% circa dei possessori per poi rialzarsi di qualche punto percentuale nell'ultima classe "oltre 75". Anche qui, come visto in precedenza con la variabile *inbank*, all'aumentare dell'età c'è una diminuzione del possesso di almeno una delle tre carte. A differenza della variabile precedente però la variazione percentuale di possesso tra le varie classi è maggiore, in particolare tra la terza e la quarta, mentre c'è addirittura un leggero aumento tra la penultima fascia e l'ultima.

Ora vedremo come si distribuiscono le variabili *inbank* e *almeno* al modificarsi della variabile *sex*, cioè il genere del cliente.

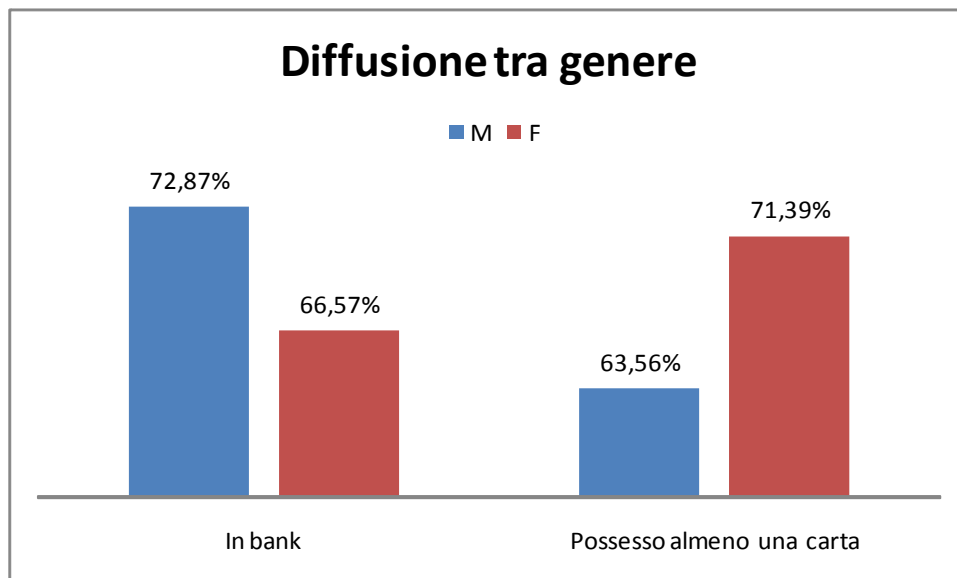


Grafico 18: diffusione accesso home banking e possesso di una tra bancomat, carta di credito e "carta ricaricabile" tra il genere.

Nel grafico 18 notiamo come 2 donne su 3 abbiano accesso all'home banking contro circa un 73% degli uomini. Per quanto riguarda il possesso di almeno una carta vediamo invece come la situazione sia invertita, con una diffusione del 71,39% tra le donne e del 63,56% tra gli uomini.

3.2 Approccio econometrico ai dati

Passiamo ora a testare con *Stata*[®] le ipotesi fino a qui formulate. Le variabili dipendenti che verranno utilizzate nelle varie analisi sono *inbank*, *uso_ib*, *almeno* e *ib_and*. Per quanto riguarda le variabili indipendenti, si è dovuto utilizzare qualche accorgimento, trasformando le variabili non dicotomiche in variabili dummy, le quali possono assumere 0 o 1. In questo modo si è potuto avere una migliore interpretazione dei risultati ottenuti. Si è dovuto inoltre omettere una variabile dummy, per ogni gruppo di variabili riguardanti lo stesso carattere, per evitare la cosiddetta trappola delle dummy, in quanto con la presenza della costante si verrebbe a creare collinearità perfetta. Per quest'ultimo motivo si è

dovuto scegliere inoltre tra la variabile *tipo* e la variabile *sotto*, in quanto la loro interazione causava proprio questo tipo di problematica. Le variabili indipendenti che sono state utilizzate sono quindi *sex*, *socio*, *tipo1*, *tipo2*, *gruppo2*, *gruppo3*, *gruppo4*, *gruppo5*, *gruppo6*, *prof1*, *prof2*, *età_fasce2*, *età_fasce3*, *età_fasce4*, *età_fasce5* oltre ad *inbank* e *uso_ib* per alcuni tipi di modello. Date queste variabili indipendenti e il significato che acquisiscono al variare degli 1 o 0, l'individuo di riferimento è una femmina, non socio, fa riferimento ad un ente o associazione, non ha un valore specifico nella categoria "Gruppo economico", è un lavoratore dipendente con un'età fino a 29 anni.

Spieghiamo adesso come verranno interpretati gli output dei modelli analizzati; per quanto riguarda i coefficienti risultanti dal modello probit, ciò che ci interesserà osservare sarà solamente il segno in quanto, come spiegato nei paragrafi precedenti, il valore risultante non è direttamente interpretabile. Verranno poi calcolati gli effetti marginali per il valore medio di ogni variabile, i quali daranno un peso, seppur riguardante il valore medio, all'apporto della singola variabile. Per quanto concerne la significatività dei coefficienti e degli effetti marginali bisognerà osservare il valore della probabilità ($P > |z|$): se maggiore di 0,05 la variabile non sarà ritenuta significativa e quindi, per il modello, l'apporto di quest'ultima sarà praticamente pari a zero secondo la seguente ipotesi:

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

Vedremo infine una statistica che descriverà la percentuale di valori predetti correttamente rispetto ai valori della variabile dipendente analizzata.

Iniziamo ad analizzare come interagiscono le variabili indipendenti descritte in precedenza con la variabile *inbank*, per cercare di capire quali influenzino di più l'accesso all'home banking.

Testando il modello in *Stata*[®] si ottiene il seguente risultato:

```

Probit regression                               Number of obs   =           858
                                                LR chi2(15)     =           98.21
                                                Prob > chi2     =           0.0000
Log likelihood = -472.96748                    Pseudo R2      =           0.0941

```

inbank	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
maschio	.1665485	.0974326	1.71	0.087	-.0244158	.3575129
socio	.3087006	.1212412	2.55	0.011	.0710723	.5463289
privato	.2863912	.2904839	0.99	0.324	-.2829468	.8557293
impresa	.9494792	.361617	2.63	0.009	.2407229	1.658235
gen. alimen.	-.0990636	.2768977	-0.36	0.721	-.6417732	.4436459
lav. metalli	.5543743	.3570352	1.55	0.120	-.1454018	1.25415
serv. personal	.3718356	.2239708	1.66	0.097	-.067139	.8108103
abitazioni	.0797846	.2784642	0.29	0.774	-.4659952	.6255644
non alimen.	.1066221	.2597546	0.41	0.681	-.4024877	.6157318
non occupato	-.4901328	.1453895	-3.37	0.001	-.7750909	-.2051746
lav. autonomo	-.323455	.1265544	-2.56	0.011	-.5714971	-.0754129
18-29 anni	.6874313	.1686733	4.08	0.000	.3568376	1.018025
30-44 anni	.2344443	.1123338	2.09	0.037	.0142741	.4546146
65-74 anni	-.2576485	.1865673	-1.38	0.167	-.6233136	.1080167
oltre 75 anni	-.4518848	.2536118	-1.78	0.075	-.9489548	.0451853
costante	.1144962	.3015402	0.38	0.704	-.4765119	.7055042

Il modello non risulta spiegare in maniera adeguata il comportamento delle variabili indipendenti; questo si può verificare dal basso Pseudo R2 ottenuto. Le variabili evidenziate risultano significative al 5%; da queste si può capire che essere giovani aumenta la probabilità di avere accesso all'home banking; lo stesso si può dire per l'essere socio ed essere un'impresa, mentre essere non occupato o lavoratore indipendente diminuisce questa probabilità.

Proviamo ora ad analizzare il comportamento rispetto all'utilizzo dell'home banking che abbiamo testato in precedenza, tramite l'utilizzo della variabile *uso_ib*. Il modello che si ottiene è:

```

Probit regression                               Number of obs   =       858
                                                LR chi2(15)    =       137.02
                                                Prob > chi2    =       0.0000
Log likelihood = -338.30076                    Pseudo R2      =       0.1684

```

uso_ib	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
maschio	.1172424	.1153928	1.02	0.310	-.1089233	.3434081
socio	.1015583	.1284032	0.79	0.429	-.1501074	.3532241
privato	-1.161457	.2935588	-3.96	0.000	-1.736821	-.5860918
impresa	-.3984899	.3437244	-1.16	0.246	-1.072177	.2751976
gen. alimen.	.2700267	.2860639	0.94	0.345	-.2906483	.8307016
lav. Metalli	.3751403	.2860208	1.31	0.190	-.1854502	.9357308
serv. persona	.6283585	.2048354	3.07	0.002	.2268884	1.029829
abitazioni	.690007	.2648817	2.60	0.009	.1708484	1.209166
non alimen.	.407526	.2467451	1.65	0.099	-.0760855	.8911375
non occupato	-.5007558	.2097334	-2.39	0.017	-.9118257	-.0896859
lav. autonomo	.1386001	.1413569	0.98	0.327	-.1384544	.4156546
18-29 anni	.0133014	.2038153	0.07	0.948	-.3861692	.4127719
30-44 anni	.1366098	.1259074	1.09	0.278	-.1101641	.3833837
60-74 anni	.3612929	.2211986	1.63	0.102	-.0722485	.7948342
oltre 75 anni	-.1240524	.4017244	-0.31	0.757	-.9114177	.663313
costante	-.2341786	.3070766	-0.76	0.446	-.8360376	.3676804

Il modello spiegato sembra essere migliore di quello precedente anche se ci sono solo quattro variabili con significatività al 5% e una al 10%. Proviamo quindi ad utilizzare lo stesso modello, questa volta però restringendo la popolazione d'interesse ai soli clienti con accesso all'home banking.

Il modello ottenuto è il seguente:

```

Probit regression                               Number of obs   =           603
                                                LR chi2(15)     =           123.00
                                                Prob > chi2     =           0.0000
Log likelihood = -281.10965                    Pseudo R2      =           0.1795
  
```

uso_ib	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
maschio	.0035154	.1313686	0.03	0.979	-.2539623	.2609931
socio	-.0010375	.141783	-0.01	0.994	-.2789271	.2768521
privato	-1.843851	.4563717	-4.04	0.000	-2.738323	-.9493789
impresa	-1.16754	.4933266	-2.37	0.018	-2.134443	-.2006379
gen. alimen.	.3283492	.3304685	0.99	0.320	-.3193571	.9760555
lav. metalli	.2325107	.3051703	0.76	0.446	-.3656122	.8306336
serv. persona	.5389424	.2263161	2.38	0.017	.0953709	.9825139
abitazioni	.7520681	.3001762	2.51	0.012	.1637336	1.340403
non alimen.	.4255438	.2689104	1.58	0.114	-.1015108	.9525985
non occupato	-.5104293	.2433719	-2.10	0.036	-.9874294	-.0334292
lav. autonomo	.3214146	.1561153	2.06	0.040	.0154342	.627395
18-29 anni	-.1003378	.216827	-0.46	0.644	-.525311	.3246354
30-44 anni	.0235337	.1381565	0.17	0.865	-.247248	.2943154
60-74 anni	.3868775	.273659	1.41	0.157	-.1494844	.9232393
oltre 75 anni	.0791865	.474885	0.17	0.868	-.851571	1.009944
costante	.7572473	.4744481	1.60	0.110	-.1726539	1.687148

Il modello, analizzando la statistica Pseudo R2, risulta leggermente migliore rispetto al precedente, contando 6 variabili significative al 5%. Proviamo ora a calcolare gli effetti marginali condizionati alla media delle variabili esplicative.

Si ottiene dunque:

```

Conditional marginal effects          Number of obs   =          603
Model VCE      : OIM
Expression     : Pr(uso_ib), predict()
-----
|                               Delta-method
|                               dy/dx   Std. Err.    z    P>|z|    [95% Conf. Interval]
-----+-----
maschio |    .0010464   .0391043    0.03   0.979   - .0755966   .0776894
socio   |   -.0003088   .0422042   -0.01   0.994   - .0830276   .0824099
privato |   -.5488521   .1374514   -3.99   0.000   - .8182519  -.2794523
impresa |   -.3475372   .147335    -2.36   0.018   - .6363085  -.058766
gen. alimen. | .0977385   .0983794    0.99   0.320   - .0950816   .2905586
lav. metalli | .0692106   .0908089    0.76   0.446   - .1087715   .2471927
serv. persona | .1604249   .0674008    2.38   0.017   .0283219   .292528
abitazioni | .2238653   .0897008    2.50   0.013   .048055    .3996756
non alimen. |    .12667    .080033    1.58   0.113   - .0301918   .2835318
non occupato | -.1519376   .0717891   -2.12   0.034   - .2926416  -.0112335
lav. autonomo | .0956743   .0464038    2.06   0.039   .0047244   .1866241
18-29 anni | -.0298672   .0644958   -0.46   0.643   - .1562766   .0965422
30-44 anni | .0070052   .0411205    0.17   0.865   - .0735896    .0876
60-74 anni | .1151604   .0813522    1.42   0.157   - .0442871   .2746078
oltre 75 anni | .0235712   .1413496    0.17   0.868   - .253469    .3006113

```

Il fatto di essere privato o impresa riduce la probabilità di utilizzare l'home banking rispettivamente del 54,89% e 34,75%. Appartenere ai gruppi "Servizi alla persona vari" e "Abitazioni/prefabbricati e relativi materiali" aumenta invece del 16,04% e 22,39% rispettivamente. Infine essere non occupato diminuisce la probabilità del 15,19% mentre essere un lavoratore indipendente la fa aumentare del 9,57%.

Vediamo ora l'adattamento del modello attraverso la percentuale di risultati predetti correttamente:

```

Probit model for uso_ib
----- True -----
Classified |          D          ~D |      Total
-----+-----+-----+-----
      +   |          63          34 |          97
      -   |          91         415 |         506
-----+-----+-----+-----
    Total |         154         449 |         603

```

Classified + if predicted $\Pr(D) \geq .5$
 True D defined as $\text{uso_ib} \neq 0$

```

-----
Sensitivity                Pr( + | D)    40.91%
Specificity                Pr( - | ~D)   92.43%
Positive predictive value  Pr( D | +)   64.95%
Negative predictive value  Pr(~D | -)   82.02%
-----
False + rate for true ~D   Pr( + | ~D)   7.57%
False - rate for true D    Pr( - | D)   59.09%
False + rate for classified + Pr(~D | +)   35.05%
False - rate for classified - Pr( D | -)   17.98%
-----
Correctly classified                79.27%

```

Con quasi l'80% di risultati predetti correttamente il modello stimato si dimostra soddisfacente dal punto di vista probabilistico.

Siamo ora interessati ad analizzare la variabile *almeno*, che indica il possesso o meno di uno tra bancomat, carta di credito o "carta ricarica".

Utilizzando sempre il modello probit otteniamo:

```

Probit regression                               Number of obs   =           858
                                                LR chi2(15)     =           326.24
                                                Prob > chi2     =           0.0000
Log likelihood = -382.31443                    Pseudo R2      =           0.2991

```

	almeno	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
	maschio	-.0118335	.1090985	-0.11	0.914	-.2256626 .2019955
	socio	.3608049	.1308831	2.76	0.006	.1042788 .617331
	privato	6.302391	151.1344	0.04	0.967	-289.9156 302.5204
	impresa	4.977475	151.1346	0.03	0.974	-291.2408 301.1958
	gen. alimen.	-.671241	.3005342	-2.23	0.026	-1.260277 -.0822048
	lav. metalli	-.3449735	.3125691	-1.10	0.270	-.9575977 .2676507
	serv. personal	.0074064	.2161584	0.03	0.973	-.4162563 .431069
	abitazioni	-.1634998	.2723067	-0.60	0.548	-.6972112 .3702116
	non alimen.	-.7155383	.2602467	-2.75	0.006	-1.225613 -.2054641
	non occupato	-.752331	.1669437	-4.51	0.000	-1.079535 -.4251274
	lav. autonomo	-.7895338	.1413253	-5.59	0.000	-1.066526 -.5125412
	18-29 anni	.7698851	.1999712	3.85	0.000	.3779487 1.161821
	30-44 anni	.190903	.1223036	1.56	0.119	-.0488077 .4306137
	60-74 anni	-.6174726	.2016031	-3.06	0.002	-1.012607 -.2223377
	oltre 75 anni	-.7221999	.2598525	-2.78	0.005	-1.231501 -.2128984
	costante	-5.173057	151.1344	-0.03	0.973	-301.3911 291.045

Come si può notare le variabili *tipo1*, *tipo2* e la costante mostrano standard error molto elevati e di conseguenza p-value molto bassi. Ciò potrebbe indicare la presenza di quasi-collinearità tra alcune variabili indipendenti. Per ovviare a questo problema si è deciso di eliminare la variabile *tipo1* e lasciare solamente la variabile *tipo2*, ovvero il fatto di essere un'impresa.

Testiamo quindi il modello con le nuove variabili:

```

Probit regression                               Number of obs   =           858
                                                LR chi2(14)     =           275.57
                                                Prob > chi2     =           0.0000
Log likelihood = -407.64717                    Pseudo R2      =           0.2526
  
```

almeno	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
maschio	-.0425584	.1052946	-0.40	0.686	-.2489321	.1638153
socio	.3828278	.1281719	2.99	0.003	.1316155	.6340401
impresa	-1.31887	.2103648	-6.27	0.000	-1.731178	-.9065629
gen. alimen.	-.5747399	.2997639	-1.92	0.055	-1.162266	.0127864
lav. metalli	-.2675318	.3119794	-0.86	0.391	-.8790001	.3439365
serv. personal	.088623	.2147979	0.41	0.680	-.3323732	.5096192
abitazioni	-.0755034	.2708314	-0.28	0.780	-.6063233	.4553164
non alimen.	-.6357003	.2591949	-2.45	0.014	-1.143713	-.1276877
non occupato	-.7609516	.1604896	-4.74	0.000	-1.075506	-.4463977
lav. autonomo	-.7872643	.1356766	-5.80	0.000	-1.053186	-.521343
18-29 anni	.8810999	.1969654	4.47	0.000	.4950548	1.267145
30-44 anni	.2445769	.1187987	2.06	0.040	.0117357	.4774181
60-74 anni	-.5657832	.195721	-2.89	0.004	-.9493893	-.1821772
oltre 75 anni	-.6051917	.2567273	-2.36	0.018	-1.108368	-.1020154
costante	1.02331	.1269236	8.06	0.000	.7745439	1.272075

Come si può vedere dai risultati il modello sembra avere un buon adattamento, con un Pseudo R2 pari a 0.2526 e buona parte delle variabili indipendenti significative al 5%.

Proviamo ora ad aggiungere tra le variabili di interesse la variabile *inbank*, ovvero l'accesso all'home banking, e vedere se l'inserimento di tale variabile risulta significativo, con conseguente miglioramento d'adattamento del modello:

```

Probit regression                               Number of obs   =           858
                                                LR chi2(15)    =           355.98
                                                Prob > chi2    =           0.0000
Log likelihood = -367.4416                    Pseudo R2      =           0.3263

```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
almeno						
maschio	-.1280442	.1115048	-1.15	0.251	-.3465897	.0905013
socio	.3084986	.1339072	2.30	0.021	.0460453	.5709518
impresa	-1.528512	.2153255	-7.10	0.000	-1.950542	-1.106482
gen. alimen.	-.6439641	.3128252	-2.06	0.040	-1.25709	-.030838
lav. metalli	-.4330519	.3160646	-1.37	0.171	-1.052527	.1864233
serv. persona	-.0417245	.2216978	-0.19	0.851	-.4762443	.3927952
abitazioni	-.1349799	.277738	-0.49	0.627	-.6793364	.4093766
non alimen.	-.7614754	.2664288	-2.86	0.004	-1.283666	-.2392845
non occupato	-.6571107	.1718047	-3.82	0.000	-.9938417	-.3203797
lav. autonomo	-.7578407	.1428778	-5.30	0.000	-1.037876	-.4778054
18-29 anni	.7235425	.2147729	3.37	0.001	.3025954	1.14449
30-44 anni	.1646058	.1253352	1.31	0.189	-.0810467	.4102583
60-74 anni	-.4720806	.1996451	-2.36	0.018	-.8633778	-.0807834
oltre 75 anni	-.4865134	.2745289	-1.77	0.076	-1.02458	.0515533
inbank	1.011601	.1147406	8.82	0.000	.7867133	1.236488
costante	.4964227	.1460029	3.40	0.001	.2102623	.7825831

I risultati confermano la significatività della variabile inserita, con un livello dell'1%; si ha inoltre un netto miglioramento riguardante l'adattamento del modello e un aumento delle variabili indipendenti significative al 5%.

Passiamo ora al calcolo degli effetti marginali, condizionati alla media delle variabili esplicative, in modo da poter quantificare l'influenza delle singole variabili analizzate:

```
Conditional marginal effects          Number of obs   =          858
Model VCE      : OIM
Expression    : Pr(almeno), predict()
```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
maschio	-.0429225	.0373197	-1.15	0.250	-.1160678	.0302227
socio	.1034138	.044941	2.30	0.021	.0153311	.1914966
impresa	-.5123826	.0748879	-6.84	0.000	-.6591602	-.3656051
gen. alimen.	-.2158675	.1052152	-2.05	0.040	-.4220854	-.0096495
lav. metalli	-.1451662	.1059575	-1.37	0.171	-.3528391	.0625067
serv. persona	-.0139868	.0742871	-0.19	0.851	-.1595868	.1316133
abitazioni	-.0452475	.0930734	-0.49	0.627	-.2276679	.1371729
non alimen.	-.2552592	.0896985	-2.85	0.004	-.431065	-.0794535
non occupato	-.2202744	.0568272	-3.88	0.000	-.3316537	-.1088951
lav. autonomo	-.2540408	.0470557	-5.40	0.000	-.3462684	-.1618132
18-29 anni	.2425434	.0710184	3.42	0.001	.10335	.3817369
30-44 anni	.0551786	.0420226	1.31	0.189	-.0271842	.1375414
60-74 anni	-.1582493	.0672664	-2.35	0.019	-.2900889	-.0264096
oltre 75 anni	-.1630874	.0923432	-1.77	0.077	-.3440767	.017902
inbank	.3391054	.0380069	8.92	0.000	.2646133	.4135974

Essere socio aumenta la probabilità di avere almeno una delle tre carte del 4.49%, mentre essere un'impresa la fa diminuire del 48.67%. I gruppi economici che diminuiscono la probabilità di possesso sono "Produzione/lavorazione metalli, macchinari elettromeccanici" e "Articoli non alimentari", rispettivamente con il 21.59% e il 25.53%. Il fatto di essere non occupato o lavoratore autonomo fa diminuire la probabilità del 22.03% e del 25.40%. Per quanto riguarda le fasce d'età, si può notare come la fascia più giovane, cioè "Fino a 29", comporti un aumento della probabilità del 24.25%. La fascia "dai 65 ai 74" anni invece

contribuisce ad una diminuzione di questa probabilità del 15.82%. Infine l'accesso all'home banking aumenta la probabilità di possedere almeno uno tra bancomat, carta di credito o "carta ricarica" del 33.91%. Vediamo ora come si comporta il modello nella percentuale di risultati predetti correttamente della variabile dipendente in oggetto:

Probit model for almeno

Classified	----- True -----		Total
	D	~D	
+	500	104	604
-	73	181	254
Total	573	285	858

Classified + if predicted Pr(D) >= .5

True D defined as almeno != 0

Sensitivity	Pr(+ D)	87.26%
Specificity	Pr(- ~D)	63.51%
Positive predictive value	Pr(D +)	82.78%
Negative predictive value	Pr(~D -)	71.26%
False + rate for true ~D	Pr(+ ~D)	36.49%
False - rate for true D	Pr(- D)	12.74%
False + rate for classified +	Pr(~D +)	17.22%
False - rate for classified -	Pr(D -)	28.74%
Correctly classified		79.37%

Come per il modello precedente, la percentuale di risultati predetti correttamente si aggira intorno all'80%, confermando la bontà del modello in oggetto.

Passiamo ora all'analisi della variabile dipendente *ib_and*, che identifica i possessori di almeno uno tra bancomat, carta di credito o "carta ricarica" e che al contempo hanno accesso al servizio di home banking. Per il modello in oggetto è stato usato lo stesso gruppo di variabili indipendenti utilizzate per il modello precedente, a causa degli stessi problemi riscontrati in quest'ultimo.

Analizziamo ora i risultati ottenuti:

```

Probit regression
Number of obs   =      858
LR chi2(14)     =     203.68
Prob > chi2     =      0.0000
Log likelihood = -491.17738
Pseudo R2      =      0.1717
  
```

ib_and	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
maschio	.0332011	.0968242	0.34	0.732	-.1565709	.2229731
socio	.375487	.1178349	3.19	0.001	.1445348	.6064392
impresa	-1.019574	.2076902	-4.91	0.000	-1.426639	-.6125088
gen. alimen.	-.2974906	.3020972	-0.98	0.325	-.8895902	.2946091
lav. metalli	-.1118853	.3073191	-0.36	0.716	-.7142198	.4904492
serv. personal	.2428595	.2051239	1.18	0.236	-.1591759	.6448949
abitazioni	.0573129	.2667445	0.21	0.830	-.4654967	.5801226
non alimen.	-.4602088	.2608406	-1.76	0.078	-.971447	.0510294
non occupato	-.5615332	.1458535	-3.85	0.000	-.8474007	-.2756657
lav. autonomo	-.5625752	.1224225	-4.60	0.000	-.8025188	-.3226316
18-29 anni	.8619398	.1634033	5.27	0.000	.5416753	1.182204
30-44 anni	.2697049	.1072706	2.51	0.012	.0594584	.4799514
60-74 anni	-.8198753	.216369	-3.79	0.000	-1.243951	-.3957999
oltre 75 anni	-.4147735	.2603387	-1.59	0.111	-.925028	.095481
costante	.3166811	.1096968	2.89	0.004	.1016793	.5316829

Proviamo ora ad aggiungere la variabile *uso_ib* tra le variabili indipendenti, cioè l'utilizzo dell'home banking nel periodo considerato:

```

Probit regression                               Number of obs   =           858
                                                LR chi2(15)     =           213.73
                                                Prob > chi2     =           0.0000
Log likelihood = -486.15408                    Pseudo R2      =           0.1802

```

ib_and	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
maschio	.0194317	.0974829	0.20	0.842	-.1716313	.2104947
socio	.3801895	.1187061	3.20	0.001	.1475298	.6128491
impresa	-1.164746	.2168002	-5.37	0.000	-1.589667	-.7398257
gen. alimen.	-.3463481	.3128132	-1.11	0.268	-.9594507	.2667546
lav. metalli	-.1483066	.3106285	-0.48	0.633	-.7571273	.460514
serv. persona	.1731864	.2082929	0.83	0.406	-.2350602	.581433
abitazioni	-.0265443	.2689134	-0.10	0.921	-.5536048	.5005162
non alimen.	-.4838503	.2603388	-1.86	0.063	-.9941049	.0264042
non occupato	-.5412357	.1462734	-3.70	0.000	-.8279262	-.2545452
lav. autonomo	-.5876226	.1235396	-4.76	0.000	-.8297558	-.3454894
18-29 anni	.8669112	.1637773	5.29	0.000	.5459137	1.187909
30-44 anni	.2555922	.1081514	2.36	0.018	.0436193	.467565
60-74 anni	-.8580578	.2173585	-3.95	0.000	-1.284073	-.4320431
oltre 75 anni	-.4095802	.2611599	-1.57	0.117	-.9214442	.1022839
uso_ib	.4362021	.1395519	3.13	0.002	.1626855	.7097187
costante	.2868421	.1107008	2.59	0.010	.0698725	.5038117

L'aggiunta della variabile *uso_ib*, essendo quest'ultima significativa, migliora la bontà del modello. Le variabili significative sono quasi le stesse di quelle osservate nel modello in cui veniva analizzata la variabile *almeno*.

Vediamo ora gli effetti marginali condizionati:

```

Conditional marginal effects          Number of obs   =          858
Model VCE      : OIM
Expression     : Pr(ib_and), predict()
-----
          |              Delta-method
          |      dy/dx   Std. Err.      z    P>|z|    [95% Conf. Interval]
-----+-----
maschio |   .0077368   .0388136    0.20   0.842   -.0683365   .0838101
socio |   .1513738   .0472744    3.20   0.001   .0587178   .2440299
impresa |  -.4637481   .0864829   -5.36   0.000  -.6332514  -.2942447
gen. alimen. | -.1378998   .1245653   -1.11   0.268  -.3820432   .1062437
lav. metalli | -.0590488   .1236782   -0.48   0.633  -.3014537   .183356
serv. personal | .0689548   .0829406    0.83   0.406  -.0936058   .2315154
abitazioni | -.0105687   .1070684   -0.10   0.921  -.220419   .1992816
non alimen. | -.1926468   .1036758   -1.86   0.063  -.3958477   .0105541
non occupato | -.215495   .0582122   -3.70   0.000  -.3295887  -.1014013
lav. autonomo | -.2339641   .0491721   -4.76   0.000  -.3303397  -.1375885
18-29 anni |   .3451639   .0651425    5.30   0.000   .2174869   .4728409
30-44 anni |   .101765   .0430592    2.36   0.018   .0173705   .1861594
60-74 anni |  -.3416389   .0866609   -3.94   0.000  -.5114912  -.1717866
oltre 75 anni | -.1630759   .1039987   -1.57   0.117  -.3669095   .0407578
uso_ib |   .1736755   .0555627    3.13   0.002   .0647746   .2825764

```

Essere socio della banca aumenta del 15.14% la probabilità di avere accesso all'home banking e uno tra bancomat, carta di credito e "carta ricarica". Essere un'impresa diminuisce questa probabilità del 46.37%. Come visto con i modelli precedenti anche in questo caso, essere non occupato o lavoratore autonomo diminuiscono la probabilità rispettivamente del 21.55% e del 23.40%. Anche per quanto riguarda le fasce d'età il trend osservato sembra essere simile, almeno come segno, ai modelli analizzati in precedenza. Nello specifico appartenere alla categoria "Fino a 29 anni" aumenta la probabilità di avere accesso al servizio home banking e uno tra bancomat, carta di credito e "carta ricarica" del 34.52%. Un aumento positivo, seppur inferiore e pari al 10.18%, si ha anche nella

categoria “dai 30 ai 44” anni. Per la fascia “dai 65 ai 74” si ha invece una diminuzione di questa probabilità del 34.16%. L'ultima variabile analizzata è l'utilizzo di home banking, che fa aumentare la probabilità del 17.37%. Vediamo ora la bontà di adattamento attraverso la percentuale di risultati previsti correttamente:

Probit model for ib_and

Classified	----- True -----		Total
	D	~D	
+	356	149	505
-	100	253	353
Total	456	402	858

Classified + if predicted $\Pr(D) \geq .5$

True D defined as ib_and != 0

Sensitivity	$\Pr(+ D)$	78.07%
Specificity	$\Pr(- \sim D)$	62.94%
Positive predictive value	$\Pr(D +)$	70.50%
Negative predictive value	$\Pr(\sim D -)$	71.67%
False + rate for true ~D	$\Pr(+ \sim D)$	37.06%
False - rate for true D	$\Pr(- D)$	21.93%
False + rate for classified +	$\Pr(\sim D +)$	29.50%
False - rate for classified -	$\Pr(D -)$	28.33%
Correctly classified		70.98%

La percentuale di risultati predetti correttamente, seppur inferiore ai modelli precedenti, può essere ritenuta soddisfacente.

Conclusioni

In questa relazione ho analizzato dei dati sulle movimentazioni bancarie di una parte della clientela di una banca cooperativa presso cui ho svolto uno stage di tre mesi (Banca San Giorgio Quinto Valle Agno). La stima econometrica di diversi modelli probit mi ha permesso di evidenziare quali variabili influiscono, in positivo o in negativo, sul possesso o utilizzo dei principali strumenti di pagamento come home banking, bancomat, carta di credito e “carta ricarica”. Uno dei risultati che potrebbe sorprendere è che in tutti i modelli analizzati il genere del titolare della posizione non ha un effetto significativamente diverso da zero: ciò significa che il sesso del cliente non influenza l'utilizzo o il possesso degli strumenti in oggetto. Essere socio della banca risulta significativo solo in due modelli e comunque ha un impatto relativamente modesto. Appartenere alla categoria privato o essere un'impresa diminuisce la probabilità di utilizzo dell'home banking. Per quanto riguarda l'analisi dei modelli di possesso degli strumenti di pagamento, essendo presente la sola variabile relativa all'essere un'impresa, quest'ultima incide negativamente sul loro possesso. Il fatto di essere non occupato o lavoratore autonomo fa diminuire in modo significativo la probabilità di possedere carte di pagamento o di avere accesso a home banking. Ma per chi ha accesso a quest'ultimo, l'essere lavoratore autonomo ha un effetto positivo sulla probabilità di utilizzarlo. Per quanto riguarda l'età, già l'analisi descrittiva mostra come nelle fasce più giovani ci sia un più frequente possesso e utilizzo degli strumenti di pagamento in oggetto rispetto alle fasce più anziane. Questa osservazione viene confermata anche dai risultati ottenuti tramite i modelli probit. Possiamo notare infine come l'accesso all'home banking incida positivamente sul possesso delle carte di pagamento, così come l'utilizzo dell'home banking stesso.

Bibliografia

Banca San Giorgio Quinto Valle Agno, www.bancasangiorgio.it

Belloc Marianna. *Modelli a risposta discreta: Lezioni dottorato – Econometria III*, Giugno 2009.

Bucciol Alessandro. *Lezioni di introduzione all'econometria: Modelli per variabili dipendenti limitate*, 2012.

Cappuccio N. e Orsi R. *Econometria*, Bologna, il Mulino, 2005.

Cramer J.S. *Logit models from economics and other fields*, Cambridge, Cambridge University, 2003.

Maddala G.S. *Introduction to econometrics*, New York (USA), Macmillan Publishing Company, 1992.

Verbeek Marno. *A guide to modern econometrics*, Chichester West Sussex (UK), John Wiley & Sons Ltd, 2004.