Indice

Introduzione

Gran parte del lavoro svolto dagli intermediari finanziari risiede nell'assunzione di rischi e tra questi il principale è il rischio di credito, che si configura come il rischio di insolvenza di un soggetto richiedente un prestito. Data la centralità dell'attività di prestito, non solo all'interno della singola banca, bensì all'interno dell'intero ecosistema bancario, il Comitato di Basilea per la vigilanza bancaria ha stabilito che le banche sono tenute a misurare la probabilità di insolvenza a un anno di ogni cliente, ai fini del calcolo dell'esposizione azionaria dei prestiti e, a tal fine, possono avvalersi di sistemi di rating interni calcolando le proprie stime per la *probability of default* (PD). È da qui che prende le mosse il presente lavoro, il quale ha voluto indagare l'utilizzo dei modelli GLM binomiali per la previsione della probabilità di insolvenza. La regressione logistica, basata su una funzione di link simmetrica, sebbene sia uno dei modelli maggiormente utilizzati dalla letteratura accademica per il calcolo della probabilità di default, mostra notevoli svantaggi negli studi su dati fortemente sbilanciati, un tratto tipico dei dati in questo contesto. Pertanto, le sue performance vengono messe a confronto con quelle della regressione *complementary log-log*, la quale utilizza una funzione di link asimmetrica derivata dalla funzione di ripartizione della variabile casuale *Generalized Extreme Value*, meglio nota come GEV.

La presente trattazione si articola su tre capitoli, strutturati come segue. Il primo capitolo si apre con una panoramica sulla centralità dell'attività di prestito sia all'interno della società, come ingrediente fondamentale per lo sviluppo, che all'interno dell'ecosistema bancario, contribuendo a redditività e liquidità. Si prosegue poi con la definizione del rischio di credito e delle sue diverse componenti, la cui gestione assorbe la quasi totalità del patrimonio di vigilanza delle banche. Successivamente sono stati presentati i principi degli accordi di Basilea in materia di rischio di credito. Il capitolo si conclude con la definizione del *credit scoring*, l'introduzione alle principali tecniche statistiche utilizzate e i limiti di queste ultime in presenza di dati fortemente sbilanciati.

Il secondo capitolo introduce i modelli GLM e le loro componenti principali, quella casuale e quella sistematica, per poi concentrarsi sui modelli GLM binomiali. Di questi, viene poi posta attenzione sulla scelta della funzione di link più appropriata. Vengono, quindi, presentate le funzioni di ripartizione che verranno utilizzate per la costruzione della funzione di link: quella della variabile casuale logistica per il modello di regressione *logit* e quella della variabile casuale di Gumbel per il modello di regressione *complementary-*

loglog. La variabile di Gumbel viene introdotta all'interno del contesto della teoria del valore estremo, in quanto deriva da un caso particolare della distribuzione GEV, *Generalized Extreme Value*, di cui viene studiata sia la distribuzione *for maxima* che la distribuzione *for minima*. Successivamente si è passati alla stima dei parametri di regressione per poi procedere con gli strumenti per la validazione e la selezione del modello. Il capitolo si conclude con il passaggio dalla regressione alla classificazione per mezzo di una *threshold* appropriata e con la presentazione degli strumenti per la valutazione delle capacità previsive del modello.

Il terzo capitolo ha voluto testare i modelli presentati al secondo capitolo su dei dati reali e, nello specifico, è stato selezionato un database contenente dati finanziari di una banca ceca. La prima fase è stata quella di ricodifica delle variabili delle 8 relazioni di cui il database è composto, relazioni di cui poi è stato fatto il merge ricavando tutte le informazioni opportune per ogni account richiedente un prestito. Successivamente ad un'opportuna fase di pre-processing, i dati sono stati suddivisi in training set e test set. Ciò ha permesso di addestrare i modelli *logit* e *complementary log-log* sul training set e testare le previsioni ottenute sul test set. Lo scopo di questo lavoro è quello di confrontare le performance, sia in termini di bontà di adattamento che in termini di accuratezza previsiva, di questi due modelli GLM binomiali, il primo basato su una funzione di link simmetrica e il secondo basato su una funzione di link asimmetrica.