

Studi e ricerche
N. 2/2010



**AGENZIA
DELLE
DOGANE**



Giulio Dal Savio Gianluca Dari Davide Russo

ANALISI DEI RISCHI: un approccio matematico-probabilistico

Ufficio Centrale Antifrode - Ufficio Studi economico-fiscali



ANALISI DEI RISCHI:
un approccio matematico-probabilistico

Giulio Dal Savio

Gianluca Dari

Davide Russo

INDICE

1	Introduzione	1
2	Dinamica ambientale, dualità e obiettivi conflittuali	5
3	Tipologia dei controlli.....	8
4	Una possibile formalizzazione: il modello logit	11
4.1	Il modello statistico	11
4.2	Stima della probabilità di accadimento di una frode.....	12
4.3	Stima delle probabilità di non essere controllato e di esito negativo.....	14
5	Individuazione delle aree di rischio	15
5.1	Gestione degli eventi.....	15
5.2	La matrice di relazioni	15
5.3	La matrice di relazioni: il caso generale	17
6	Stime e risultati dei modelli logit su dati simulati	19
7	Conclusioni	30
8	Bibliografia	31

1 Introduzione

Il processo d'integrazione e d'interdipendenza dei sistemi economici rappresenta l'espressione più ampia delle politiche di liberalizzazione adottate dagli Stati. La nascita di aree di libero scambio, unioni doganali, mercati comuni e unioni economiche, si pone quale naturale conseguenza del progressivo abbandono delle politiche protezionistiche. Il fenomeno evolutivo, quantitativamente illustrato dalle statistiche sul commercio internazionale in continua espansione, reca con sé diverse problematiche che per complessità ed entità non hanno precedenti nella storia. Quello che si registra è infatti una vera rivoluzione che, data la sua dimensione, incide sugli assetti socio - economici - politici degli Stati. Nel paradigma evolutivo, così sinteticamente descritto, si svilisce il ruolo storico della dogana. Se tradizionalmente, per esigenze di ordine economico, il controllo delle frontiere trovava proprie ragioni di esistenza, il progressivo allentamento dei dazi e contingentamenti sulle operazioni di import - export ha snaturato il suo tradizionale compito istituzionale: non più sentinella dedita a proteggere l'industria nazionale ed assicurare introiti alle casse statali, ma soggetto investito di compiti del tutto nuovi. E ciò non deve sorprendere.

Con la vertiginosa espansione del commercio internazionale è consequenziale attribuire alla dogana una funzione del tutto nuova che si esplicita in almeno due direzioni. La prima concerne la tutela del bene pubblico nelle sue diverse modalità. A titolo indicativo, e non esaustivo, si ricorda la tutela della salute pubblica e dell'ambiente, tutela che necessariamente deve avvenire con un'opera di prevenzione tesa a scongiurare il traffico di sostanze nocive per scopi illeciti, o comunque di sostanze non accompagnate da standard di sicurezza comunemente accettati e riconosciuti. La seconda direttrice involge aspetti di ordine economico. L'opera di contrasto e di deterrenza avverso le contraffazioni dei marchi, le opere dell'ingegno, la proprietà intellettuale, oltre a cautelare i titolari di tali diritti, salvaguarda gli stessi consumatori che in buona fede hanno acquistato quel bene perché ritenuto autentico. E' infatti giocoforza dedurre che dagli innumerevoli illeciti, concernenti le merci scambiate, si intacca sia la credibilità dei mercati, sia la fiducia dei consumatori.

C'è infine un particolare aspetto da sottolineare: il rispetto delle regole competitive. Non va infatti dimenticato che il progressivo allentamento dei vincoli al commercio internazionale non può andare a discapito delle regole concorrenziali a fondamento dell'UE e dei trattati stipulati in più ampi consessi quali il World Trade Organization. Le normative antidumping, il divieto di commercializzare prodotti contraffatti, sono esempi di regole atte a impedire distorsioni di mercato

solo se applicate. La necessità dei controlli sui flussi di import-export trova la sua ragione di esistenza negli effetti deleteri, fortemente correlati, delle frodi commerciali, delle pratiche collusive e distorsive della concorrenza. Ponendosi quale presenza stabile, e titolare delle innumerevoli attività di monitoraggio, la dogana assurge a dignità di istituzione irrinunciabile con funzioni tutelanti gli interessi mondiali e quello dei singoli paesi.

Ciò detto, è pressoché infinito l'insieme dei fattori che interagiscono in realtà sempre più interdipendenti ed integrate (c.d. globalizzazione). La variabilità ambientale, e le sue complesse dinamiche, non possono essere analizzate se non ricorrendo allo studio delle relazioni di causa-effetto e alle correlazioni fra eventi e risultati. Oggi più che mai è dunque necessario utilizzare gli strumenti propri dell'analisi matematica, della disciplina economica ed econometrica, delle scienze statistiche. Questa necessità, però, non deve apparire come una semplice disquisizione accademica nella quale trova spazio anche la più stravagante delle idee; tutt'altro. I risvolti operativi che esse generano sono fondamentali, soprattutto quando le suddette scienze si fondono in una disciplina unitaria che analizza i comportamenti strategici degli agenti. Modellizzare il contesto ambientale alla stregua di un puzzle avente un'enormità di componenti è operazione certamente ardua; ma ancor più arduo è capire come le componenti si incastrano, si aggiungono, o si sottraggono, e quindi, in definitiva, quali siano le complesse dinamiche che si instaurano. Quest'ultime, peraltro, possono condurre a risultati talmente imprevedibili da modificare, anche totalmente, la fisionomia di quel mosaico originario che, con notevole sforzo concettuale, si era descritto. L'analisi dei rischi è figlia della variabilità ambientale, non potrebbe esistere senza variabilità. In essa si rinvengono innumerevoli elementi i quali, anche se presi singolarmente, sono di difficile modellizzazione. Vi è poi la loro interazione, una sorta di miscela avente una reazione chimica non sempre decifrabile, o nota a priori. Come si cercherà di evidenziare nel presente lavoro, alle variabili economiche si devono aggiungere quelle relative ai comportamenti dei soggetti aventi, spesso, obiettivi conflittuali fra loro. Tale conflittualità (che rende come strategici i suddetti comportamenti) implica lo studio delle interazioni fra variabili obiettive e soggettive. L'analisi dei rischi, in definitiva, raccogliendo in se una serie di argomentazioni che vanno valutate congiuntamente, costituisce un tema dall'enorme complessità che non può essere affrontata senza un'adeguata attività di collaborazione fra Paesi e senza un'adeguata attività di intelligence integrata fra Stati. Il punto è di cruciale importanza. L'uso razionale delle informazione (c.d. intelligence) implica dapprima la scelta o l'individuazione di quelle significative e, successivamente, l'integrazione a sistema degli input informativi selezionati. La finalità è presto individuata nell'esigenza di aumentare il bagaglio

informativo per fronteggiare al meglio le insidie connesse al proprio ambiente di riferimento (che può essere territorialmente limitato oppure esteso a livello planetario).

Tuttavia, o perché l'informazione è distorta, o perché essa non è mai piena, o perché atteggiamenti di reticenza o di gelosa custodia permangono (è un fenomeno facilmente riscontrabile quello di sapere il più possibile dagli altri e diffondere il meno possibile del proprio), l'informazione è sempre incompleta e/o asimmetrica.

Quando si affronta il tema dell'analisi dei rischi si ha sempre l'impressione di analizzare qualcosa di evanescente, privo di una robustezza di base dalla quale partire. Sebbene possa apparire singolare, è bene partire dal seguente interrogativo: perché ha senso il tema dell'analisi dei rischi? Per la notevole potenza esplicativa, la risposta alla domanda costituisce un punto fermo dal quale iniziare. Esaminando la frase, o meglio la locuzione, è immediato osservare che l'analisi riguarda il rischio, ovvero lo studio di eventi che possono a priori manifestarsi. Dunque, il rischio esiste in quanto l'incertezza è presente (in altri termini, il tema non sussisterebbe se a priori si conoscessero tutte le variabili, come si muovono, e come si correlano). Venendo all'esame sul tipo di incertezza, è riduttivo riferirsi esclusivamente a quella di tipo obiettivo, quale, ad esempio, l'incertezza relativa all'uscita di un dato numero su una ruota del lotto, o della faccia che si presenta prima del lancio di un dado. Bisogna allora riferirsi all'incertezza in senso più ampio, ovvero quella che contempla anche, e soprattutto, il comportamento strategico dei soggetti. Così facendo, è immediato dedurre che ogni scelta, di natura socio – economica - politica degli agenti (governi, enti, istituzioni, imprese e operatori) è, al tempo stesso, input - output. I soggetti, in altri termini, prendono decisioni (output) sulla base di ciò che osservano; le stesse decisioni producono effetti su quelle successive, costituendone pertanto l'input. Anche l'Autorità Doganale, in quanto agente, non si sottrae all'alone di incertezza permeante le sue scelte e le sue azioni; anch'essa è costretta al continuo monitoraggio ambientale nel tentativo di ridurre l'alea. Procedendo secondo un ordine logico, è necessario fissare subito le linee generali. In primis, l'Autorità Doganale opera in un contesto ambientale più vasto di quello relativo ad un altro agente privato (esempio l'impresa). Infatti, in seguito alle politiche di liberalizzazione delle economie, l'ambito d'intervento della Dogana si va estendendo su scala planetaria. In questa cornice di carattere generale, v'è poi il problema dell'informazione incompleta che riguarda tutti gli agenti. Nello specifico, il set informativo¹ della Dogana si caratterizza per

¹ Il set di cui si discorre, è propriamente definito come un sottoinsieme di tutti i possibili stati di natura e connesse probabilità.

incompletezza e asimmetria nella forma di Moral Hazard². Tra l'altro, sull'A.D. incorre anche la possibile asimmetria della selezione avversa sul sistema di controllo del traffico.

Il presente lavoro percorre l'arduo sentiero dell'analisi dei rischi caratterizzata da ostacoli e incertezze. La formalizzazione della dualità e obiettivi conflittuale sarà immediatamente discussa al paragrafo 1. Come si vedrà, la questione verterà esclusivamente sulla probabilità che un evento accada. Non meno importante è la discussione sulla tipologia dei controlli espletati tenuto conto del vincolo delle risorse. Problemi concernenti la produttività degli input (lavoro e capitale) non sono stati affrontati in quanto esulano dagli scopi del presente lavoro. Il modello implementato per la stima delle probabilità è di tipo logistico. Attraverso i Logit presentati sono emerse numerose questioni; alcune "risolte", le altre invece meritano un'analisi molto più accurata al fine di trovare la metodologia "risolutiva" migliore. Per quest'ultime si è fatto rinvio ad altra sede.

Sempre nel corso del lavoro si è sviluppata un'idea fortemente ancorata alle relazioni che possono instaurarsi nell'ambiente investigato. Come sempre accade in ogni lavoro mirante a formalizzare una realtà caratterizzata da leggi fisiche immutabili (il tempo ad esempio) e da comportamenti umani (che sono diversi e variabili nel tempo e nello spazio), anche il presente studio si espone fortemente alle critiche che possono essergli mosse. Chi ci ha lavorato ne è ben consapevole. L'obiettivo, infatti, non è stato quello di risolvere qualcosa che ci appare irrisolvibile, ma di fornire un contributo. Se ci sarà, sarà per noi fonte di immensa soddisfazione.

² Il concetto di "azzardo morale" è ben conosciuto dalla teoria economica e si manifesta in numerosi aspetti della vita. Tipici esempi si rinvengono nei contratti di assicurazione e nei rapporti tra fisco e contribuenti nei quali l'agente più svantaggiato è colui che riceve la dichiarazione (furto o sinistro, denuncia dei redditi) senza sapere, a priori, se quanto asserito corrisponda a realtà.

2 Dinamica ambientale, dualità e obiettivi conflittuali

Su un piano di assoluta generalità, l'ambiente si caratterizza per la presenza di almeno due agenti che interagiscono fra loro con finalità conflittuali. Usando proprie variabili strategiche, ciascun soggetto persegue l'obiettivo di ottimizzare una data funzione tenendo conto delle possibili reazioni dell'avversario. Dal punto di vista eminentemente pratico i titolari delle opposte finalità sono individuati nell'Autorità Doganale e nell'operatore economico frodante³.

A mezzo della semplice introduzione di payoff antagonisti, la dinamica ambientale si caratterizza per un'accesa sfida che non terminerà mai con un verdetto definitivo⁴. All'aumento del numero delle frodi, l'Autorità Doganale reagirà con un'intensificazione dei controlli; viceversa, una riduzione del monitoraggio sulle merci, oppure una perdita di efficacia, costituirà linfa vitale per la proliferazione delle attività frodanti. Senza perdita di generalità, iniziamo proprio dall'obiettivo economico (massimizzazione dei profitti) in violazioni di norme e regolamenti perseguito dall'operatore. E' del tutto naturale dedurre che la sua strategia è finalizzata a minimizzare la funzione di perdita (ovvero il rischio di essere scoperto) o, equivalentemente, a massimizzare la probabilità di non essere intercettato. In definitiva, massimizzare un guadagno o minimizzare una perdita hanno per soluzione il medesimo risultato (c.d. dualità).

Parallelamente, la dualità è altresì presente nella funzione obiettivo dell'Autorità Doganale: scoprire il maggior numero di frodi è equivalente a minimizzare il rischio che una di esse non sia intercettata.

Si supponga ora di estraniarsi dall'ambiente per poter osservare i soggetti indicati da una posizione terza, caratterizzata cioè da assoluta imparzialità.

All'osservazione non potrà sfuggire che l'operatore frodante produrrà una documentazione (la dichiarazione doganale) con variabili che riducono al minimo la probabilità di essere controllato (o se si vuole massimizzano la probabilità di sfuggire al controllo).

Al tempo stesso, la sua strategia non prescindere certamente da un secondo calcolo, ovvero minimizzare la possibilità di risultare positivo qualora fosse controllato. Per l'analista

³ Per l'analisi dei rischi è infatti del tutto irrilevante captare la strategia dell'operatore economico che agisce nel rispetto delle regole.

⁴ In altri termini, non ci sarà mai un vinto e un vincitore. Questa tipologia di comportamenti strategici, che si esplicano in ambienti territoriali più o meno estesi, può ben rappresentarsi con il modello biologico preda-predatore (Equazioni di Volterra-Lotka). Il sistema dinamico dà vita ad un ciclo (ascesa e discesa) determinato da un punto di equilibrio stabile. L'estinzione di entrambe le specie rappresenta invece un punto di equilibrio instabile (0,0).

(l'osservatore) le modalità strategiche comportamentali non possono assumere infiniti valori. Presto si renderà conto che esse sono in numero finito e ciascuna con una determinata frequenza.

Per esempio, potrà osservare che l'operatore indirizza il suo traffico dove i controlli sono fatti con scarsa attenzione; oppure, supponendo che la necessità di evitare collassi (intasamenti e code) predomini sulla necessità di effettuare controlli accurati, indirizza le sue operazioni laddove si registra un'alta densità del traffico di merci.

Se, dunque, l'obiettivo del frodante è quello di minimizzare l'esito di risultare positivo al controllo (invero è l'unico obiettivo possibile), è giocoforza introdurre due probabilità che sintetizzano il suo comportamento. La prima può essere definita come $P(\text{NC})$, da interpretarsi come la probabilità P che l'operazione sfugga al controllo (NC = non controllo); l'altra probabilità è indicata con $P(0|\text{C})$, ovvero la probabilità che la merce superi il controllo con positività zero.

Entrambe le probabilità possono essere stimate servendosi di tecniche appropriate. Fra queste si segnalano quelle fatte con modelli non lineari. E' fondamentale rimarcare un punto già sollevato. Riguarda proprio $P(\text{NC})$ e $P(0|\text{C})$. Si è già detto che esse sintetizzano una determinata strategia frodante. Rimane solo un'ulteriore qualificazione: $P(\text{NC})$ e $P(0|\text{C})$ non sono fra loro indipendenti. Esse, piuttosto, manifestano correlazione positiva fra loro (almeno nell'intenzione dell'operatore), nonché correlazione con due variabili comunemente presenti: la dichiarazione del soggetto e la dogana scelta (lo spazio).

L'autorità doganale, affetta da asimmetria informativa, non può conoscere con esattezza $P(\text{NC})$ e $P(0|\text{C})$. Tuttavia, per definizione, ne conosce l'esistenza; conseguentemente è in grado di captare le stesse variabili sulle quali ha operato la strategia frodante. In definitiva, massimizzare il numero di frodi scoperte equivale a ridurre $P(\text{NC})$ nonché $P(0|\text{C})$. La strategia del controllo "ottimo" dovrà quindi esplicitarsi in un aumento dell'efficienza - maggiori controlli riducono $P(\text{NC})$ - e in un aumento dell'efficacia. Come? Dal punto di vista concettuale il problema è di immediata soluzione: riducendo l'asimmetria informativa. Il vero rebus risiede però nell'aspetto metodologico, pragmatico. Procedendo in questa direzione, in primis, ad ogni operazione può essere associata una probabilità (che dovrà essere stimata) di non essere controllata $P(\text{NC})$ ed una di risultare negativa al controllo $P(0|\text{C})$ allorché l'operazione sia effettivamente frodante. Ad una casistica a parte vanno annoverati i falsi positivi⁵.

Da queste brevi note emergono alcune questioni di sintesi. I due agenti sono in contrapposizione fra loro e ciascuno adotterà la strategia reputata vincente. In tale dinamica, l'asimmetria informativa

⁵ E' il caso di merce regolare che per errore risulta positiva al controllo. Tale eventualità (che non può essere esclusa) fornisce anch'essa risultati apprezzabili per l'analisi dei rischi, se non altro perché consente di analizzare le cause di insuccesso. Inoltre, cosa non secondaria, ha sicuramente un impatto sui costi sopportati e susseguenti contenziosi.

gioca un ruolo fondamentale. Su quella che incombe in capo all'autorità doganale si è già detto. Ci si domanda se essa avvolga anche il soggetto frodante. La questione sollevata è di centrale importanza e non può essere licenziata senza sottolinearne le relative implicazioni. A tal fine si supponga che il soggetto operi in assenza di asimmetria. Cosa implica tale eventualità? Conoscendo con esattezza il tipo o i tipi di controllo che vengono svolti, l'operatore frodante adotterà sicuramente la strategia tesa ad eluderli. In siffatta circostanza, $P(NC)$ o $P(0|C)$ assumeranno valori vicini all'unità⁶. I risultati emergenti dai controlli doganali manifesteranno così una efficacia molto ridotta. L'ipotesi alternativa, conduce evidentemente a conclusioni opposte. Per tale ragione, impedire al soggetto di conoscere a priori quale tipo di controllo (leggasi strategia) sarà utilizzato, è un obiettivo che deve essere perseguito con assoluta priorità. Come? Evitando che il sistema di monitoraggio si fondi su una standardizzazione dei controlli. Cicli di routine, abitudinari, se da una parte aumentano l'efficienza, dall'altra sono di facile apprendimento e conducono inevitabilmente ad una riduzione dell'asimmetria informativa a favore del soggetto frodante.

⁶ In effetti, dire che la probabilità sia uguale ad uno significa implicitamente affermare che nessun evento imprevisto si verifichi.

3 Tipologia dei controlli

Dopo aver studiato l'ambiente ed i soggetti che in esso interagiscono, si supponga che il nostro ipotetico osservatore, decida di entrare nel terreno di gioco potendo scegliere quale casacca indossare, ovvero recitare la parte del soggetto frodante o quella dell'autorità doganale. Per via della maggiore informazione acquisita sicuramente, fra quelle possibili, attuerà la strategia migliore. A questo punto, immaginiamo di intervistarlo. La risposta, o meglio le risposte che ci attendiamo fornisca, saranno sicuramente valide e non possono essere trascurate. Ci illustrerà tutti i controlli che ha osservato nonché tutti i frodatori che non sono stati scoperti. Potrebbe anche fornirci una casistica completa e accurata. Fra controlli obbligatori su determinate merci (ad esempio le armi), controlli istituzionali (derivanti, cioè, da profili di rischio esistenti), controlli random, sarà in grado di darci l'esatta mappatura della strategia doganale. Ma questa informazione è già in nostro possesso. Quello a cui saremmo interessati è sapere le strategie adottate dall'operatore frodante. Il nostro ipotetico osservatore è talmente ben informato che ci descrive la dinamica ambientale partendo da una rappresentazione in forma statica.

Indicando, rispettivamente, con x e y le variabili nella bolletta doganale e le variabili di contesto ambientale⁷, si ponga $P(NC)=f(x,y)$ e $P(0|C)=g(x,y)$.

Il soggetto frodatore ha per obiettivo:

$$\text{Max } [P(NC)+P(0|C)]$$

o, equivalentemente:

$$\text{Min } \{2 - [P(NC)+P(0|C)]\}$$

A questo punto, ci mostra che la dogana ha esattamente l'interesse opposto ovvero:

$$\text{Max } \{2 - [P(NC)+P(0|C)]\} = \text{Min } [P(NC)+P(0|C)]$$

Ciò detto, il nostro ipotetico interlocutore non aggiunge altro sostenendo di aver già esposto abbastanza. Alla luce dei ragguagli ottenuti, procediamo autonomamente.

Sapendo che nella dichiarazione doganale sono presenti le variabili x ed y , è naturale procedere alla stima delle probabilità associate all'evento frode. Come? Attraverso un modello a risposta binaria, ad esempio un logit.

Il primo passo da compiere concerne l'estrazione di una certa quantità di operazioni oppure (se possibile) si considerano tutte in modo da eliminare qualsiasi arbitrarietà nella scelta.

⁷ x ed y possono essere due variabili o, più realisticamente, due vettori di variabili.

L'informazione che si vuole ottenere concerne la significatività delle variabili in bolletta; la significatività si riferisce all'effetto marginale di ciascuna di esse.

Il passo successivo è quello di ottimizzare i diversi tipi di controllo sotto la restrizione imposta dal vincolo delle risorse.

In particolare, sia C l'ammontare massimo di controlli esperibili nell'unità di tempo.

Il problema riguarda la ripartizione di C in modo che:

$$C_1 + C_2 + C_3 + \dots + C_n \leq C$$

dove, rispettivamente:

C_1 indica la quantità di controlli obbligatori;

C_2 indica la quantità di controlli istituzionali;

C_3 indica la quantità di controlli casuali;

Per tali controlli è assunta la stretta positività (con ciò si intende dire che sono imprescindibili).

I restanti C_i (con $i = 4 \dots n$) saranno caratterizzati da non negatività: $C_i \geq 0$.

In queste altre classi di controlli possono rientrare tutti quei metodi che non sono oggetto di studio nel presente lavoro ma che nel tempo, per esigenze individuate dagli analisti, potrebbero risultare utili nell'attuazione della strategia complessiva.

Definite le classi dei controlli sorgono due questioni non propriamente secondarie. La prima concerne le variazioni della partizione dei controlli per effetto di ciclicità o stagionalità degli eventi.

L'altra questione attiene alla dimensione di C , direttamente dipendente dalla produttività degli input (lavoro e capitale) impiegati. Ambedue, comunque, non saranno analizzati in questo lavoro.

L'ipotesi che C sia costante non elimina comunque il problema relativo all'ottima ripartizione dei controlli e qualche regola che limita l'arbitrarietà della scelta deve essere introdotta. Si potrebbe infatti dimostrare che il problema di massimo dell'autorità doganale dipende esclusivamente dalla ripartizione di C . Qui forniamo l'asserzione in termini intuitivi, omettendo la relativa dimostrazione scientifica.

Avendo definito con $\text{Max } 2 - [P(\text{NC}) + P(0|C)]$ l'obiettivo dell'autorità doganale, è del tutto ovvio arguire che una distorta allocazione di C_1 , C_2 e C_3 impedirà il raggiungimento dell'obiettivo prefissato.

La prima regola, piuttosto intuitiva, risiede quindi nell'aggiornamento sistematico dei profili di rischio in modo da variare la quantità dei controlli ovvero delle risorse destinate al monitoraggio del traffico. Un modo tutto sommato abbastanza logico è quello di testare i valori delle variabili su cui la funzione di perdita (definita come una funzione di danno derivante da inefficacia del sistema di controllo) assume il massimo o i massimi valori.

4 Una possibile formalizzazione: il modello logit

4.1 Il modello statistico

Un modo per formalizzare un legame tra gli eventi e la loro probabilità di accadere è la scelta di un modello di probabilità non lineare a risposta discreta (precisamente dicotomica).

In più di un'occasione, fra i modelli non lineari si è fatta menzione dei logit. L'importanza di questo modello è testimoniata dal successo che riscuote in numerose applicazioni. L'esempio più famoso (nonché il più semplice da esporre) è il modello logistico a risposta binaria. Per chiarire i termini della questione, si supponga che Y sia una variabile dipendente da una matrice X di regressori, in modo che Y può assumere solo due valori: zero o uno.

Tale caratteristica o proprietà può essere validamente applicata a qualsiasi operazione doganale poichè due, e solamente due, possono essere gli eventi ad essa associati. Nel caso di specie, ponendo ad esempio $Y =$ operazione in dogana, l'evento che $Y = 1$ (l'operazione è frodante) ed $Y = 0$ (l'operazione è regolare) può essere studiata validamente con un modello logit. La matrice X dei regressori sarà invece costituita da un set di variabili (informazioni), ciascuna della quale esercita un'influenza (positiva o negativa) sulla variabile Y .

Oltre alla semplicità insita nella risposta dicotomica, c'è anche un'altra ottima ragione a favore di un modello logistico. Si tratta della sua estensione applicativa. In effetti, se da una parte si è interessati al campionamento casuale di eventi per decifrare il contesto di realizzazione delle nuove frodi, dall'altra non meno interessante è l'analisi di aree a maggiore incertezza, ovvero di quelle aree ove la probabilità del controllo diminuisce fortemente e/o l'operazione contiene spiccati elementi di pericolosità⁸. Le probabilità che si analizzeranno nel corso del presente lavoro, sono

⁸ In realtà la stima delle probabilità di accadimento di uno o più eventi non è l'unico obiettivo che ci si può porre. La pericolosità di una frode si caratterizza per il danno potenziale insita in essa. Può, cioè, aversi una frode con alta probabilità di manifestarsi ma con un potenziale di danno di molto inferiore ad un'altra avente una bassa probabilità ma un elevato danno. A titolo d'esempio, si supponga che Z_1 e Z_2 siano due possibili danni con probabilità p_1 e p_2 di manifestarsi. La pericolosità potenziale è data da $W_i = p_i Z_i$ con $i = 1, 2$. Per rispondere al quesito se una sia più pericolosa dell'altra, occorre non soltanto stimare la probabilità di accadimento, ma anche e soprattutto Z . Quello del danno atteso è certamente un criterio più robusto per classificare la rischiosità di un evento, ma è anche di difficile determinazione e stima. Ogni analisi che punta in tale direzione è sicuramente apprezzabile qualora fornisca un miglioramento conoscitivo della realtà investigata.

tutte probabilità condizionate, cioè misurano il verificarsi dell'evento in dipendenza di altre variabili.

Allo scopo si renderanno utili due modelli. Uno volto a studiare la probabilità di accadimento delle frodi alla luce di variabili che il modello stesso considera significative, mentre l'altro analizza quanto sia improbabile che un controllo venga effettuato su una determinata operazione. Il primo modello deve individuare le strategie di frode in atto, il secondo deve isolare quelle aree ad alta rischiosità per realizzarne delle nuove.

Una sintesi opportuna dei risultati emergenti fornisce gli elementi informativi per aggiornare i profili di rischio esistenti ed introdurne dei nuovi. Ciò detto, il modello logistico (logit), che analizza la probabilità di accadimento di un determinato evento condizionatamente ad altre variabili, ha la seguente rappresentazione:

$$P(1|x) = \frac{e^{\beta x}}{1 + e^{\beta x}}.$$

Tale modello sarà impiegato nel corso del lavoro.

4.2 Stima della probabilità di accadimento di una frode

Il primo modello ha origine dal campionamento casuale ed ha come obiettivo lo studio delle probabilità di essere frode per ogni evento controllato. Il campionamento casuale, cui basare la stima della probabilità di essere frode, è fondamentale: infatti, se da una parte costituisce un valido deterrente a non commettere frodi, dall'altra permette di dirigere i controlli su nuovi o mutati scenari. La dichiarazione, o bolletta doganale, rappresenta l'ineliminabile set informativo su cui basarsi. Fermo restando che la scelta delle variabili è prerogativa dell'analista, di seguito sono indicate quelle che potrebbero essere ritenute più importanti:

numero dei colli (col)

quantità della merce (qm)

valore statistico (vs)

trasporto in container (is_c)

capitolo (cap)

origine (orig)

provenienza (prov)
dazio (dazio)
eventualità che origine e provenienza non coincidano (rot)
tipo registro (regis)

Ovviamente la quantità di variabili a sistema può essere aumentata, conto tenuto degli inconvenienti che essa comporta. Infatti, da un punto di vista statistico (ma anche utilitaristico), l'introduzione di molte variabili aumenta la complessità di lettura dei risultati e contemporaneamente diminuisce l'efficienza del modello⁹.

Ad eccezione di quelle quantitative, tutte le altre sono variabili dummy, cioè indicatrici di una particolarità associata all'evento. A titolo d'esempio, si consideri la dummy "origine della merce". Nel modello essa si rappresenta con un vettore colonna che vale 1 (se l'origine della merce è del Paese A) oppure zero (se l'origine della merce è diversa dal Paese A). La modalità binaria zero – uno consente una prima importante qualificazione: se la merce ha origine in A, non può avere origine anche in B. Tecnicamente ciò implica l'assenza di collinearità, ovvero che i due vettori colonna siano linearmente indipendenti. Altro aspetto piuttosto importante da considerare concerne alcune operazioni logiche applicabili. Se ad esempio si volesse individuare quegli eventi che abbiano origine in A ma provenienza da B, basterà moltiplicare i rispettivi valori per ogni riga. Se il prodotto è pari ad "1" allora la condizione è vera; qualora, invece, fosse "0", vale la considerazione opposta. In definitiva, riepilogando, se D_1 e D_2 sono due dummy che indicano, rispettivamente, origine e provenienza, allora il prodotto (elemento per elemento) $D_3 = D_1 * D_2$ è a sua volta una variabile dummy a valori zero e uno (in termini di logica si afferma che si stanno cercando le osservazioni che ammettono due caratteristiche contemporaneamente, c.d. condizione AND).

⁹ All'aumento delle variabili, aumenta il rischio di introdurre ridondanza o problemi di collinearità.

4.3 Stima delle probabilità di non essere controllato e di esito negativo

Nel secondo modello logit si è interessati alla stima della probabilità di non essere controllato indipendentemente dal tipo di controlli adottati (casuali o mirati). Tale probabilità si riferisce a quegli eventi per i quali il controllo è meno frequente. L'esito o gli esiti che si vogliono studiare concerne quindi la probabilità totale di non essere controllato. In questo caso, la variabile geografica assume spesso un peso decisivo (in taluni casi è la sola che spiega l'intera variabilità del fenomeno). Al pari del precedente modello, anche per quest'ultimo (stimante la probabilità di non essere controllato) si può utilizzare lo stesso set informativo, tenuto conto delle medesime avvertenze già esplicitate. Ma oltre alla stima di $P(NC)$, è possibile quantificare anche $P(0|C)$ ovvero la probabilità che una determinata merce superi il controllo a cui è sottoposta. Per entrambe, comunque, le variabili di contesto geografiche-territoriali, possono essere catturate da dummy a qualsiasi livello di dettaglio (direzione regionale, circoscrizioni doganali, provincia di destinazione) o di generalizzazione.

5 Individuazione delle aree di rischio

5.1 Gestione degli eventi

Una volta che ad ogni evento sia attribuita una probabilità (ed eventualmente un danno), dobbiamo essere in grado di raggruppare gli eventi secondo caratteristiche comuni, in modo che siano sintetizzati in profili di rischio. Tale raggruppamento, possibile proprio grazie al metodo inferenziale scelto nei controlli, mira a contenere problemi di ingovernabilità dovuti ad un'eccessiva proliferazione dei profili di rischio. L'intento non è quindi di analizzare ogni evento, quanto quello di individuare cause simili di determinazione delle frodi.

Una possibilità è rappresentata dalla distanza degli eventi, ad esempio trattando i vettori riga (gli eventi collezionati nei controlli) come elementi di uno spazio euclideo n-dimensionale, dove le n dimensioni sono tutte le variabili considerate interessanti.

Altro aspetto di non secondaria importanza è lo studio del segno dei coefficienti delle variabili significative. Nei modelli logit, il segno (positivo o negativo) fornisce indicazioni sull'impatto che la variabile esercita sulla probabilità di accadimento di un evento. Infatti la derivata parziale:

$$\frac{\partial p(1|x)}{\partial x_i} = f(\beta x) \beta_i.$$

fornisce la misura dell'effetto marginale per variazioni infinitesimali.

Si noti che l'effetto marginale è funzione di x . Ma il suo segno, essendo $f(x)$ una funzione di densità e quindi non negativa, è legato a quello di β . A volte si cerca di specificare un effetto marginale medio, ad esempio calcolando la densità per i valori medi delle variabili x , ma la rilevanza maggiore è data sempre al segno del coefficiente.

Infine, è del tutto intuitivo rivolgere l'analisi a quelle variabili aventi un impatto al margine più elevato.

5.2 La matrice di relazioni

Nel corso dell'introduzione al presente lavoro, si è fatto cenno al trattamento razionale delle informazioni (c.d. intelligence). Senza dubbio il significato è talmente generale che potrà apparire vago o sfuggente. Qualche chiarimento in proposito può quindi tornare utile. La realtà ambientale

fornisce una quantità impressionante di informazioni: lo stato del tempo, lo stato del traffico stradale, le notizie dei mass-media, ne sono qualche esempio. Ma quali, fra le migliaia di informazioni, sono da selezionare? Fra le innumerevoli informazioni insite nel contesto ambientale, evidentemente un criterio selettivo valido (o anche più criteri selettivi), si fonderà sull'utilità di ciascuna di esse. Ma l'utilità non ha alcun senso se non viene preventivamente specificato l'obiettivo che si intende perseguire. Ad esempio, per un individuo può avere grande utilità conoscere la situazione del traffico stradale qualora debba recarsi in una determinata zona o località.

Dunque, le informazioni debbono essere estratte secondo l'utilità che arrecano per il conseguimento di un obiettivo prefissato. Nel campo della prevenzione, la questione è sempre stata di estrema importanza: un conto è analizzare le cause che hanno prodotto un determinato evento; ben altro è analizzare le cause che possono produrre un determinato evento.

In quest'ultima ottica la matrice delle relazioni ha una considerevole importanza. Ma in cosa consiste esattamente? Un modo per sintetizzare in modo schematico gli input informativi (o eventi) stabilendo i legami che si instaurano fra essi. La sua forma più semplice è definita da una matrice quadrata M di cui m_{ij} rappresenta il generico elemento e rappresenta la presenza di una relazione tra gli elementi i e j . In primis, se la relazione esiste, allora essa può essere ipotizzata simmetrica¹⁰. Tale proprietà implica che se l'evento i è relazionato all'evento j , allora è vero anche il contrario. Inoltre, è del tutto ovvio porre ogni evento in relazione con se stesso (c.d. proprietà riflessiva della relazione). Simmetria e riflessività implicano che la matrice sia costituita da valori unitari sulla diagonale principale e che $m_{ij} = m_{ji}$. Come si vedrà a breve, è altresì possibile l'estensione al caso generale. A questo stadio d'analisi, emerge già qualche risultato utile. Considerati due eventi i e j , se fra essi esiste una relazione binaria diretta allora si può porre $m_{ij}=1$; diversamente, m_{ij} assumerà valore zero. E' possibile però andare ben oltre. Si supponga che fra i e j la relazione diretta intercorre k volte. In questo caso si può porre $m_{ij} = k$, dove k rappresenta la frequenza. A titolo di esempio, se Tizio e Caio si sono incontrati in un bar 10 volte nell'arco di un mese allora, sulla base di quanto asserito, $m_{ij} = 10$.

¹⁰ L'ipotesi di simmetria è generalmente valida. Esistono comunque delle eccezioni che possono violarla. Per esempio Tizio può conoscere Caio, senza che sia vero il viceversa.

5.3 La matrice di relazioni: il caso generale

L'esistenza di una relazione binaria diretta tra due eventi (informazioni, set di informazioni, individui, oggetti, insiemi, entità) è svelata dal valore che assume il termine m_{ij} . La questione che ora si va ponendo riguarda la possibile estensione al caso generale, ovvero l'esistenza di relazioni fra eventi la cui numerosità sia maggiore di due. La risposta non solo è affermativa, ma addirittura trascina con sé un sorprendente risultato: non solo è possibile la generalizzazione, ma addirittura essa svela l'esistenza di eventuali relazioni indirette le quali sfuggono alla formalizzazione sin qui descritta. Per chiarire i termini della questione, si supponga di avere tre eventi i, j, h . Qualora essi sono in relazione diretta avremo m_{ijh} diverso da zero. Ma se uno dei tre non fosse direttamente relazionato agli altri due, cosa accadrà? Una rappresentazione grafica può essere di ausilio:

Figura 1

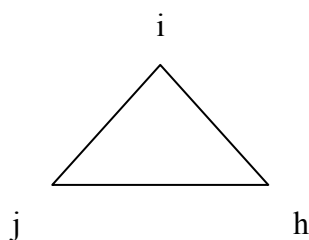
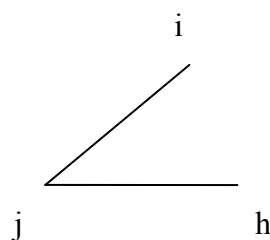


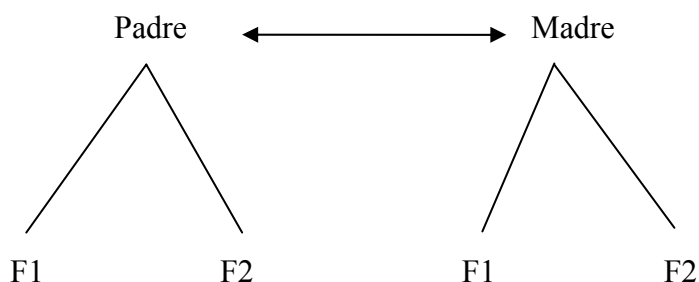
Figura 2



In figura 1 è mostrato il caso in cui ciascun evento è direttamente relazionato agli altri due. Quante volte intercorre la relazione è messa in luce dalla frequenza.

Il caso prospettato in figura 2 è invece molto più interessante. Una relazione diretta c'è sicuramente fra le coppie $(i;j)$ e $(j;h)$ mentre non si registra fra i e h . Per come è stato costruito il termine m_{ij} , dovremmo concludere che m_{ih} sia uguale a zero, corrispondente ad una mancanza di relazione. Le cose, però, non stanno esattamente così, anzi tutt'altro: la relazione esiste ma non è svelabile da un legame diretto. Il caso è interessante proprio perché si sposa perfettamente con qualsiasi realtà ambientale da investigare. Al momento è bene sottolineare il risultato dovuto all'aggiunta dell'evento h . Quest'ultimo caso, infatti, non esclude l'esistenza di collegamenti secondari. Come spesso accade gli esempi possono chiarire concetti molto complessi: si consideri il rapporto di parentela fra individui. Naturalmente fra padre e figlio esiste un rapporto diretto e inoltre l'uno implica l'altro (simmetria). Ma esiste anche la parentela acquisita (gli affini), ovvero dei legami

indiretti. Guardando ancora all'esempio dei legami familiari una relazione diretta viene a mancare persino fra fratelli e/o sorelle. Anche questa ipotesi si presta ad una rappresentazione grafica:



La schematizzazione adoperata è in grado di cogliere la complessa realtà doganale¹¹. Al momento l'analisi si concentra sui legami che possono instaurarsi. Due eventi (leggasi “operazioni doganali”) sono in relazione fra loro se registrano, ad esempio, la stessa origine geografica. L'idea può essere immediatamente estesa: due eventi sono in relazione fra loro se, oltre alla stessa origine, si riferiscono alla stessa merce. Aumentando le variabili di legame aumenta la complessità della matrice di relazioni a vantaggio però di una migliore estrazione informativa. Esistono delle relazioni più deboli (quelle indirette) o più forti (quelle dirette); tutte comunque sono originate da variabili che spiegano un determinato fatto accaduto oppure, in chiave preventiva, concorrono a rendere più o meno probabile l'accadimento di un evento. Va da sé che la matrice pur potendo svelare l'esistenza di legami a qualsiasi livello di intensità non è in grado di stabilire se essi abbiano natura lecita o illecita¹². Essa, dunque, adempie al solo compito di sistematizzare una mole impressionante di informazioni. Sebbene la questione sia in fase di studio, è fuori discussione la sua generale applicazione e la sua potenziale utilità¹³.

E' possibile concludere che sia impresa impossibile (nel senso che è priva di fondamento logico-scientifico) captare in anticipo come si muoveranno le frodi o le strategie frodanti senza dotarsi di un corretto schema relazionale.

¹¹ Se poi fossero introdotte le relazioni di ordine (che abbiano quindi la proprietà di antisimmetria) le operazioni potrebbero essere classificate in base alla loro pericolosità. Ma quello del danno potenziale è un punto successivo.

¹² La matrice di relazioni risponde solo al test “accertare l'esistenza di legami fra entità”. Qualora le entità fossero individui, allora le relazioni che si instaurano possono essere di qualsiasi tipo, ivi compreso il movente economico. In tal caso, è necessario tenere ben distinte le relazioni economiche lecite da quelle illegali e illecite (tipico caso è l'associazione a delinquere).

¹³ In effetti è applicabilissima alla realtà doganale mentre, in chiave utilitaristica, una matrice con nodi, legami diretti e indiretti di grado qualsiasi, è un formidabile strumento per ridurre, o almeno contenere, l'asimmetria informativa giocante sempre un ruolo decisivo nella scelta della strategia.

6 Stime e risultati dei modelli logit su dati simulati

Gli eventi di cui si valuta la probabilità di accadimento sono tutte le importazioni effettuate indipendentemente dal regime. La prima stima effettuata riguarda le frodi scoperte con i controlli svolti. In mancanza di un campionamento casuale, si sono generati i dati relativi ai controlli doganali di un particolare mese T.

Secondo una prima procedura di estrazione e generazione, nel mese T si sono registrate circa 470.000 operazioni di cui 60.000 circa sottoposte a controllo con positività del 3,8% (circa 2400 operazioni frodanti sono state rilevate). Le variabili di interesse sono quelle definite nei paragrafi precedenti. Fra queste le dummy si riferiscono a 30 paesi di origine ed ai 30 paesi di provenienza (che coprono oltre il 90% degli eventi osservati nel mese T) ordinate in maniera decrescente secondo la loro frequenza. Per esempio *d_orig1* indica il primo Paese di origine più frequente, *d_orig2* il secondo Paese, e così via fino ad *d_orig30*. Stesso criterio vale per le dummy con suffisso *d_prov* indicante la provenienza.

La medesima nomenclatura si è usata per tutte le dummy di categoria, anch'esse ordinate per ordine di importanza. Ad esempio *d_cap5* è il nome di una variabile dummy che vale 1 se l'evento è riferito al capitolo 5 e 0 altrimenti. Si noti bene che con capitolo 5 non si intende il capitolo numerato secondo la convenzione taric, ma il capitolo che si pone al 5° posto nella graduatoria dei più frequentemente osservati negli eventi di importazione. Anche per il tipo di registro si è usato lo stesso criterio. In sintesi, il primo logit implementato è condizionato a 30 dummy per il paese di origine, 30 dummy per il paese di provenienza, 8 per il tipo registro, 97 per i capitoli, 3 per trasporto, 2 per la destinazione italiana e San Marino, una per la merce trasportata in container, un'altra afferente al dazio, e infine una dummy per il caso di rotta non diretta (cioè origine e provenienza non coincidono). Il valore statistico, la quantità della merce, il numero dei colli, il dazio, sono le ulteriori variabili esplicative. Gli eventi osservati (in questo caso i sottoposti a controllo) sono circa 60.000.

A seguito di alcune iterazioni per semplificare il modello, cioè procedure di test del rapporto di verosimiglianze, si arriva ad un primo prospetto sintetico:

Tabella 1

Logistic regression		Number of obs = 62448			
LR chi2(68) = 1183.51					
Prob > chi2 = 0.0000					
Log likelihood = -9502.1743		Pseudo R2 = 0.0586			
esito	Coef.	Std. Err.	Z	P> z	[95% Conf. Interval]
d_orig1	.4996055	.1064219	4.69	0.000	.2910223 .7081886
d_orig2	1.314941	.1028388	12.79	0.000	1.11338 1.516501
d_orig11	1.5255	.4501617	3.39	0.001	.6431994 2.407801
d_orig30	1.089316	.3195964	3.41	0.001	.4629183 1.715713
d_prov9	.8124458	.2246794	3.62	0.000	.3720823 1.252809
d_prov12	1.519345	.2168293	7.01	0.000	1.094367 1.944322
d_prov15	1.159213	.6130512	1.89	0.059	-.0423456 2.360771
d_prov16	1.036816	.4474472	2.32	0.020	.1598354 1.913796
d_prov21	1.74574	.5618578	3.11	0.002	.6445188 2.846961
d_cap7	.0035779	.0959656	0.04	0.970	-.1845111 .191667
qm	1.31e-07	5.89e-08	2.23	0.026	1.59e-08 2.47e-07
is_d	.3836458	.0536207	7.15	0.000	.2785511 .4887405

In grassetto sono stati riportati i coefficienti positivi, interpretabili anche come l'effetto marginale. Il segno dei coefficienti è importante per studiare come varia la probabilità al variare dei regressori. Qualche utile informazione è possibile trarla in termini di statica comparata studiando il segno delle derivali parziali prime e seconde.

Ad esempio è possibile calcolare la differenza di impatto sulla probabilità che esercita la variabile origine in relazione alla provenienza della merce¹⁴.

In sintesi, si studia il verso dell'impatto che ha una variabile sulla probabilità che un evento sia frode. L'attenzione agli effetti marginali, pur limitata al loro segno, è il modo più evidente per evidenziare in quale direzione si orienta il rischio prescindendo dalla modalità con la quale si manifesta la frode e del danno che essa comporta.

¹⁴ Per esempio, se $x_1 = d_orig1$ e $x_2 = d_prov1$, allora la seguente derivata parziale $\frac{\partial^2 p(1|x)}{\partial x_1 \partial x_2}$ misura la variazione delle probabilità.

Ulteriori considerazioni si traggono dall' "odds ratio" che misura il rapporto tra due probabilità. Nel caso di specie sia P la probabilità associata all'evento frodante, l'odds ratio è definito come $\frac{P}{1-P}$.

Esso è pari ad 1 se le due probabilità sono uguali fra loro; diversamente, valori non unitari, attestano il predominio di una probabilità sull'altra. Anche per questo coefficiente è possibile studiarne la variazione.

Di seguito si riportano i più significativi (Tabella 2).

Tabella 2

Odds of: 1 vs 0						
esito	b	z	P> z	OR	OR/StdX	SDofX
d_regis3	-2.15284	-9.819	0.000	0.1162	0.5769	0.2555
D_orig1	0.49961	4.695	0.000	1.6481	1.2761	0.4879
D_orig2	1.31494	12.786	0.000	3.7245	1.5424	0.3295
d_orig3	-0.92859	-4.479	0.000	0.3951	0.8161	0.2188
d_orig11	1.52550	3.389	0.001	4.5974	1.1904	0.1142
d_orig14	-1.83914	-3.053	0.002	0.1590	0.8304	0.1010
d_orig17	-0.86568	-3.104	0.002	0.4208	0.9186	0.0981
d_orig30	1.08932	3.408	0.001	2.9722	1.0645	0.0573
d_prov1	-0.35621	-3.077	0.002	0.7003	0.8491	0.4594
d_prov2	-0.34631	-3.445	0.001	0.7073	0.8775	0.3773
d_prov3	-1.24847	-10.214	0.000	0.2869	0.6941	0.2924
d_prov5	-0.38860	-2.955	0.003	0.6780	0.9188	0.2181
d_prov9	0.81245	3.616	0.000	2.2534	1.1185	0.1378
d_prov11	-1.80486	-3.632	0.000	0.1645	0.8091	0.1173
d_prov12	1.51934	7.007	0.000	4.5692	1.1777	0.1076
d_prov14	-1.26428	-3.047	0.002	0.2824	0.8765	0.1043
d_prov21	1.74574	3.107	0.002	5.7301	1.1470	0.0786
d_cap3	-0.24889	-2.694	0.007	0.7797	0.9402	0.2479
d_cap4	-0.53444	-5.525	0.000	0.5860	0.8788	0.2417
d_cap5	-0.48191	-4.501	0.000	0.6176	0.9047	0.2079
d_cap8	-1.26956	-6.113	0.000	0.2810	0.7841	0.1915
d_cap12	-0.37704	-2.712	0.007	0.6859	0.9387	0.1677
d_cap13	-1.75070	-5.682	0.000	0.1737	0.7689	0.1501
d_cap15	-0.51912	-2.782	0.005	0.5950	0.9278	0.1443
d_cap22	-0.96767	-3.391	0.001	0.3800	0.8995	0.1095
d_cap25	-1.34245	-2.944	0.003	0.2612	0.8807	0.0946
d_cap26	-0.82129	-2.768	0.006	0.4399	0.9254	0.0944
d_cap42	-2.71909	-2.709	0.007	0.0659	0.8388	0.0646
vs	-0.00000	-2.493	0.013	1.0000	0.3432950356.2446	
qm	0.00000	2.231	0.026	1.0000	1.2057 1.4235e+06	
Is_d	0.38365	7.155	0.000	1.4676	1.2114	0.4998

Il secondo logit studia la probabilità dell'evento "non controllo" utilizzando le stesse variabili definite in precedenza.

Il meccanismo è sempre il medesimo: dopo la specificazione delle dummy, si deve osservare il segno dei coefficienti e le variazioni degli odds. Di seguito si riportano i coefficienti delle variabili con significatività al 10% (Tabella 3).

Tabella 3

ncont	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
col	.0000272	6.17e-06	4.41	0.000	.0000151	.0000393
is_d	.1443582	.0124467	11.60	0.000	.1199631	.1687533
d_regis1	1.255164	.0442376	28.37	0.000	1.16846	1.341869
d_regis2	1.039991	.0470974	22.08	0.000	.9476817	1.1323
d_regis3	.8889319	.0476097	18.67	0.000	.7956186	.9822451
d_regis4	.3885748	.0606891	6.40	0.000	.2696264	.5075232
d_regis5	.4556489	.0697568	6.53	0.000	.3189281	.5923698
d_regis6	.9016476	.0720645	12.51	0.000	.7604038	1.042891
d_regis7	1.253379	.1017276	12.32	0.000	1.053997	1.452761
d_orig2	.7127751	.0322749	22.08	0.000	.6495175	.7760327
d_orig9	.5211734	.1103741	4.72	0.000	.3048441	.7375028
d_orig13	.420997	.1266672	3.32	0.001	.1727339	.6692602
d_orig16	.5490152	.1425339	3.85	0.000	.2696538	.8283765
d_orig19	.4475023	.1057203	4.23	0.000	.2402944	.6547102
d_orig23	.5513681	.1594394	3.46	0.001	.2388726	.8638636
d_orig28	1.310007	.1374055	9.53	0.000	1.040697	1.579317
d_prov3	.2066502	.0388658	5.32	0.000	.1304746	.2828259
d_prov5	.1444251	.0442584	3.26	0.001	.0576802	.23117
d_prov7	.9882265	.0514293	19.22	0.000	.8874269	1.089026
d_prov9	.6306858	.0659001	9.57	0.000	.5015241	.7598476
d_prov14	1.060645	.0996809	10.64	0.000	.8652739	1.256016
d_prov15	.5542508	.1190118	4.66	0.000	.320992	.7875096
d_prov16	.9603538	.0882355	10.88	0.000	.7874154	1.133292
d_prov24	.2711847	.0965795	2.81	0.005	.0818923	.4604771
d_prov26	.5652251	.0894507	6.32	0.000	.3899049	.7405453
d_prov29	.5040238	.0853175	5.91	0.000	.3368046	.671243
d_prov30	.6778217	.0971246	6.98	0.000	.4874609	.8681825
rot	.3354881	.0258554	12.98	0.000	.2848123	.3861638
qm	7.85e-09	4.63e-09	1.70	0.090	-1.23e-09	1.69e-08
_cons	2.537733	1.027175	2.47	0.013	.5245062	4.55096

Ricordando che i coefficienti rappresentano l'effetto al margine (positivo) delle variabili, l'obiettivo è stato quello di studiare la probabilità di “non controllo”, poiché essa è sicuramente incorporata nella strategia del frodante. Valgono in questa sede tutte le ipotesi precedentemente fatte, ivi comprese la considerazioni sul danno potenziale.

Detto ciò si riporta la tabella che evidenzia l'impatto sull'odds ratio di ogni variabile (Tabella 4):

Tabella 4

Odds of: 1 vs 0						
ncont	b	z	P> z	OR	e^bStdX	SDofX
col	0.00003	4.412	0.000	1.0000	1.0187	678.3356
is_c	-0.45826	-36.787	0.000	0.6324	0.8235	0.4237
is_d	0.14436	11.598	0.000	1.1553	1.0748	0.4998
d_regis1	1.25516	28.373	0.000	3.5084	1.5677	0.3582
d_regis2	1.03999	22.082	0.000	2.8292	1.2839	0.2403
d_regis3	0.88893	18.671	0.000	2.4325	1.2252	0.2285
d_regis4	0.38857	6.403	0.000	1.4749	1.0388	0.0979
d_regis5	0.45565	6.532	0.000	1.5772	1.0375	0.0808
d_regis6	0.90165	12.512	0.000	2.4637	1.0677	0.0727
d_regis7	1.25338	12.321	0.000	3.5022	1.0694	0.0535
d_orig1	-0.54079	-15.431	0.000	0.5823	0.7835	0.4510
d_orig2	0.71278	22.085	0.000	2.0396	1.3261	0.3960
d_orig4	-0.12480	-3.331	0.001	0.8827	0.9746	0.2060
d_orig5	-0.80454	-20.066	0.000	0.4473	0.8500	0.2020
d_orig8	-0.59953	-10.695	0.000	0.5491	0.9036	0.1691
d_orig9	0.52117	4.722	0.000	1.6840	1.0681	0.1264
d_orig10	0.29967	5.126	0.000	1.3494	1.0377	0.1234
d_orig11	-0.49698	-5.960	0.000	0.6084	0.9416	0.1211
d_orig13	0.42100	3.324	0.001	1.5235	1.0505	0.1170
d_orig14	-0.45803	-5.453	0.000	0.6325	0.9508	0.1102
d_orig15	-0.43606	-3.984	0.000	0.6466	0.9565	0.1020
d_orig16	0.54902	3.852	0.000	1.7315	1.0547	0.0970
d_orig17	-0.55570	-8.531	0.000	0.5737	0.9487	0.0948
d_orig18	-0.43355	-5.500	0.000	0.6482	0.9615	0.0905
d_orig19	0.44750	4.233	0.000	1.5644	1.0381	0.0835
d_orig20	-3.10054	-32.287	0.000	0.0450	0.7775	0.0811
d_orig22	-0.21984	-2.872	0.004	0.8027	0.9831	0.0777
d_orig23	0.55137	3.458	0.001	1.7356	1.0436	0.0774

d_orig24	-0.66002	-3.700	0.000	0.5168	0.9539	0.0715
d_orig28	1.31001	9.534	0.000	3.7062	1.0911	0.0666
d_prov2	-0.18945	-5.375	0.000	0.8274	0.9299	0.3838
d_prov3	0.20665	5.317	0.000	1.2296	1.0678	0.3172
d_prov4	-1.70543	-41.304	0.000	0.1817	0.6243	0.2762
d_prov5	0.14443	3.263	0.001	1.1554	1.0368	0.2504
d_prov7	0.98823	19.215	0.000	2.6865	1.1898	0.1759
d_prov9	0.63069	9.570	0.000	1.8789	1.1016	0.1534
d_prov11	-0.44035	-3.562	0.000	0.6438	0.9485	0.1200
d_prov13	0.67595	7.211	0.000	1.9659	1.0791	0.1127
d_prov14	1.06064	10.640	0.000	2.8882	1.1170	0.1043
d_prov15	0.55425	4.657	0.000	1.7406	1.0547	0.0961
d_prov16	0.96035	10.884	0.000	2.6126	1.0957	0.0952
d_prov17	-0.50295	-3.417	0.001	0.6047	0.9559	0.0897
d_prov18	-0.55164	-5.379	0.000	0.5760	0.9517	0.0897
d_prov21	-1.04027	-10.497	0.000	0.3534	0.9219	0.0782
d_prov24	0.27118	2.808	0.005	1.3115	1.0199	0.0725
d_prov26	0.56523	6.319	0.000	1.7598	1.0389	0.0674
d_prov28	-0.49566	-3.648	0.000	0.6092	0.9681	0.0654
d_prov29	0.50402	5.908	0.000	1.6554	1.0308	0.0601
d_prov30	0.67782	6.979	0.000	1.9696	1.0403	0.0583
d_cap49	-2.83479	-2.774	0.006	0.0587	0.8492	0.0576
d_cap62	-2.73927	-2.676	0.007	0.0646	0.8951	0.0404
d_cap76	-4.07057	-3.966	0.000	0.0171	0.8904	0.0285
d_cap77	-3.71110	-3.617	0.000	0.0245	0.9001	0.0284
d_cap83	-5.94500	-5.726	0.000	0.0026	0.8691	0.0236
d_cap86	-3.16577	-3.073	0.002	0.0422	0.9337	0.0217
d_cap95	-4.44767	-4.179	0.000	0.0117	0.9521	0.0110
d_cap96	-3.09287	-2.904	0.004	0.0454	0.9695	0.0100
rot	0.33549	12.976	0.000	1.3986	1.1590	0.4399
vs	-0.00000	-2.487	0.013	1.0000	0.9788951857	0.0255
qm	0.00000	1.695	0.090	1.0000	1.0164	2.0697e+06
dazio	-0.00001	-6.127	0.000	1.0000	0.9780	4001.9540

Infine si riportano i risultati di una stima di un modello logit in cui si studia la probabilità di non risultare positivo a controllo. Ciò significa che si intende studiare tutti quegli eventi per cui è stato disposto un controllo ma non è stata rilevata la presenza di comportamento frodante.

Per questo logit, a differenza dei precedenti, le variabili scelte sono di tipo contestuale, più precisamente si è definita una dummy per ogni circoscrizione doganale.

Per ogni evento è possibile una sola circoscrizione doganale di registrazione, e quindi è valorizzata una sola variabile dicotomica: le operazioni possono essere partizionate secondo la dogana in cui vengono registrate.

E' evidente che si potrebbe procedere all'acquisizione di altre informazioni, quali la direzione regionale cui appartiene la circoscrizione oppure altri indicatori specifici del contesto in cui l'evento si è manifestato (presenza di scali nella rotta seguita, o tempi medi di sosta, ecc...).

Ma ancora di più, in una ottica e in un ambito Europei, si potrebbero studiare gli effetti dei paesi di ingresso delle merci, considerando anche il possibile regime comunitario a cui sono sottoposte le merci.

Pertanto è sempre possibile una descrizione più dettagliata dell'ambiente in cui opera la dogana: ma il nostro obiettivo primario rimane quello di indicare l'opportunità di definire quantitativamente tale ambiente.

Tabella 5

C0	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
d_circ1	-3.263863	.6067679	-5.38	0.000	-4.453106	-2.07462
d_circ2	-3.618953	.6118574	-5.91	0.000	-4.818172	-2.419735
d_circ3	-4.095922	.6165122	-6.64	0.000	-5.304264	-2.887581
d_circ4	-2.556286	.6081332	-4.20	0.000	-3.748205	-1.364367
d_circ5	-3.315763	.6136054	-5.40	0.000	-4.518407	-2.113118
d_circ6	-3.001466	.6109742	-4.91	0.000	-4.198953	-1.803978
d_circ7	-4.132157	.6257209	-6.60	0.000	-5.358548	-2.905767
d_circ8	-3.489874	.618363	-5.64	0.000	-4.701843	-2.277905
d_circ9	-3.31459	.6366469	-5.21	0.000	-4.562395	-2.066785
d_circ10	-2.744889	.6146307	-4.47	0.000	-3.949543	-1.540235
d_circ11	-3.777876	.6269062	-6.03	0.000	-5.00659	-2.549162
d_circ12	-3.255267	.611379	-5.32	0.000	-4.453548	-2.056986
d_circ13	-4.759239	.6624424	-7.18	0.000	-6.057603	-3.460876
d_circ14	-4.864453	.6665598	-7.30	0.000	-6.170886	-3.55802
d_circ16	-3.479841	.6432743	-5.41	0.000	-4.740635	-2.219046
d_circ17	-4.478635	.6626221	-6.76	0.000	-5.77735	-3.179919
d_circ18	-4.118061	.6592688	-6.25	0.000	-5.410204	-2.825918
d_circ19	-4.435567	.8394875	-5.28	0.000	-6.080933	-2.790202

d_circ20	-3.74526	.6597149	-5.68	0.000	-5.038277	-2.452242
d_circ21	-2.34632	.6189198	-3.79	0.000	-3.559381	-1.13326
d_circ22	-2.96448	.6221639	-4.76	0.000	-4.183899	-1.745061
d_circ23	-3.829883	.6446134	-5.94	0.000	-5.093302	-2.566464
d_circ24	-2.961141	.6478983	-4.57	0.000	-4.230998	-1.691283
d_circ27	-3.449988	.7045769	-4.90	0.000	-4.830933	-2.069042
d_circ28	-4.46935	.7024136	-6.36	0.000	-5.846056	-3.092645
d_circ29	-2.569248	.6405803	-4.01	0.000	-3.824762	-1.313733
d_circ30	-4.532599	.7873325	-5.76	0.000	-6.075743	-2.989456
d_circ31	-3.118304	.6393667	-4.88	0.000	-4.37144	-1.865168
d_circ32	-3.7458	.7166397	-5.23	0.000	-5.150388	-2.341212
d_circ33	-5.049856	1.172331	-4.31	0.000	-7.347582	-2.75213
d_circ34	-3.58827	.7171214	-5.00	0.000	-4.993802	-2.182737
d_circ35	-4.628887	.8389912	-5.52	0.000	-6.273279	-2.984494
d_circ36	-3.478158	.7345108	-4.74	0.000	-4.917773	-2.038544
d_circ37	-4.094345	.7884584	-5.19	0.000	-5.639695	-2.548994
d_circ38	-4.486387	.8393477	-5.35	0.000	-6.131478	-2.841295
d_circ39	-2.600217	.69834	-3.72	0.000	-3.968939	-1.231496
d_circ40	-3.524678	.7173384	-4.91	0.000	-4.930636	-2.118721
d_circ41	-4.725616	.9338018	-5.06	0.000	-6.555834	-2.895398
d_circ43	-3.455686	.7577878	-4.56	0.000	-4.940922	-1.970449
d_circ44	-4.189655	1.176796	-3.56	0.000	-6.496132	-1.883177
d_circ46	-2.899851	.7203872	-4.03	0.000	-4.311784	-1.487917
d_circ47	-3.161247	.8467285	-3.73	0.000	-4.820804	-1.501689
d_circ48	-3.107436	.6800396	-4.57	0.000	-4.440289	-1.774583
d_circ49	-4.659658	.933996	-4.99	0.000	-6.490257	-2.82906
d_circ52	-2.580217	.7647539	-3.37	0.001	-4.079107	-1.081327
d_circ54	-2.95491	.9475846	-3.12	0.002	-4.812142	-1.097678
d_circ56	-4.458988	1.17497	-3.79	0.000	-6.761887	-2.156088
d_circ57	-3.663562	.9391759	-3.90	0.000	-5.504312	-1.822811
d_circ59	-4.543295	.8391992	-5.41	0.000	-6.188095	-2.898495
d_circ60	-3.226844	.8460928	-3.81	0.000	-4.885155	-1.568533
d_circ62	-3.152736	.9446204	-3.34	0.001	-5.004158	-1.301314
d_circ63	-4.171306	1.176939	-3.54	0.000	-6.478063	-1.864548
d_circ64	-4.207673	1.176658	-3.58	0.000	-6.51388	-1.901466
_cons	.1823216	.6055301	0.30	0.763	-1.004496	1.369139

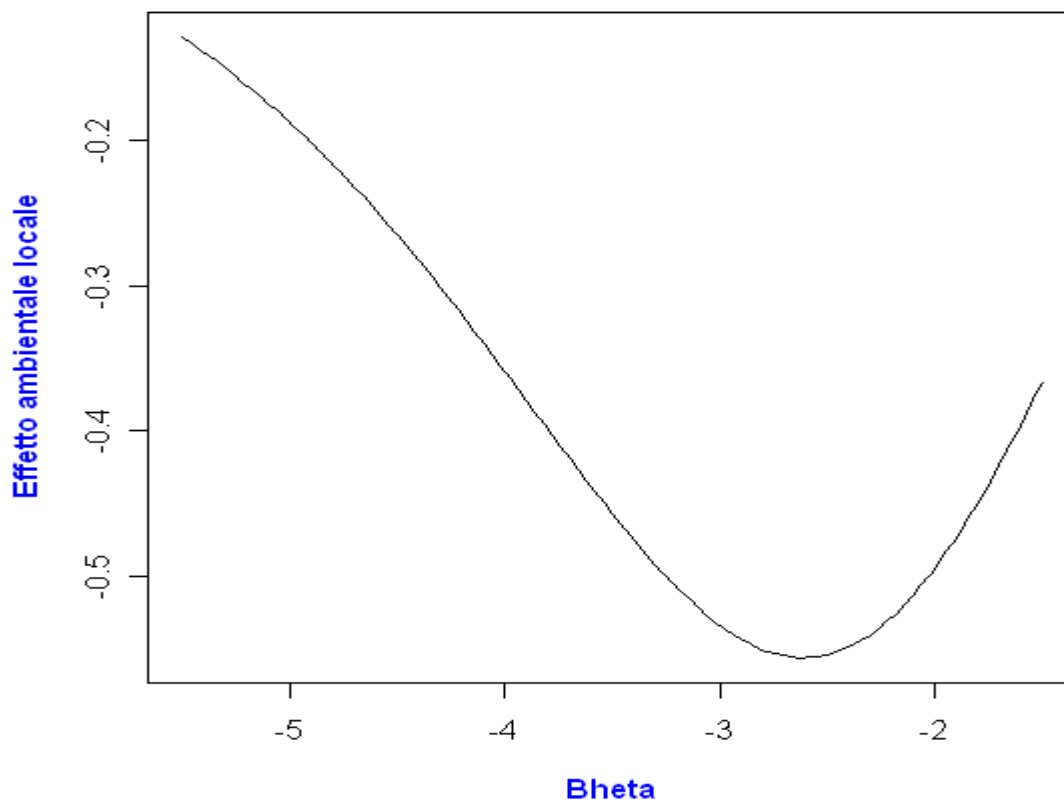
I dati in questa tabella sono stime di sole variabili dicotomiche relative ad alcune circoscrizioni doganali. La nomenclatura scelta è identica a quella usata per gli altri modelli e l'indice numerico è calcolato dalla circoscrizione più frequente fino a quella meno frequente sui dati simulati per un generico mese T. Sono riportate le circoscrizioni più significative.

Sottolineiamo che in questo caso è immediata la lettura degli effetti marginali: infatti, ricordando che questi differiscono da punto a punto, quindi da circoscrizione a circoscrizione, e tenendo conto che i regressori sono tutte variabili dummy mutuamente esclusive (partizione della popolazione osservata), l'effetto diventa:

$$f(\beta_i + \beta_0)\beta_i = \frac{e^{\beta_i + \beta_0}}{(1 + e^{\beta_i + \beta_0})^2} \beta_i.$$

Immaginando che β_i vari in modo continuo è possibile dare una rappresentazione grafica dell'effetto marginale (Figura 3).

Figura 3



Si nota il minimo effetto in corrispondenza del valore -2,6. Quindi in corrispondenza di un tale valore del parametro (ad esempio per la circoscrizione 39) si ha un impatto molto negativo sulla probabilità di non avere esito di frode al controllo. Equivalentemente, nella formulazione duale del problema, possiamo dire che la circoscrizione 39 ha un effetto positivo sulla probabilità che il controllo sia ben mirato. Ovviamente le conclusioni sui valori estremi del grafico sono di natura opposta: i controlli nelle circoscrizioni che abbiano il coefficiente sugli estremi del dominio sono poco efficaci.

Ricordiamo che sono state scelte variabili arbitrarie e che, probabilmente, si potrebbe giungere ad alte conclusioni con una descrizione diversa dell'ambiente (altre variabili): ribadiamo che il nostro interesse primario è quello di indicare soltanto una possibile rappresentazione quantitativa del contesto operativo della dogana.

7 Conclusioni

La rotta tracciata è stata quella di individuare le variabili che hanno un rilevante impatto sulla probabilità che un evento sia frodante, nonché le variabili significative che determinano la probabilità del non controllo. L'applicazione dei modelli logistici è sembrata la più consona in termini di informazione che se ne ricava.

Capire cosa favorisce la frode è una questione che non può essere svincolata dal contesto ambientale. In questo senso, la matrice di relazione fornisce utili indicazioni per la selezione delle variabili significative. Dall'analisi emerge comunque un risultato sicuro: esistono aree geografico-territoriali ad elevato rischio. L'idea però è quella di coglierne l'interazione (si pensi al binomio origine-provenienza della merce). Come? Attraverso variabili soggettive comportamentali. Molto c'è da fare in questa direzione. Allo stato attuale non v'è altro da aggiungere. Solo una riflessione rimane: solo il tempo dirà se la rotta tracciata è giusta o è almeno una buona rotta.

8 Bibliografia

Paolo Baldi, (2006), “Calcolo delle Probabilità”, *McGraw-Hill*.

Giorgio Koch, (1998), “La matematica del probabile”, *Aracne*.

Jeffrey M. Wooldridge, (2001), “Econometric Analysis of Cross Section and Panel Data”, *MIT Press*.

Damodar Gujarati, (2004), “Basic Econometrics”, *McGraw-Hill*.

Gianluca Dari, (2002), “Frodi doganali e Profili di Rischio: un approccio matematico-probabilistico”, *Agenzia delle Dogane*.

Davide Russo, (2006), “Sistema Aracne: dalla relazione all’informazione”, *Agenzia delle Dogane*.

Christopher Baum, (2006), “Introduction to STATA”, *Boston College FMRC*.

IBM, (2001), “SQL Reference”, <http://www.ibm.com>.

Akerloff G., (1970), “The market for lemons: quality uncertainty and the market mechanism”, *Quarterly Journal of economics* 89 pg. 488 – 500

Fudenberg D and J. Tirole, (1992), “Game Theory”, *Cambridge, Mass: MIT Press*.

Green J.R and J. Laffont, (1979), “Incentives in Public Decision making”, *Amsterdam: North-Holland*