



SDA Bocconi School of Management
Personalized Pricing in Insurance: The Promises and Perils of
Big Data

Il punto di vista del regolatore

Stefano De Polis
Segretario Generale IVASS

Milano, 17 marzo 2023

Ringrazio la SDA Bocconi e il Prof Rossi per l'invito ad intervenire a un dibattito su un tema di concreta rilevanza per chi, come noi, ha il compito di contribuire ad aumentare l'efficienza dei mercati, l'equità dei contratti e al contempo assicurare la solidità patrimoniale delle compagnie assicurative e la stabilità finanziaria del sistema nel contesto dinamico della competizione.

Partiamo dalla prima delle due parole chiave dell'evento di oggi: le promesse dei *big data* per il comparto assicurativo. I *big data* hanno un'accezione più estesa rispetto agli *advanced analytics* di grandi fonti di dati già disponibili ma non utilizzate per carenze di capacità tecnologiche e metodologiche; i *big data* si riferiscono ad informazioni di qualunque tipo acquisite con qualsiasi strumento, da quelli tradizionali a quelli più innovativi, quali ad esempio i social network. Il concetto di *big data* si basa infatti sulle cosiddette "3V" che si riferiscono a volume, velocità e varietà, ad intendere che si parla di una gran quantità di dati, strutturati e non, generati con alta frequenza e con un formato sovente non strutturato.

Un esempio del cambiamento ma anche della complessità che introduce l'utilizzo di *big data* viene da alcune esperienze avanzate del settore agricolo. Oggi tra le nuove applicazioni della tecnologia figurano l'utilizzo congiunto della geo-localizzazione dei terreni; l'acquisizione nel continuo di dati meteorologici; la rilevazione di dati fenologici mediante sensori digitali: dai pochi dati sinora utilizzati per stimare il rischio in agricoltura ad una matrice tridimensionale di informazioni in tempo reale!

Mi soffermerò su tre manifestazioni di queste potenzialità e cercherò di illustrare come, nel mondo assicurativo di oggi, esse si siano concretamente manifestate.

La prima grande promessa è quella di ridurre se non eliminare le inefficienze generate dalle asimmetrie informative. Nei mercati in cui l'assicurazione è obbligatoria, ad esempio la r.c. auto, è verosimile che le poche variabili tradizionalmente utilizzate per definire il premio (ad esempio: età, provincia di residenza, caratteristiche del veicolo) non siano sufficienti a misurare il rischio individuale. La conseguenza in questo caso è una redistribuzione dei premi tra assicurati che può non essere

desiderabile. Nei mercati in cui l'assicurazione non è obbligatoria (la r.c., le assicurazioni contro i rischi catastrofali) le conseguenze delle asimmetrie informative, in particolare l'*adverse selection*, possono essere tali da non permettere lo sviluppo del mercato lasciando prive di copertura ampie fasce della popolazione potenzialmente interessate a stipulare contratti con premi "attuarialmente equi"¹.

Nel momento in cui gli assicuratori dispongono di molte nuove e ricche variabili sul rischio individuale - nel caso della assicurazione auto la velocità, la distanza percorsa e molteplici ulteriori informazioni sugli stili di guida - la chimera del premio attuarialmente equo, un paio di decenni fa pura teoria, si avvicina.

La seconda promessa: l'efficienza dei risarcimenti. I *big data* hanno rivelato uno straordinario potenziale al fine di valutare l'entità del danno in tempo reale: nella r.c. auto attraverso tecniche di *machine learning* (ML) con l'utilizzo di foto scattate dall'assicurato e informazioni sulle caratteristiche del veicolo è oggi possibile ricevere un risarcimento "giusto" a poche ore dal verificarsi di un sinistro, la *real time indemnification*, riducendo in modo drastico i tempi di liquidazione. Guardando oltre la r.c. auto, i dati utilizzati per la telemedicina sono potenzialmente utili per ridurre i costi delle cure mediche, per decidere diagnosi e terapie e per mitigare i rischi della non autosufficienza (LTC). Una maggiore accuratezza nella valutazione del danno riduce il *nitpicking*, la pratica di contestare l'entità del danno avvalendosi delle circostanze del sinistro e delle rivalse specificate nei contratti². Al contempo, le nuove informazioni sulle circostanze del sinistro mettono in condizione le compagnie di contrastare meglio le frodi. Dunque, l'utilizzo dei *big data* dopo l'accadimento di un sinistro può creare maggiore fiducia tra gli attori del mercato.

La terza promessa, forse la più avveniristica: l'assicurazione degli stili di vita e il *bundling* dei rischi. La possibilità di osservare le abitudini di vita di ciascuno, connettendo informazioni rilevate da più fonti e *device* (il tragitto percorso per recarsi al lavoro, l'utilizzo del *car sharing*, la frequenza dei viaggi in aereo, il tempo dedicato

¹ La sottoassicurazione (*under insurance*) generata dalle asimmetrie informative è tale giustificare misure di policy come l'obbligatorietà dell'assicurazione e/o i sussidi per alcuni segmenti di mercato. Si veda Einav e Finkelstein (2021).

² Cfr. Bourgeon e Picard (2014).

all'esercizio fisico) permetterebbe di offrire un pacchetto assicurativo che copra in modo proporzionale ciascun rischio. Il *bundling* di rischi eterogenei può favorire la ricerca di soluzioni più convenienti in grado di coprire anche (micro) eventi spesso non singolarmente assicurabili.

Questa breve panoramica dei potenziali benefici del ricorso a dati aggiuntivi e modelli di analisi avanzati è utile a dare la cifra del cambio di passo rispetto al modello tradizionale di *business*: ci confrontiamo oggi con *insurance 2.0*.

Osserviamo, a questo punto, che l'assicurazione assume (soprattutto in caso di obbligatorietà) anche una valenza "solidaristica": applicare ai soggetti poco rischiosi premi che eccedono il costo medio (atteso) permette di rendere accessibile l'assicurazione per i segmenti più rischiosi, che in presenza di un *pricing* altamente personalizzato sarebbero di fatto tagliati fuori dal mercato.

In definitiva: in assenza di *pricing* personalizzato i soggetti meno rischiosi non sono potenzialmente disposti ad assicurarsi perché il premio che dovrebbero pagare eccede il loro rischio. Quando invece i *big data* vengono utilizzati per personalizzare il premio il pericolo della non assicurazione riguarda i soggetti più rischiosi. In ambo i casi la sotto-assicurazione è un'ipotesi reale, e il bilanciamento tra due esigenze di fatto contrapposte - solidarietà vs *pricing* "efficiente" - appare complesso.

Veniamo allora alla seconda parola chiave del nostro confronto: i pericoli dell'utilizzo dei *big data*. Pensiamo, in modo non esaustivo, a quattro aspetti piuttosto controversi: la trasparenza dei contratti, gli effetti anti-competitivi (in particolare i fenomeni di *lock-in*), lo stravolgimento delle caratteristiche dell'auto-selezione anche (ma non solo) in funzione del ruolo che può assumere la *privacy*, e la crescente complessità dell'attività di vigilanza.

Cominciamo dalla crescente importanza della trasparenza rispetto ai nuovi contratti telematici, che appaiono assai più complessi di quelli tradizionali. Si pensi ai contratti r.c. auto con scatola nera: la distanza percorsa può determinare, anche in funzione dell'ora del giorno/notte e della tipologia di strade percorse, uno sconto sul

premio. Un'eccessiva complessità del *pricing* potrebbe³ indurre i consumatori a focalizzarsi solo su alcuni degli aspetti "salienti" del prezzo, ad esempio lo sconto dopo l'installazione della scatola nera. Esiste dunque un *trade-off* tra la comprensione della relazione tra i (molteplici) comportamenti monitorati dalle nuove tecnologie e il premio e un *pricing* particolarmente personalizzato: in presenza di *bias* cognitivi non necessariamente un contratto "teoricamente equo" dal punto di vista attuariale seleziona e prezza il rischio in modo efficiente⁴. Più che in passato la trasparenza dei contratti è un elemento chiave per fare sì che i consumatori possano selezionare l'opzione più confacente alle proprie esigenze/preferenze.

Il secondo pericolo è quello del *lock-in* dei consumatori: poiché i contratti basati sui *big data* legano il premio ai comportamenti osservati in passato, cambiare compagnia senza potersi portare dietro la storia dei comportamenti implica una perdita di informazioni che il nuovo assicuratore non può utilizzare per prezzare il rischio. Come conseguenza, l'incentivo a cambiare compagnia si riduce e la competizione per attrarre nuovi clienti si attenua. Evidenze da una recente indagine IVASS indicano che le compagnie iniziano a utilizzare algoritmi di ML in grado di per identificare i clienti con la maggiore probabilità di abbandono, con l'obiettivo di modificare l'offerta per trattenere i clienti maggiormente mobili penalizzando di fatto quelli più fedeli

Il terzo tema su cui spesso ci interroghiamo è come i nuovi contratti basati sui *big data* cambino le caratteristiche della selezione dei rischi. Il modello base di *insurance* suggerisce che in presenza di *adverse selection* i consumatori che scelgono di assicurarsi per un dato livello di premio sono mediamente più rischiosi di quelli che sono disposti a farlo per un livello di premio più basso. L'utilizzo dei *big data* cambia questo meccanismo attraverso due canali. Il primo: atteso che la decisione di essere monitorati per ricevere un prezzo personalizzato è volontaria, questa possibilità risente anche del peso che i consumatori attribuiscono alla propria *privacy*: la "nuova" domanda di assicurazione introduce un elemento di ulteriore complessità rispetto al passato. La seconda questione è forse ancora più fondamentale: è plausibile che i *big data* sovvertano il vantaggio informativo a favore dell'assicuratore. In altre parole, è

³ Sul tema si veda Miravete (2013) e Sydnor, Bhargava, Loewenstein (2017).

⁴ Cfr. Handel and Jonathan T. Kolstad (2015), e Jin and Yu (2021).

ben possibile che l'assicurato conosca il proprio rischio meno dell'assicuratore, la cosiddetta "*inverse selection*"⁵. La teoria suggerisce che questo fenomeno aumenta i profitti delle imprese in presenza di *bias* cognitivi dei consumatori.

Il ruolo dei Big data nell'attività di Vigilanza.

I processi di selezione e tariffazione (o *pricing*) dei rischi e quelli di riservazione e liquidazione dei sinistri rappresentano i presupposti di una sana e prudente gestione di una compagnia. Il Codice delle Assicurazioni Private prevede che le tariffe debbano essere coerenti con le basi statistiche in possesso dell'impresa.

Il processo di tariffazione rappresenta l'attività fondamentale nella definizione di un contratto assicurativo. Definire un premio è il risultato di una combinazione di analisi, esperienza e giudizi attuariali. Come detto, l'utilizzo di *big data* e *advanced analytics* può consentire all'assicuratore di valutare in modo analitico i rischi potenziali; rende il compito del regolatore – la verifica dei processi di *pricing* – assai più complesso. Resta fermo che per migliorare efficacemente le previsioni l'informazione deve però essere rilevante, accurata e appropriata.

Nell'utilizzo dei *big data* le imprese devono essere in grado di garantire requisiti di qualità delle informazioni, di correttezza delle metodologie e replicabilità e tracciatura della metodologia di calcolo, nonché di *governance* del processo, in termini di ruoli e responsabilità, di reportistica e di corretta gestione della *privacy*. Un aspetto rilevante è l'attendibilità della fonte di informazioni: una base dati tariffaria deve essere necessariamente strutturata e statisticamente robusta.

Sotto il profilo metodologico, i dati relativi alle variabili utilizzate per profilare il *target* di mercato ai fini della determinazione delle tariffe dovranno essere effettivamente esplicativi del rischio. Resta poi l'esigenza che le compagnie individuino gruppi omogenei di rischio sufficientemente numerosi al fine di garantire i principi di mutualità peculiari del settore assicurativo nonché la stabilità dei risultati.

⁵ Cfr. Brunnermeir, Lambda e Carlos Segura (2021).

Le metodologie di vigilanza, in assenza delle più tradizionali tabelle di serie storica, dovranno basarsi su metodologie alternative necessariamente meno complesse di quelle aziendali, ma comunque utili per confronti e *back-testing* delle *performance* del processo di tariffazione.

In sintesi, specie in questa prima fase di sviluppo della *insurtech*, approcci regolarmente utilizzati dalla vigilanza sono volti a verificare che le compagnie abbiano strutturati processi di governo e controllo dell'utilizzo dei *big data* e a sviluppare competenze e strumenti per continuare a verificare l'affidabilità dei processi di tariffazione.

In questa prospettiva assumono rilevanza anche tre ulteriori dimensioni organizzative:

- l'adeguatezza dell'architettura informatica e della gestione dei dati delle imprese (con sistemi che dovranno essere in grado di archiviare dati di qualsiasi formato e in modalità *real time - Data Lake*);
- la disponibilità di personale specializzato: il *data scientist* del settore dovrà avere competenze informatiche ma anche di business e statistico-attuariali. La necessità di personale con una formazione specifica si estende naturalmente anche all'Autorità di Vigilanza;
- le politiche di *outsourcing* delle compagnie nel campo informatico e dei fornitori di analisi di *big data*. L'utilizzo di *big data* non può scaturire da *software "black box"* di imprese fornitrici terze che non consentono un'adeguata comprensione del modello di valutazione, in *primis* all'impresa e poi all'occorrenza alla Vigilanza. Per la rilevanza delle 3V dei *big data* è altresì prevedibile un aumento del ricorso delle imprese al *cloud computing*⁶. È imminente l'entrata in vigore della Direttiva DORA (*Digital Operational Resilience Act*) che richiama l'attenzione di tutti gli operatori su questi temi e amplia i poteri di vigilanza delle Autorità sui fornitori di servizi informatici.

⁶ L'EIOPA ha recentemente fornito Orientamenti in materia:
https://www.eiopa.europa.eu/publications/guidelines-outsourcing-cloud-service-providers_en

Promettenti applicazioni dei *big data* riguardano anche le valutazioni delle riserve sinistri. Con i modelli cosiddetti di *claim analytics* viene costantemente seguito e studiato il comportamento di ogni singolo sinistro in base al suo stato iniziale ed alla successiva evoluzione (pagato nell'anno, dopo un anno, dopo 2, ... etc). L'analisi viene effettuata per stimare gli esborsi futuri sinistro per sinistro e quindi l'entità della riserva sinistri nelle sue componenti più analitiche. La tecnica si presta anche ad effettuare *peer review* attraverso una metodologia di stima indipendente delle riserve sinistri. Anche in questa applicazione l'attenzione va posta sulla qualità dei dati, sui processi e sulle metodologie di analisi, aspetti tutti sui quali anche la Vigilanza deve acquisire adeguata esperienza.

#

Concludo tornando su un tema di fondo. Sebbene il ricorso ai *big data* per definire un'offerta assicurativa personalizzata riduca alcune inefficienze, introduce al contempo potenziali distorsioni verso cui il regolatore deve tenere alta la guardia, per assicurare che una competizione sana possa soddisfare le esigenze di protezione dei consumatori. Una condizione necessaria per svolgere questo compito è lo studio rigoroso del *personalized pricing in insurance* anche in relazione alle esigenze di mutualità connaturate al metodo assicurativo: solo in questo modo saremo in grado di produrre quella che qui in Bocconi chiamereste *knowledge that matters*.

Grazie per l'attenzione e buon proseguimento dei lavori.

Riferimenti bibliografici

1. “Selection in Insurance Markets: Theory and Empirics in Pictures”, Liran Einav and Amy Finkelstein
Journal of Economics Perspectives, 25(1), Winter 2011, 115-138
2. “Inverse Selection”, Markus Brunnermeier, Rohit Lamba, and Carlos Segura-Rodriguez
NBER Working Paper n. 25406, 2021
3. “Health Insurance for “Humans”: Information Frictions, Plan Choice, and Consumer Welfare”, Benjamin R. Handel and Jonathan T. Kolstad
American Economic Review 2015, 105(8): 2449–2500
4. “Fraudulent Claims and Nitpicky Insurers”, Jean-Marc Bourgeon and Pierre Picard
American Economic Review 2014, 104(9): 2900-2917
5. “Choosing the Wrong Calling Plan? Ignorance and Learning.” Eugenio Miravete
American Economic Review, 2013, 93, 297-310.
6. Choose to Lose: Health Plan Choices from a Menu with Dominated Options, Justin Sydnor, Saurabh Bhargava and George Loewenstein,
Quarterly Journal of Economics, Vol 132:3, 1319-1372, 2017.
7. How to Prevent Traffic Accidents, Yizhou Jin e Thomas Yu (2021)
Working paper
8. EIOPA: Big data in motor e health insurance
https://www.eiopa.europa.eu/document/download/becbbe3a-ba4c-47b9-870a-63872fef3986_en?filename=Big%20Data%20Analytics%20in%20motor%20and%20health%20insurance%3A%20A%20thematic%20review