

Il motore inferenziale montato sulla banca dati di Quattroruote per prevedere il valore residuo delle auto

Claudio Milani – Editoriale Domus
Guido Cuzzocrea – Nunatac

“CarPreview” è il nuovo Sistema di Previsione dei Valori Residui delle auto messo a punto da Quattroruote. Il sistema permette di prevedere, con ragionevole probabilità, quale sarà il valore di una certa vettura dopo 1, 2, 3, 4, 5 anni dall’acquisto.

Il cuore di “CarPreview” è un sofisticato motore statistico che scandaglia la Banca dati di Quattroruote contenente tutte le quotazioni dell’usato degli ultimi sei anni, allo scopo di trovarvi delle costanti: ad esempio, che una certa marca “tiene” meglio di un’altra, che il gasolio è meglio della benzina, che una utilitaria si svaluta meno di una media e così via. Partendo da queste e numerose altre costanti (che coinvolgono una quarantina di caratteristiche), il sistema classifica tutte le auto in vendita in Italia in 15 diverse tipologie, dalla più alla meno nella tenuta del valore nel tempo, associando a ciascuna di esse una diversa curva di svalutazione. Considerando una specifica versione di un dato modello, la curva di svalutazione della sua tipologia viene corretta per tener conto delle caratteristiche specifiche di quella vettura.

Ogni mese il sistema aggiorna le previsioni, che vengono poi esaminate a fondo dagli analisti di Quattroruote e integrate con le informazioni in loro possesso su eventi che si verificheranno in futuro: per esempio, un restyling, o l’arrivo di un nuovo modello, oppure la prossima adozione di nuove norme sulla sicurezza o la difesa dell’ambiente.

Il sofisticato sistema di previsione è stato realizzato con tecnologia SAS e con il supporto degli specialisti di Nunatac.

1. Introduzione

L’andamento delle quotazioni dell’usato nel mercato delle automobili è un tema complesso influenzato da innumerevoli aspetti; vanno certamente considerati: le caratteristiche tecniche del prodotto auto, la qualità percepita dai consumatori, la notorietà e immagine delle case costruttrici, l’andamento dell’economia di settore e il differenziale tra domanda ed offerta, le regole istituzionali che governano i mercati. L’efficacia di un modello inferenziale nel prevedere, con un margine di errore limitato, un fenomeno di tale complessità dipende da alcuni importanti elementi:

- A) la capacità di definire in modo chiaro gli obiettivi che si intendono perseguire con la costruzione del modello, così come di valutarne l’impatto a livello organizzativo e commerciale;
- B) la qualità e completezza dei dati a disposizione per stimare delle regole empiriche significativamente esplicative;
- C) la metodologia seguita per costruire un modello previsionale basato su:
 - la serie storica dei dati passati,
 - eventuali fonti informative esogene disponibili,
 - le conoscenze degli esperti di settore, opportunamente tradotte in numeri;

- l’orizzonte temporale di previsione e i margini di errore ritenuti accettabili.

2. Presentazione di Editoriale Domus e Nunatac*

2.1. Editoriale Domus

Viene fondata a Milano da Gianni Mazzocchi nel 1929. Giovanna Mazzocchi ha continuato l’opera paterna promuovendo nuove testate, potenziando tutte le pubblicazioni con supporti tecnologici e informatici d’avanguardia, servizi utili ed estendendone la fruibilità a Internet. Oggi, con riviste e CD-ROM seguiti da oltre 10 milioni di lettori, un network di siti verticali consultati da oltre 600.000 visitatori e oltre 320.000 utenti registrati, con più di 300 dipendenti e un fatturato superiore ai 150 miliardi annui, la filosofia di Editoriale Domus è ancora l’attenzione ai bisogni e agli interessi dell’utente, nonché la sua salvaguardia e l’aggiornamento come consumatore.

2.2. Nunatac

NUNATAC è una Società Quality Partner di SAS, composta di un gruppo di consulenti specialisti in informatica, statistica e marketing.

La caratteristica peculiare di NUNATAC è la combinazione di abilità specifiche e un "know-how" consolidato nel campo del Database Marketing.

Nel corso degli ultimi nove anni, Nunatac ha realizzato importanti progetti di data mining nei seguenti settori: Banche, Assicurazioni, V.P.C., Automotive, Telecomunicazioni e Farmaceutico.

3. Obiettivi del Progetto

Lo studio del fenomeno "valori residui dell'auto", a partire dalla serie storica delle quotazioni presenti nella Banca Dati di Quattroruote, ha condotto all'individuazione di molti elementi oggettivi per spiegare l'evoluzione del prezzo di un'auto sul mercato dell'usato.

In questo procedimento, sono state integrate l'esperienza settoriale, in capo agli esperti di Quattroruote e l'analisi dei dati di mercato, attraverso la metodologia statistica realizzata da Nunatac.

Il progetto ha perseguito diverse direttrici di analisi. Tale percorso ha condotto all'identificazione degli elementi fondamentali di spiegazione del fenomeno "valori residui" delle auto ed alla definizione di altrettanti moduli statistici che costituiscono i "mattoni" fondamentali del sistema SPVR (Sistema Previsione dei Valori Residui).

In particolare sono state svolte 5 differenti analisi:

1. Classificazione degli Infocar (codice che identifica ciascuna combinazione di marca, modello, allestimento) in 15 tipologie di Valori Residui;
2. Modellazione di 15 "Curve Maestre";
3. Modelli di scarto sui Parametri Specifici;
4. Aggiustamento della previsione per singolo Infocar;
5. Stima dei pesi medi per i parametri di un'applicazione di simulazione.

3.1 La definizione di 15 tipologie

Dopo aver sintetizzato in un **Fattore di Svalutazione** il fenomeno V.R., abbiamo proceduto alla selezione dell'insieme di caratteristiche di prodotto e di mercato, maggiormente discriminanti i differenti comportamenti osservati in passato.

Le caratteristiche risultate significative sono:

1. Marca
2. Segmento di Mercato
3. Alimentazione
4. Potenza e Cilindrata
5. Carrozzeria
6. Anzianità prodotto
7. Restyling

La classificazione in Modelli non è stata utilizzata nella struttura esplicativa del modello per due sostanziali motivi:

- l'apprezzamento di uno specifico Modello è strettamente correlato a valutazioni soggettive;
- la dinamica di ingresso e uscita dei Modelli dal mercato del nuovo è estremamente rapida e ciò provocherebbe una velocità di obsolescenza eccessiva del sistema, salvo continui ed onerosi interventi sui parametri di impostazione dei modelli.

Il modello interpretativo stimato considera:

1. il fattore di svalutazione ==> come variabile da spiegare,
2. le 7 variabili riclassificate ==> come patrimonio esplicativo,
3. la variabile Anno di Acquisto dell'auto ==> come variabile di classificazione strumentale.

- ☛ Nell'ambito di un Modello Lineare Generalizzato, il "fattore di svalutazione" teorico (stimato dal modello) è la somma algebrica del valore dei parametri (Marca, Cilindrata, ...) associati ad un'auto.

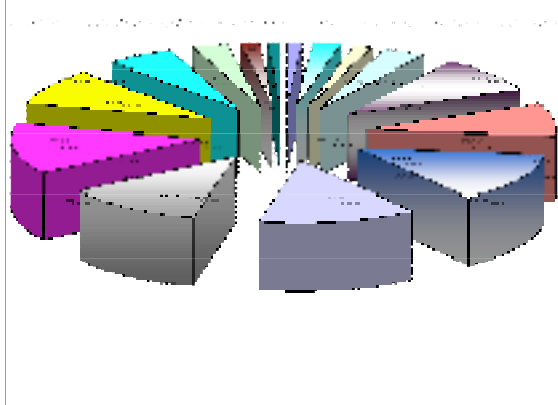
Analizzando i parametri stimati dal modello, valgono alcune importanti considerazioni:

- a parità di altre condizioni, nel corso degli anni (dal 1996 al 1999), il fenomeno V.R. ha subito una dinamica di peggioramento. Come dire che oggi il valore residuo di un'auto si abbatte più rapidamente di ieri;
- le marche sono state riclassificate proprio in base al criterio di performance decrescenti sui V.R.;
- l'anzianità dell'auto sul mercato è un fattore peggiorativo, così come l'aver subito dei restyling. Plausibilmente, questi due elementi costituiscono una misura indiretta del ciclo di vita del prodotto;
- sempre a parità di altre condizioni, le auto a gasolio si deprezzano meno di quelle benzina;
- a parità di altre condizioni, e in particolare a parità di potenza, le cilindrato inferiori si deprezzano meno;
- le berlina risentono maggiormente del processo di invecchiamento, rispetto alle station-wagon, alle monovolume e alle multispazio.

Sulla base del valore stimato dal modello del fattore di svalutazione, abbiamo operato una clusterizzazione di tutte le casistiche considerate: 15 cluster sintetizzano al 99% la variabile di analisi, costituendo un'eccellente sintesi dei risultati del modello interpretativo.

Il grafico che segue riporta la distribuzione percentuale dei 15 cluster (le 15 tipologie di auto stimate) per l'universo di auto con anno di acquisto 1999.

Figura 1: Distribuzione dei Cluster



Ciascun CLUSTER identifica una Tipologia di auto con equiparabili dinamiche di Valore Residuo.

3.2 Modelli generali di previsione per la tipologia

In un secondo passaggio dell'analisi sono stati individuati 15 modelli generali di previsione dei valori residui per l'auto media di ciascuna tipologia (15 modelli) e anche delle regole di aggiornamento "automatico" della curva "maestra" stimata per la tipologia.

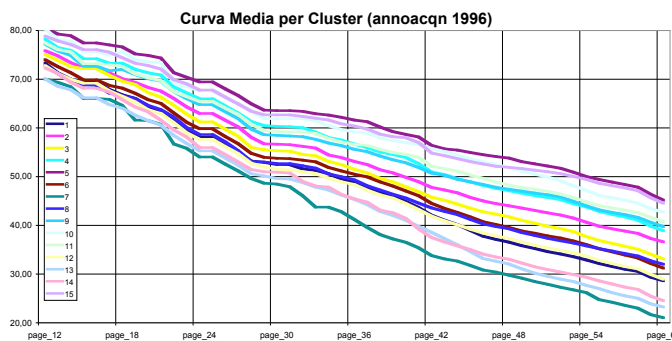
1. Determinazione della "curva media" per tipologia
2. La correzione delle curve sulla base dei dati recenti
3. La Modellazione di 15 "Curve Maestre"
4. Il meccanismo di aggiornamento delle "Curve Maestre" al periodo.

Le curve medie per tipologia sono state ricostruite sui valori osservati di mercato, rilevati attraverso i dati presenti nella Banca Dati di Quattroruote.

Le considerazioni sulle distribuzioni degli Infocar nei 15 cluster sono avvalorate dalle rappresentazioni grafiche delle "15 curve base" nel corso dei successivi anni di acquisto: la classificazione costruita sembra effettivamente rappresentare in modo esaustivo il fenomeno, descrivendo traiettorie di V.R. ben differenziate.

Per ogni cluster, abbiamo quindi ricostruito tante curve (di profondità temporale variabile) quanti sono i diversi anni d'acquisto (nell'attuale base dati, 1996, 1997, 1998, 1999),

Queste quattro curve (per ciascun cluster) devono essere integrate in un'unica **interpolante** che consente di ottenere 15 curve che sintetizzano al meglio (secondo criteri di robustezza statistica) le

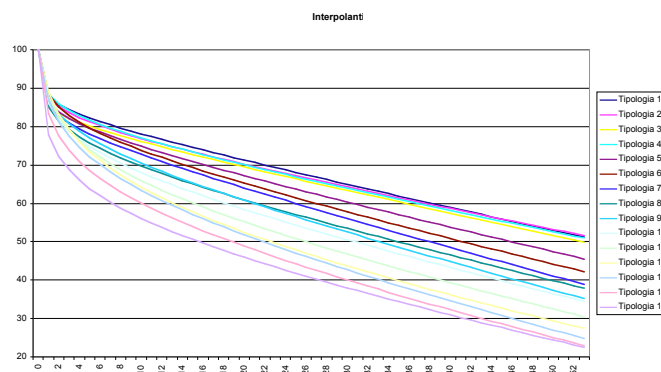


informazioni contenute nella banca dati dei valori residui osservati.

E' possibile, per ciascuna tipologia, individuare **la forma matematica della curva interpolante**: se esiste una funzione matematica che descrive l'andamento delle risultanti (osservate nella storia) e che rispetta alcune intuitive condizioni, allora saremo in grado di utilizzare questa funzione come **curva maestra**, come previsione generale per ogni tipologia.

Il grafico sottostante raffigura le 15 curve maestre individuate:

QuotAuto (sistema di gestione delle quotazioni delle auto usate di Quattroruote) genera mensilmente il nuovo listino dell'usato e, dopo circa 6 mesi dall'introduzione in commercio, anche i modelli nuovi ricevono una valutazione dell'usato.



Nel rispetto della logica secondo cui il modello di previsione dovrà poter disporre delle informazioni di mercato più recenti possibili è stato studiato un meccanismo automatico di correzione della forma delle curve maestre. Questo meccanismo dovrebbe garantire la stabilità delle previsioni generate dal modello di stima, anche in seguito a variazioni nel mercato dell'usato.

3.3 L'assegnazione del peso ai parametri specifici di un Infocar

A partire dalle valutazioni storiche osservate, la curva di svalutazione di ciascun Infocar è messa a confronto

con la curva media storica del proprio cluster di appartenenza.

Le differenze definiscono il **delta** tra curve dei V.R. osservate per ogni Infocar e “curva media” della tipologia cui l’Infocar appartiene. Questa differenza rappresenta la parte di informazione, non catturata dai modelli statistici sulle curve maestre. Per ogni tipologia, un modello statistico stima i pesi dei parametri specifici utilizzando il delta come variabile obiettivo.

Nel disegno generale di SPVR, questo modello statistico consente di passare dal generale, cioè dalle curve maestre, al particolare cioè all’Infocar.

3.4 Correzione di un V.R. di un Infocar in base alle quotazioni più recenti

Con un meccanismo simile a quello di correzione delle curve maestre, anche le quotazioni relative a singoli Infocar, consentono di aggiornare la curva di previsione del singolo Infocar.

Questo meccanismo di aggiornamento è valido unicamente per gli Infocar che hanno quotazioni già presenti in banca dati.

3.5 L’applicazione di simulazione per gli esperti di Quattroruote

Il Sistema di Previsione dei Valori Residui, allo stato attuale, prevede un’interfaccia utente che permette agli analisti di Quattroruote di intervenire sulle previsioni, in base alle informazioni di settore più aggiornate.

E’ in fase di messa a punto una versione di questa applicazione di simulazione per gli operatori del settore (Case Costruttrici, Società di Leasing e Noleggio a lungo termine, Dealers).

5. Conclusioni

I moduli statistici, definiti nelle analisi svolte sino ad ora, costituiscono la “parte intelligente” che produrrà le stime dei valori residui delle auto.

Essi costituiscono comunque, soltanto una delle componenti (fondamentali) del processo informatico che dovrà produrre, in modo automatico, i valori di stima per le quotazioni future delle auto.

Con cadenza mensile, l’aggiornamento delle basi dati di QuotAuto, si integra con i processi di generazione dei valori residui delle auto.

Un modello interpretativo, come quello realizzato, differisce da un motore di previsione proprio perché considera determinante l’aspetto interpretativo e la possibilità di individuare le unità elementari che compongono il fenomeno oggetto di studio.

La differenza tra i due approcci sta nella volontà di rendere trasparenti e comprensibili le caratteristiche del

fenomeno e le relazioni che partecipano alla generazione dei valori previsti. Venendo meno questa volontà, è possibile spostare l’enfasi sul solo grado di precisione delle stime, attraverso l’utilizzo di algoritmi matematici di crescente complessità. Va sottolineato che la precisione delle stime è da intendersi come capacità di riprodurre, in modo più o meno perfetto, una moltitudine di esempi osservati nel passato. Il che non è necessariamente garanzia di capacità di generalizzare, rispetto a nuove casistiche.

Il modello di previsione non è una sfera di cristallo.

Il compito del modello di previsione è quello di:

- sintetizzare in maniera coerente una grande massa di dati osservati nel tempo;
- estrapolare, dalle sintesi ottenute, delle leggi empiriche che siano interpretabili;
- utilizzare queste leggi per produrre delle stime di valori residui futuri, con un margine di errore misurabile.

* **EDITORIALE DOMUS Via Mazzocchi
1/3 Rozzano MI – (Italy) Tel.+39 02
82472.1**

* **NUNATAC Via Crocefisso 5, 20122
Milano – (Italy) Tel.+39 02 86996848**