

Correlazione tra errori di previsione dei tassi di default: implicazioni sulle perdite inattese di portafoglio.

*Correlation in forecast errors of default rates:
aspects in portfolio unexpected losses.*

Paolo Chirico

Dip. di Statistica e Matematica applicate alle Scienze Umane “Diego de Castro”
Università degli Studi di Torino – chiricop@katamail.com

Abstract: The aim of this paper is to prove the importance of correlation in forecast errors of cluster default rates in evaluation of unexpected losses. Besides it is suggested some guideline in evaluation of the correlation.

Keywords: Perdite attese/inattese, tassi di default, cluster, correlazioni.

1. Introduzione al problema.

La valutazione delle perdite inattese, derivanti da default (insolvenza) dei richiedenti i prestiti, è una problematica di rilevante crucialità nelle politiche di gestione del rischio del portafoglio clienti di una banca. In tal senso, il presente lavoro vuole, innanzitutto, dimostrare come ai fini di una corretta valutazione di tali perdite sia essenziale la considerazione delle correlazioni, non tra tassi di default come generalmente la letteratura propone, bensì tra gli errori di previsione dei tassi stessi. Secondariamente propone alcune linee guida per affrontare il problema della valutazione di dette correlazioni.

E' facile comprendere come l'ammontare di perdite da default (insolvenza) relative al portafoglio clienti di una banca è, per diverse ragioni, in ogni esercizio t , una variabile casuale. Per far fronte al rischio di default dei suoi clienti, sembrerebbe quindi logico che la banca predisponga riserve pari alla media delle perdite totali, L_t , o *perdite attese*. Tuttavia le perdite effettive possono essere superiori a quelle attese generando *perdite inattese*, per la cui valutazione può essere logico considerare quanto in media le perdite effettive si discostano da quelle attese. E' evidente che la banca, oltre a tener conto delle perdite attese, debba considerare *le perdite inattese*, se non vuole correre il rischio di subirle. Il problema delle perdite inattese può essere ricondotto ovviamente in termini di tassi di default, tenendo presente che le perdite (effettive, attese e non) si ottengono moltiplicando il totale del capitale esposto per il corrispondente tasso (medio) di default¹. Pertanto le perdite inattese sono riconducibili allo scostamento tra tasso effettivo di default e tasso atteso di default.

Il tasso generale di default al tempo t , $d(t)$, di una banca può essere espresso come:

$$d(t) = \sum_i d(t)_i \frac{n_i}{n} \quad (1)$$

¹ Per semplicità si suppone che l'unica fonte di perdita sia il default

dove $d_i(t)$, sono i tassi di effettivi di default nei vari cluster, n_i le numerosità dei cluster ed n la numerosità totale del portafoglio clienti.

Il tasso $d_i(t)$, inteso come percentuale di default delle posizioni del cluster i è, per quanto detto, una variabile casuale così definibile:

$$d_i(t) = p_i(t) + u_i(t) \quad (2)$$

dove $p_i(t)$ è il tasso atteso, ovvero la probabilità di default delle posizioni nel cluster i al tempo t e $u_i(t)$ una componente accidentale², tale che $E[u_i(t)] = 0$ e $\text{Var}[u_i(t)] = p_i(t)(1 - p_i(t))/n_i(t)$.

Tuttavia in fase di previsione $p_i(t)$ non è nota, ma stimata con $\hat{p}_i(t)$, che la banca formula sulla base delle informazioni a sua disposizione. Con riferimento a detto tasso previsto, gli scostamenti tra tasso effettivo e tasso previsto sono quindi:

$$d_i(t) - \hat{p}_i(t) = e_i(t) + u_i(t) \quad (3)$$

dove $e_i(t) = p_i(t) - \hat{p}_i(t)$ rappresenta l'errore di previsione del tasso atteso. Tale errore, in quanto tale, è da supporre con media 0 e incorrelato alle componenti accidentali $u_i(t)$. Coerentemente con quanto appena detto, il tasso inatteso di perdita *ULR* può essere calcolato come media degli scarti al quadrato tra tassi previsti e tassi effettivi:

$$ULR = \sqrt{E\left[\sum_i \left(d_i(t) - \hat{p}_i(t)\right) \frac{n_i}{n}\right]^2} \quad (4)$$

che per quanto detto circa la natura di $e_i(t)$ e $u_i(t)$, può risciversi approssimativamente:

$$ULR = \sqrt{\sum_i \frac{n_i^2}{n^2} \text{Var}[e_i(t)] + \sum_i \sum_j \frac{n_i n_j}{n^2} \text{Cov}[e_i(t) e_j(t)] + \sum_i \frac{n_i}{n^2} \hat{p}_i(t) [1 - \hat{p}_i(t)]} \quad (5)$$

2. La correlazione tra gli errori di previsione.

Appare chiaro come il tasso inatteso di perdita *URL* dipenda dalla correlazione “inter-cluster” tra gli errori di previsione dei tassi attesi di default³.

Detta $\sigma_i^2(t)$ la varianza degli errori $e_i(t)$, le suddette correlazioni possono essere quindi formulate come:

$$\rho_{i,j}(t) = E\left[\frac{p_i(t) - \hat{p}_i(t)}{\sigma_i(t)} \frac{p_j(t) - \hat{p}_j(t)}{\sigma_j(t)}\right] \quad (6)$$

² Le deviazioni $u_i(t)$ sono concettualmente riconducibili alla “sfortuna” di avere clienti che falliscono e pertanto sono da presumersi incorrelate tra di loro.

³ $\text{Cov}[e_i(t) e_j(t)] = \rho_{i,j}(t) \sqrt{\text{Var}[e_i(t)] \text{Var}[e_j(t)]}$

La formulazione 6) pone in evidenza il problema della dipendenza temporale dei coefficienti di correlazione. In effetti i coefficienti di correlazione potrebbero variare nel tempo: in una situazione di crisi generale del sistema economico è logico ipotizzare che in tutti i cluster i tassi effettivi siano ben superiori a quelli previsti determinando correlazioni positive, mentre in una situazione di crisi solo in alcuni settori o territori e sviluppo in altri, le correlazioni possono essere miste. Volendo formalizzare questa osservazione, potremmo dire che:

$$\rho_{i,j}(t) = \rho_{i,j} + \xi_{i,j}(t) \quad (7)$$

Tuttavia, la componente $\xi_{i,j}(t)$ ha natura squisitamente congiunturale e la sua inferibilità è ben lungi dall'essere abordabile: si tratterebbe infatti di inferire come errori di previsione, frutto cioè dell'incapacità (impossibilità) di considerare correttamente tutti i fattori determinanti nel tempo t , possano essere correlati specificatamente nello stesso tempo. Tutto quello che pare realistico tentare è di stimare eventuali mutue sistematicità negli errori di previsione, ovvero concentrarci su $\rho_{i,j}$.

Anche questa stima necessita di qualche assunzione circa la variabilità degli errori $e_i(t)$, più precisamente la stazionarietà in varianza dei processi da loro identificati. Tale assunzione è risultata verosimile con la valutazione a posteriori dei tracciati degli errori nei vari cluster.

Pertanto, data una serie temporale di dati, una stima della correlazione tra $e_i(t)$ e $e_j(t)$ potrebbe essere:

$$\rho_{i,j} = \frac{\sum_t (p_i(t) - \hat{p}_i(t))(p_j(t) - \hat{p}_j(t))}{\sqrt{\sum_t (p_i(t) - \hat{p}_i(t))^2 \sum_t (p_j(t) - \hat{p}_j(t))^2}} \quad (8)$$

Appare chiaro come la questione cruciale in (3) sia l'individuazione delle previsioni $\hat{p}_i(t)$ ⁴. Le possibili strade sono sostanzialmente due:

- 1) assumere stazionarie in media i processi delle variabili tasso $p_i(t)$;
- 2) assumere un opportuno modello che spieghi l'andamento temporale del tasso medio di default per ogni cluster.

Nel primo caso la stima della correlazione risulta molto semplice, ovvero:

$$\rho_{i,j} = \frac{\sum_t (p_i(t) - \bar{p}_i(t))(p_j(t) - \bar{p}_j(t))}{\sqrt{\sum_t (p_i(t) - \bar{p}_i(t))^2 \sum_t (p_j(t) - \bar{p}_j(t))^2}} \quad (9)$$

dove $\bar{p}_i(t)$ è la media aritmetica della serie delle probabilità $p_i(t)$.

Nel secondo caso il discorso è più complesso. La gamma dei modelli di spiegazione e previsione delle probabilità di default è chiaramente vasta. Tuttavia, poiché il numero di cluster è decisamente elevato e variegato, è auspicabile l'adozione di una modellistica

⁴ $p_i(t)$ è invece desumibile dalla "Base Informativa Pubblica" della B.d.I.

standard, adottabile per ogni tasso di default. E' altresì auspicabile che il modello non abbia molti parametri in quanto le serie $p_i(t)$ a disposizione sono relativamente brevi. In tal senso si possono delineare due approcci:

- 1) *Approccio Serie Storiche*, ovvero modelli che si basano sulla sola variabile in esame (tasso medio di default per cluster $p_i(t)$), quali i modelli ARIMA unidimensionali. Con tale impostazione, per ogni tasso di default, il modello previsivo deve essere un ARIMA(p,d,q) con numero totale di parametri contenuto.
- 2) *Approccio Econometrico*, ovvero modelli econometrici in cui il tasso medio di default è spiegato da un insieme comune di variabili retrodatate:

$$p_i(t) = f(x_1(s_1), \dots, x_k(s_k)) + e_i(t) \quad s_1, \dots, s_k < t$$

3. Osservazioni finali.

In fase d'analisi si sono seguiti entrambi gli approcci sopra prospettati e risultati generalmente migliori si sono ottenuti con l'approccio econometrico, anche se in diversi casi il modello ARIMA è risultato preferibile. Tale circostanza porta a concludere che l'andamento storico, da solo, non è generalmente molto informativo dell'andamento futuro dei tassi di default. Più opportuno desumere questo in funzione di altre variabili quali ad esempio il tasso di sconto applicato dalla Banca d'Italia. Tuttavia l'approccio econometrico ha evidenziato la grossa difficoltà di individuare un set comune di variabili esplicative. Tale obiettivo costituisce una logica necessità operativa, visto che l'incrocio settore-provincia determina un numero decisamente elevato di cluster, per ognuno dei quali non è pensabile un'analisi econometrica ad hoc.

In tal senso, in quasi tutti i cluster, è risultato adeguato considerare:

- il tasso medio di sconto della Banca d'Italia, nell'anno precedente;
- il tasso di default effettivo del cluster nell'anno precedente.

E' altresì evidente che il vincolo di operare con modellizzazioni standard, per spiegare tassi di default in cluster molto differenti fra di loro, penalizza chiaramente la capacità esplicativa dei modelli stessi. In alternativa si possono, al più, individuare gruppi di variabili comuni per categorie di cluster, ma distinti da categoria a categoria.

Tuttavia, benché un'accurata previsione dei tassi di default sia logicamente fondamentale per una proficua gestione del rischio, ai fini della stima delle correlazioni tra errori di previsione si è riscontrata una certa robustezza rispetto ai modelli previsivi.

Bibliografia.

- Banca d'Italia (2001) *Base Informativa Pubblica*, Roma
- Resti A. (a cura di) (2001) *Misurare il rischio di credito, una guida metodologica*, Fondo Interbancario di Tutela dei Depositi, Roma
- Sironi A., (1999) I modelli per la stima dei tassi di insolvenza basati sui dati del mercato dei capitali, *La misurazione e gestione del rischio di credito*, Bancaria Editrice, Roma
- Szego G., Varetto F., (1999) *Il rischio di credito: misura e controllo*, Utet, Torino