

La classe dei modelli ACD per i dati ad altissima frequenza: un'applicazione al titolo Generali

Paola Zuccolotto

Dipartimento di Metodi Quantitativi, Università di Brescia
E-mail: zuk@eco.unibs.it

Summary: Financial data analysis has recently focused its interest in intradaily models, thanks to the availability of data collected at ever higher frequencies. The natural limit in increasing frequency is achieved when *all transactions* are recorded: these are the so called *ultra-high frequency data* and models have been developed directly tailored to their features. The main feature is that observations are irregularly spaced, in fact these models are an alternative to the usual custom to aggregate data to some fixed time interval. Aim of this paper is to introduce ultra-high frequency data in general, their description with statistical instruments and the problems possibly arising from their analysis. After the presentation of the new context of work, the basic ACD model, a proposal developed by Engle and Russel (1995, 1998), will be examined. Finally an application to italian Stock Market, with particular attention to QML estimation of parameters and semiparametric hazard estimation, will be presented.

Key words: Ultra-high frequency data, Duration, Hazard, ACD Models.

1. Introduzione

La volatilità dei mercati finanziari impone la raccolta dei dati ad intervalli temporali sempre più ravvicinati, nell'intento di minimizzare la naturale perdita di informazione insita nella rilevazione ad intervalli fissi. Tale

perdita di informazione si annulla solo nel momento in cui vengono registrate tutte le transazioni avvenute. I dati che derivano dalla rilevazione esaustiva sono detti *dati ad altissima frequenza* e la loro caratteristica principale è di non essere equispaziati nel tempo.

L'ambito metodologico per l'analisi di questo tipo di dati è quello dei processi puntuali, definiti come la successione $\{t_1, t_2, \dots, t_i, \dots\}$ dei tempi in cui si è verificato un certo evento casuale. I processi puntuali possono adattarsi al campo finanziario nel momento in cui l'evento che si verifica ai tempi $t_1, t_2, \dots, t_i, \dots$ è rappresentato da una transazione, di modo che il tempo t_i indica il momento in cui lo scambio i -esimo viene perfezionato, ma possono ipotizzarsi anche processi generati da altri eventi del mercato finanziario.

Tale contesto viene descritto equivalentemente attraverso due diversi tipi di variabili casuali: la variabile conteggio N_t , che indica il numero di eventi avvenuti entro il tempo t e la variabile durata X_i che indica il tempo intercorso tra l'evento $(i - 1)$ -esimo e i -esimo, alla quale si farà riferimento nel seguito. La differenza fondamentale tra le due variabili è che risultano definite rispettivamente nel *tempo di calendario* e nel *tempo degli eventi*: entrambe sono equispaziate nel loro asse temporale ma parlare in termini di durate, e quindi di tempo degli eventi, implica una deformazione del consueto asse temporale, costituito dal tempo di calendario. La trattazione del problema in termini di durate fornisce quindi un agile approccio ai dati ad altissima frequenza (che sono appunto equispaziati nel tempo degli eventi ma non nel tempo di calendario) e allo stesso tempo consente interessanti interpretazioni dei risultati nell'ambito dei modelli di deformazione temporale introdotti da Clark (1973). Tale deformazione temporale può incorporare, tra l'altro, la nota relazione tra volatilità e volumi, rendendo l'approccio qui descritto estremamente attuale.

In quest'ambito Engle e Russel (1998) hanno recentemente proposto la classe dei modelli ACD (Autoregressive Conditional Duration) che promette di essere uno strumento affidabile ed efficace per il trattamento dei dati ad altissima frequenza. Scopo di questo lavoro è delineare i tratti fondamentali dei modelli ACD e mostrarne il funzionamento attraverso un'applicazione a dati reali. Nel secondo paragrafo saranno richiamati

i fondamenti probabilistici dei processi puntuali con variabili esplicative e sarà introdotta l'ipotesi di base dalla quale prendono forma i modelli ACD, che saranno brevemente descritti nel paragrafo successivo. Il quarto paragrafo darà spazio infine all'applicazione ai dati del titolo Generali, con particolare riguardo alle problematiche dell'ottenimento di stime di quasi-massima verosimiglianza ed alla stima semiparametrica della funzione di rischio.

2. Trattamento statistico dei dati ad altissima frequenza

Sia $x_i = t_i - t_{i-1}$ ($x_i > 0$), la durata dell'evento i -esimo, e $\mathbf{x}_{i-1} = \{x_{i-1}, x_{i-2}, \dots\}$ il vettore contenente le durate degli eventi passati. Nei processi puntuali con variabili esplicative al generico evento i -esimo, oltre che una durata x_i , è associato un vettore \mathbf{y}_i di variabili che hanno la funzione di fornire un'ulteriore descrizione dell'evento avvenuto (ad esempio il prezzo a cui è si è perfezionata la transazione, il numero di azioni vendute, etc.). Con $\bar{\mathbf{y}}_{i-1}$ si indica il passato delle variabili esplicative. A differenza dei processi puntuali semplici, in presenza di variabili esplicative si dispone quindi di due variabili casuali, X_i e \mathbf{Y}_i , e viene presa in considerazione la loro funzione di densità congiunta, condizionata all'informazione passata \mathcal{I}_{i-1} , indicizzata da un vettore di parametri $\boldsymbol{\theta}_i$:

$$(X_i, \mathbf{Y}_i) | \mathcal{I}_{i-1} \sim g(x_i, \mathbf{y}_i | \mathbf{x}_{i-1}, \bar{\mathbf{y}}_{i-1}; \boldsymbol{\theta}_i)$$

che può essere scritta anche come:

$$g(x_i, \mathbf{y}_i | \mathbf{x}_{i-1}, \bar{\mathbf{y}}_{i-1}; \boldsymbol{\theta}_i) = f(x_i | \mathbf{x}_{i-1}, \bar{\mathbf{y}}_{i-1}; \boldsymbol{\theta}_{i1}) q(\mathbf{y}_i | x_i, \mathbf{x}_{i-1}, \bar{\mathbf{y}}_{i-1}; \boldsymbol{\theta}_{i2})$$

espressione talvolta utile per la stima dei parametri con il metodo della massima verosimiglianza. La funzione di densità marginale della durata i -esima, calcolabile a partire dalla congiunta come:

$$f(x_i | \mathbf{x}_{i-1}, \bar{\mathbf{y}}_{i-1}; \boldsymbol{\theta}_{i1}) = \int_{\mathbf{y}_i \in \Omega_{\mathbf{y}_i}} g(x_i, \mathbf{y}_i | \mathbf{x}_{i-1}, \bar{\mathbf{y}}_{i-1}; \boldsymbol{\theta}_i) d\mathbf{y}_i$$

consente di ricavare, in corrispondenza di ogni istante temporale $t > t_{i-1}$, le usuali grandezze definite nel contesto dei processi puntuali, come la

funzione di sopravvivenza

$$\begin{aligned} S(t|\mathbf{x}_{i-1}, \bar{\mathbf{y}}_{i-1}; \boldsymbol{\theta}_{i1}) &= 1 - F(t|\mathbf{x}_{i-1}, \bar{\mathbf{y}}_{i-1}; \boldsymbol{\theta}_{i1}) = \\ &= \int_{t-t_{i-1}}^{\infty} f(x_i|\mathbf{x}_{i-1}, \bar{\mathbf{y}}_{i-1}; \boldsymbol{\theta}_{i1}) dx_i \end{aligned}$$

dove

$$F(t|\mathbf{x}_{i-1}, \bar{\mathbf{y}}_{i-1}; \boldsymbol{\theta}_{i1}) = \int_0^{t-t_{i-1}} f(x_i|\mathbf{x}_{i-1}, \bar{\mathbf{y}}_{i-1}; \boldsymbol{\theta}_{i1}) dx_i$$

è la funzione di ripartizione¹, ed il tasso istantaneo di accadimento dell'evento, detto rischio o *hazard*

$$h(t|\mathbf{x}_{i-1}, \bar{\mathbf{y}}_{i-1}; \boldsymbol{\theta}_{i1}) = \frac{f(t-t_{i-1}|\mathbf{x}_{i-1}, \bar{\mathbf{y}}_{i-1}; \boldsymbol{\theta}_{i1})}{S(t|\mathbf{x}_{i-1}, \bar{\mathbf{y}}_{i-1}; \boldsymbol{\theta}_{i1})}.$$

Per maggiori dettagli su queste funzioni si veda Lancaster (1990). Se la densità congiunta è nota, la stima dei parametri si effettua massimizzando la funzione di log-verosimiglianza

$$\mathcal{L}(\boldsymbol{\eta}; X, \mathbf{Y}) = \sum_{i=1}^N \log g(x_i, \mathbf{y}_i|\mathbf{x}_{i-1}, \bar{\mathbf{y}}_{i-1}; \boldsymbol{\theta}_i(\boldsymbol{\eta}))$$

dopo aver assunto che g dipenda da un numero finito di parametri $\boldsymbol{\eta}$ invarianti sugli eventi, attraverso una funzione $\boldsymbol{\theta}_i$ degli stessi. Nel seguito, per non appesantire la notazione, si utilizzerà semplicemente il simbolo $\boldsymbol{\eta}$ in luogo di $\boldsymbol{\theta}_i(\boldsymbol{\eta})$. Engle e Russel (1994, 1998) hanno introdotto in questo contesto un'ipotesi semplificatrice non nuova nella teoria dei processi puntuali. Detta

$$\psi_i = E(x_i|\mathbf{x}_{i-1}, \bar{\mathbf{y}}_{i-1}; \boldsymbol{\eta}_1) = \int_0^{\infty} x_i f(x_i|\mathbf{x}_{i-1}, \bar{\mathbf{y}}_{i-1}; \boldsymbol{\eta}_1) dx_i$$

¹Si noti che per la funzione di ripartizione viene utilizzata la notazione $F(t) = Pr\{x_i < t - t_{i-1}\}$ anziché quella usuale secondo cui si dovrebbe scrivere $F(t - t_{i-1}) = Pr\{x_i < t - t_{i-1}\}$. Formalmente $F(t)$ risulta definita per $t \in (t_{i-1}; \infty)$, anziché per $t \in (0; \infty)$ come imporrebbe il supporto della variabile X_i di cui F è funzione di ripartizione, e ciò corrisponde ad una conveniente traslazione di F . In questo contesto si preferisce utilizzare l'espressione $F(t)$ in quanto ne rende più chiaro il significato: $Pr\{x_i < t - t_{i-1}\}$ corrisponde, infatti, alla probabilità che l' i -esimo evento accada entro il tempo t , dato che l' $(i-1)$ -esimo evento è accaduto esattamente al tempo t_{i-1} .

la media condizionata della i -esima durata, si ipotizza che essa riassume tutta la dipendenza temporale presente nelle durate, attraverso l'assunzione:

$$\frac{x_i}{\psi_i} = \epsilon_i \sim i.i.d. \quad \text{con} \quad E(\epsilon_i) = 1 \quad (1)$$

in base alla quale si può scrivere $f(x_i | \mathbf{x}_{i-1}, \bar{\mathbf{y}}_{i-1}; \boldsymbol{\eta}_1) = f(x_i | \psi_i; \boldsymbol{\eta}_1)$. Valendo la (??), e ricordando la fattorizzazione della g nelle due funzioni f e q , la log-verosimiglianza può essere espressa come

$$\mathcal{L}(\boldsymbol{\eta}; X, \mathbf{Y}) = \sum_{i=1}^N [\log f(x_i | \psi_i; \boldsymbol{\eta}_1) + \log q(\mathbf{y}_i | x_i, \mathbf{x}_{i-1}, \bar{\mathbf{y}}_{i-1}; \boldsymbol{\eta}_2)] \quad (2)$$

espressione che ha il vantaggio di prescindere dalla specificazione della densità g (che incorpora dinamiche congiunte di durate e variabili spesso difficili da esplicitare a priori) e di consentire, se i parametri $\boldsymbol{\eta}_1$ e $\boldsymbol{\eta}_2$ non si influenzano a vicenda, la massimizzazione dei due termini separatamente (Engle, 1996). I modelli ACD nella loro formulazione originaria definiscono un modello per le durate, quindi focalizzano l'attenzione sulla sola funzione f : l'aggiunta di variabili esplicative avviene in un secondo momento e non sarà trattata in questo lavoro.

3. I modelli ACD

L'ipotesi di lavoro (??) risulta fondamentale in quanto permette di modellare le durate semplicemente stabilendo la forma funzionale della media condizionata ψ_i e la distribuzione degli errori ϵ_i . La classe dei modelli ACD ipotizza per ψ_i una struttura autoregressiva e diversi modelli della classe si ottengono al variare dell'assunzione su ϵ_i . La formulazione originaria del modello ACD(q, p) è data da

$$\begin{aligned} x_i &= \psi_i \epsilon_i \\ \psi_i &= \omega + \sum_{j=1}^q \alpha_j x_{i-j} + \sum_{j=1}^p \beta_j \psi_{i-j} \end{aligned} \quad (3)$$

dove, al fine di garantire la positività delle medie ψ_i e la stazionarietà del processo deve essere $\omega > 0$, $\alpha_j, \beta_j \geq 0$ e $\sum_j (\alpha_j + \beta_j) < 1$. Per quanto riguarda la distribuzione degli errori, le ipotesi più frequenti sono

l'Esponenziale e la Weibull, dalle quali derivano rispettivamente i modelli EACD(q,p) e WACD(q,p) (Engle e Russel, 1998). In particolare, l'ipotesi Esponenziale trova giustificazione nella teoria dei processi di Poisson, in base alla quale l'ipotesi di durate condizionatamente indipendenti ed esponenzialmente distribuite comporta che la media ψ_i e la funzione di rischio siano costanti lungo tutto l'intervallo temporale tra t_{i-1} e t_i e siano inoltre l'una il reciproco dell'altra:

$$f(x_i|\psi_i; \boldsymbol{\eta}) = \frac{1}{\psi_i} \exp \left\{ -\frac{1}{\psi_i} x_i \right\}$$

$$E(x_i|\psi_i; \boldsymbol{\eta}) = \psi_i = \omega + \sum_{j=1}^q \alpha_j x_{i-j} + \sum_{j=1}^p \beta_j \psi_{i-j}$$

$$h(t|\psi_i; \boldsymbol{\eta}) = \frac{f(t - t_{i-1}|\psi_i; \boldsymbol{\eta})}{S(t|\psi_i; \boldsymbol{\eta})} = \frac{\frac{1}{\psi_i} \exp \left\{ -\frac{1}{\psi_i} (t - t_{i-1}) \right\}}{\int_{t-t_{i-1}}^{\infty} \frac{1}{\psi_i} \exp \left\{ -\frac{1}{\psi_i} x_i \right\} dx_i} = \frac{1}{\psi_i}$$

dove $t > t_{i-1}$ e $\boldsymbol{\eta} = (\omega, \alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_p)$ è il vettore contenente i parametri del modello, che possono essere stimati massimizzando il primo termine della (??) che in questo caso diviene:

$$\mathcal{L}(X; \boldsymbol{\eta}) = \sum_{i=1}^N [\log f(x_i|\psi_i; \boldsymbol{\eta})] = - \sum_{i=1}^N \left[\log \psi_i + \frac{x_i}{\psi_i} \right].$$

L'ipotesi di durate condizionatamente indipendenti e distribuite come una Weibull(γ_i, κ_i) costituisce una frequente alternativa alla distribuzione Esponenziale. In questo caso la media condizionata ψ_i è data da

$$\psi_i = \kappa_i \Gamma \left(1 + \frac{1}{\gamma_i} \right)$$

dove $\Gamma(\cdot)$ è la funzione Gamma. Poichè, condizionatamente a ψ_i , le x_i devono essere identicamente distribuite, è necessario fissare uno dei due parametri γ_i e κ_i (altrimenti in corrispondenza della stessa media ψ_i potrebbero aversi distribuzioni con diversi parametri):

posto $\gamma_i = \gamma$ per ogni i , si ha

$$\kappa_i = \frac{\psi_i}{\Gamma\left(1 + \frac{1}{\gamma}\right)}$$

e quindi il modello WACD(q, p) è specificato come

$$f(x_i|\psi_i; \boldsymbol{\eta}) = \frac{\gamma}{x_i} \left(\frac{x_i \Gamma\left(1 + \frac{1}{\gamma}\right)}{\psi_i} \right)^\gamma \exp \left\{ - \left(\frac{x_i \Gamma\left(1 + \frac{1}{\gamma}\right)}{\psi_i} \right)^\gamma \right\}$$

$$E(x_i|\psi_i; \boldsymbol{\eta}) = \psi_i = \omega + \sum_{j=1}^q \alpha_j x_{i-j} + \sum_{j=1}^p \beta_j \psi_{i-j}$$

$$h(t|\psi_i; \boldsymbol{\eta}) = \left(\frac{\Gamma\left(1 + \frac{1}{\gamma}\right) (t - t_{i-1})}{\psi_i} \right)^{-\gamma} \frac{\gamma}{t - t_{i-1}}$$

dove $t > t_{i-1}$ e $\boldsymbol{\eta} = (\omega, \alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_p, \gamma)$.

La log-verosimiglianza in questo caso è data da

$$\mathcal{L}(X; \boldsymbol{\eta}) = \sum_{i=1}^N \log \left(\frac{\gamma}{x_i} \right) + \gamma \log \left(\frac{x_i \Gamma\left(1 + \frac{1}{\gamma}\right)}{\psi_i} \right) - \left(\frac{x_i \Gamma\left(1 + \frac{1}{\gamma}\right)}{\psi_i} \right)^\gamma.$$

L'ipotesi di Weibull è equivalente ad assumere una distribuzione esponenziale per x_i^γ così che le durate più lunghe sono considerate più o meno probabili a seconda che γ sia rispettivamente minore o maggiore dell'unità.

Come si dirà tra breve, altre ipotesi sono comunque possibili riguardo la distribuzione degli errori e in base a questo si adeguano la funzione rischio e quella di log-verosimiglianza.

Un importante teorema (Engle e Russel, 1998; Engle, 2000) assicura peraltro la possibilità di ottenere stime di quasi-massima verosimiglianza ottimizzando la funzione

$$\mathcal{L}(\boldsymbol{\eta}; X) = - \sum_{i=1}^N \left[\log \psi_i + \frac{x_i}{\psi_i} \right]$$

che, in base a quanto appena detto, corrisponde alla funzione di log-verosimiglianza esatta sotto l'ipotesi Esponenziale, così che risulta possibile stimare i parametri senza porre alcuna ipotesi sulla distribuzione degli errori. Per i risultati dell'applicazione di questi modelli su dati simulati, con particolare riguardo alla robustezza delle stime sotto l'ipotesi Esponenziale, si veda Zuccolotto (2001).

Estensioni al modello ACD di base si ottengono sia variando l'assunzione sulla media condizionata ψ_i , che agendo sulla distribuzione degli errori.

Per quanto riguarda la forma di ψ_i , è possibile aggiungere variabili esogene in grado di spiegare la frequenza delle transazioni, inoltre sono stati proposti vari modelli non lineari come ad esempio il log-ACD (Bauwens e Giot, 1997) o l'ACD Asimmetrico (Bauwens e Giot, 1998). Per un'applicazione di quest'ultimo al titolo Generali (con gli stessi dati utilizzati nel presente lavoro) si veda De Luca (2001).

Nuove ipotesi sulla distribuzione degli errori possono includere la Gamma Generalizzata, la Log-Logistica, La Log-Normale, la distribuzione di Burr (Grammig *et al.*, 1998).

Ulteriori estensioni derivano dalla evidente relazione esistente tra i noti modelli GARCH di Engle e Borsellev ed i modelli ACD, in base alla quale sono spesso possibili per i secondi le stesse estensioni già proposte per i primi, con qualche ovvio aggiustamento. È il caso dei modelli FI-ACD introdotti da Jasiak (1999) per i quali un'applicazione ad un titolo italiano è stata proposta da Golia (2001).

4. Un'applicazione al mercato finanziario italiano: il titolo Generali

In questo paragrafo saranno esposti i risultati della stima di alcuni modelli della classe ACD sui dati derivanti dalla contrattazione del titolo Generali alla Borsa di Milano (i dati sono stati forniti dalla Borsa Italiana S.p.A.) dal 3/04/2000 al 19/04/2000. Durante tale periodo sono state registrate 87866 transazioni. L'apparato teorico è stato illustrato riferendosi alla durata x_i come al tempo intercorrente tra la $(i - 1)$ -esima e la i -esima transazione, ma in realtà, come sottolineato in precedenza, si può assumere come “evento i -esimo” qualunque evento di mercato. Nel caso del titolo Generali, ad esempio, si osservano spesso transazioni successive avvenute al medesimo prezzo. Ai fini interpretativi appare, invece, di particolare interesse estrapolare la dinamica secondo cui si succedono contrattazioni a prezzi differenti, quindi in questo caso è sensato analizzare, anziché le durate tra successive transazioni, le durate tra successive variazioni del prezzo. Spostando l'attenzione su questo diverso evento di mercato le durate da analizzare si riducono a 22638. Eliminate le durate nulle, le contrattazioni della fase di pre-apertura, le durate tra il primo evento della giornata e l'ultimo della giornata precedente, il *dataset* definitivo risulta formato da 19697 durate, rappresentate graficamente in Figura 1, dove si nota evidentemente l'alternarsi di periodi di quiete (durate elevate) a periodi di turbolenza (durate piccole). L'osservazione del correlogramma suggerisce una decisa presenza di autocorrelazione, fino a *lag* elevati. Per 4 modelli (ACD(1,1) - ACD(1,2) - ACD(2,1) - ACD(2,2)) sono state ottenute le stime di quasi-massima verosimiglianza, riportate in Tabella 1, assieme ai relativi scarti quadratici medi² e ad alcuni indici per valutare la bontà del modello, basati sull'esame dei residui $\hat{\epsilon}_i = \frac{x_i}{\hat{\psi}_i}$ (errori stimati): il test di Ljung-Box per l'ipotesi di incorrelazione dei residui e dei loro quadrati, la media e lo scarto quadratico medio dei residui, la varianza dello scarto tra la durata effettiva e la sua media stimata.

²Quando uno dei parametri stimati è risultato non significativo la massimizzazione della log-verosimiglianza è stata condotta nuovamente, vincolando detto parametro a zero, ed in luogo dello scarto quadratico medio è stato indicato il simbolo (\cdot). In nessuno dei tre casi, comunque, detta procedura ha comportato variazioni sulla stima e sul relativo scarto quadratico medio per gli altri parametri.

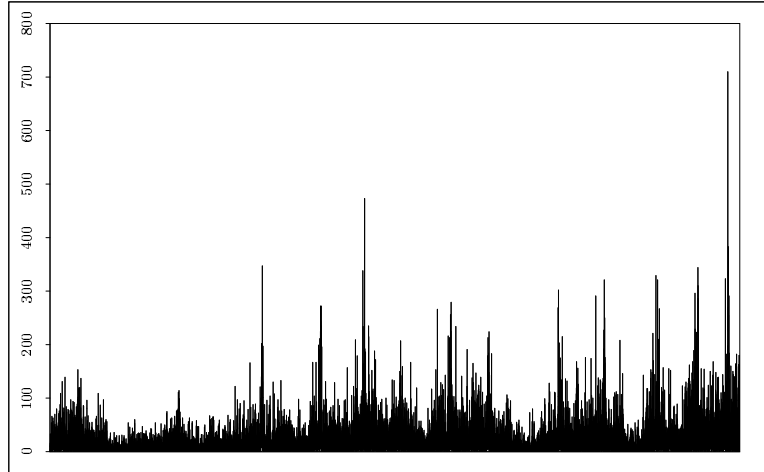


Figura 1. Durate del titolo Generali dal 4/4/00 al 19/4/00

Si nota che gli scarti quadratici medi delle stime evidenziano un'elevata significatività di tutte le stime non nulle. Inoltre il test di Ljung-Box, che va confrontato con la soglia critica $\chi_{20,5\%}^2 = 31.4$, consente di escludere con decisione i primi tre modelli. Per il modello ACD(2,2), che si configura immediatamente come quello migliore, risultano invece accettabili le ipotesi di incorrelazione sia dei residui $\hat{\epsilon}_i$ (a questo proposito si veda anche il correlogramma riportato in Figura 2) che dei loro quadrati $\hat{\epsilon}_i^2$. I residui $\hat{\epsilon}_i$ hanno media unitaria, in accordo con l'ipotesi di partenza. Lo scarto quadratico medio dovrebbe avere anch'esso valore unitario se si volesse avvalorare l'ipotesi di distribuzione Esponenziale: qui lo scostamento, pur di piccola entità, segnala la possibilità di una diversa distribuzione. Il modello ACD(2,2) è anche quello per cui è minima la varianza $\sigma^2(x_i - \hat{\psi}_i)$. Si noti che $\sigma^2(x_i) = 737.9005$, quindi l'utilizzo della stima della media condizionata $\hat{\psi}_i$ come previsore della durata x_i consente di spiegarne in parte la variabilità.

Tabella 1. Stime dei parametri di quattro modelli ACD e indicatori della bontà del modello (tra parentesi gli scarti quadratici medi delle stime)

Modello	ACD(1,1)	ACD(1,2)	ACD(2,1)	ACD(2,2)
$\hat{\omega}$	0.0519 (0.0093)	0.0509 (0.0093)	0.0620 (0.0109)	0.1225 (0.0203)
$\hat{\alpha}_1$	0.0384 (0.0020)	0.0384 (0.0020)	0 (·)	0 (·)
$\hat{\alpha}_2$	-	-	0.0447 (0.0025)	0.0842 (0.0041)
$\hat{\beta}_1$	0.9592 (0.0022)	0.9592 (0.0022)	0.9524 (0.0026)	0.0819 (0.0103)
$\hat{\beta}_2$	-	0 (·)	-	0.8281 (0.0115)
$LB_{20}(\hat{\epsilon}_i)$	288.0543	288.0388	204.1177	23.1844
$LB_{20}(\hat{\epsilon}_i^2)$	79.8931	79.8080	56.9678	24.7317
$\sigma^2(x_i - \psi_i)$	640.0199	640.0158	638.6144	632.4504
$E(\hat{\epsilon}_i)$	0.9964	0.9980	0.9975	0.9976
$\sigma(\hat{\epsilon}_i)$	1.1580	1.1599	1.1579	1.1505

Utilizzando i residui $\hat{\epsilon}_i$ dal modello ACD(2,2) è possibile stimare la funzione di rischio dell'errore $h_\epsilon(t) = \frac{f_\epsilon(t)}{S_\epsilon(t)}$ (detta *baseline hazard*), che riveste un certo interesse essenzialmente per due ordini di motivi:

- consente interessanti interpretazioni sulla dinamica del fenomeno ed anche valutazioni sulla possibile distribuzione degli errori,
- costituisce il punto di partenza per il calcolo della funzione di rischio per le durate, non altrimenti ricavabile, in quanto essa è analiticamente nota solo quando si pongono ipotesi sulla distribuzione degli errori: è semplice dimostrare (si veda in Appendice) che la funzione di rischio per le durate si ottiene a partire da quella dell'erro-

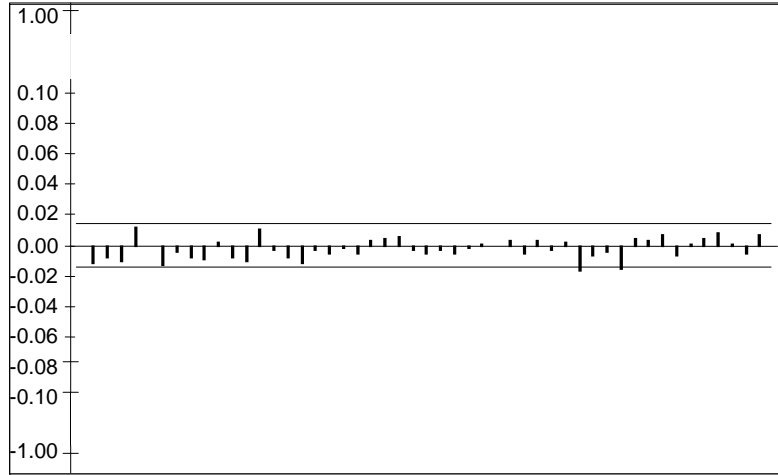


Figura 2. Correlogramma dei residui $\hat{\epsilon}_i$ dal modello ACD(2,2)

re attraverso la relazione

$$h(t|\psi_i; \boldsymbol{\eta}) = h_\epsilon \left(\frac{t - t_{i-1}}{\psi_i} \right) \frac{1}{\psi_i}.$$

Poichè non sono state fatte assunzioni sulla distribuzione degli errori, si tratta di una stima semiparametrica della funzione $h_\epsilon(t)$, che si può ottenere ad esempio nel modo suggerito da Engle (2000) : detti $\hat{\epsilon}_{(i)}$ i residui stimati disposti in ordine non decrescente si ricorre alla formula

$$\hat{h}_\epsilon(t = \hat{\epsilon}_{(i)}) = \frac{2k}{(N - i)} \frac{(\hat{\epsilon}_{(i+k)} - \hat{\epsilon}_{(i-k)})}{(N - i)} \quad i = k + 1, \dots, N - k$$

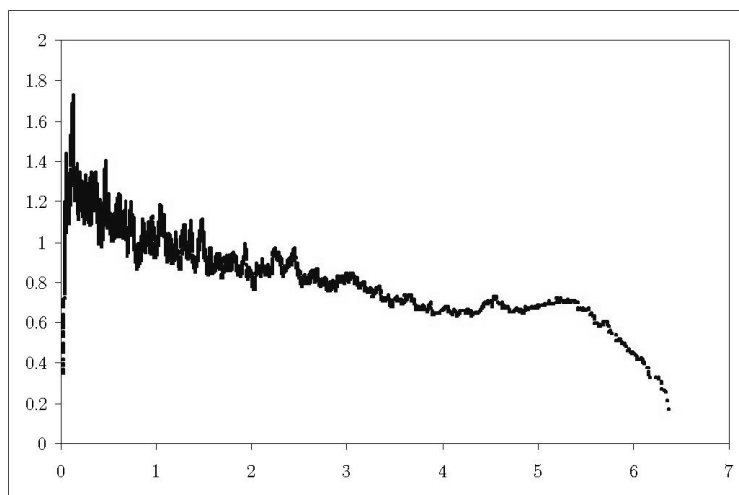


Figura 3. Rappresentazione della funzione $\hat{h}_\epsilon(t)$

che si comprende osservando che i residui $\hat{\epsilon}_i$ si possono considerare come un campione di osservazioni indipendenti ottenuto effettuando N prove su un evento casuale e registrando ad ogni prova il tempo occorso perchè l'evento si verificasse. k è un numero intero che viene fissato a seconda del grado di spianamento desiderato, qui si è posto $k = 100$, considerato anche l'elevato valore di N . In Figura 3 è rappresentata la funzione $\hat{h}_\epsilon(t)$ stimata sui residui dal modello ACD(2,2). Il rischio è elevato in corrispondenza di durate piccole, quindi procede decrescendo, prima in modo deciso poi più gradualmente. Si conferma comunque il sospetto di inadeguatezza dell'ipotesi di distribuzione Esponenziale, che imporrebbe una funzione di rischio costante pari a 1, anche se lo scostamento da tale situazione non appare particolarmente marcato. Si potrebbe ipotizzare una distribuzione Weibull con parametro γ poco distante dall'unità. Stimando sulle 19697 durate un modello WACD(2,2) si ottengono infatti i risultati riportati in Tabella 2.

Si conferma quanto ipotizzato riguardo la distribuzione degli errori, tuttavia le stime risultanti non presentano variazioni di rilievo rispetto a

Tabella 2. Stime dei parametri del modello WACD(2,2) e indicatori della bontà del modello (tra parentesi gli scarti quadratici medi delle stime)

Modello	WACD(2,2)
$\hat{\omega}$	0.1227 (0.0209)
$\hat{\alpha}_1$	0 (·)
$\hat{\alpha}_2$	0.0841 (0.0042)
$\hat{\beta}_1$	0.0821 (0.0106)
$\hat{\beta}_2$	0.8279 (0.1180)
$\hat{\gamma}$	0.9722 (0.0052)
$LB_{20}(\hat{\epsilon}_i)$	23.1661
$LB_{20}(\hat{\epsilon}_i^2)$	24.7329
$\sigma^2(x_i - \hat{\psi}_i)$	632.4315
$E(\hat{\epsilon}_i)$	0.9984
$\sigma(\hat{\epsilon}_i)$	1.1515

quelle di quasi-massima verosimiglianza ottenute in precedenza. Il leggerissimo miglioramento suggerito dagli indicatori proposti appare senz'altro trascurabile.

5. Conclusioni

Da qualche anno a questa parte la tecnologia informatica ha reso disponibile una nuova tipologia di dati, derivanti dalla registrazione di ogni singola transazione del mercato finanziario. Questi dati sono detti

“ad altissima frequenza” e la loro caratteristica fondamentale è di non essere equispaziati nel tempo. Nuovi modelli sono stati elaborati a fronte di queste attuali problematiche, al fine di sfruttare al meglio la grande potenzialità informativa insita nella rilevazione esaustiva dei movimenti avvenuti nel mercato finanziario. Attualmente la proposta più completa in questo ambito è costituita dai modelli ACD di Engle e Russel (1998), dei quali in questo lavoro è illustrata un’applicazione ad un titolo del mercato finanziario italiano.

Ringraziamenti: Desidero ringraziare il Dott. G. De Luca, la Dott.ssa S. Golia ed il Dott. M. Sandri per le stimolanti discussioni del venerdì pomeriggio, per i preziosi suggerimenti e per la simpatia che ha reso ancor più piacevole il tempo dedicato a questo lavoro.

Riferimenti bibliografici

Bauwens L., Giot P. (1997) The Logarithmic ACD Model: an application to the bid-ask quote process of three NYSE stocks, *CORE Discussion Paper - Université Catholique de Louvain*, Revised version of Discussion Paper 9789.

Bauwens L., Giot P. (1998) Asymmetric ACD Models: introducing price information in ACD Models, *CORE Discussion Paper - Université Catholique de Louvain*, Revised version of Discussion Paper 9844.

Clark P.K. (1973) A subordinated stochastic process model with finite variance for speculative prices, *Econometrica*, 41, 135-159.

De Luca G. (2001) La direzione dei prezzi di titoli azionari con i modelli ACD, *Quaderni di Statistica, Università degli Studi di Verona, Facoltà di Economia, Dipartimento di Economia, Società e Istituzioni, Sezione Statistica*, 2.

Engle R.F. (1996) The Econometrics of ultra-high frequency data, *NBER Working Paper Series*, Working Paper 5816.

Engle R.F. (2000) The Econometrics of ultra-high frequency data, *Econometrica*, 68, 1-22.

Engle R.F., Russel J.E. (1994) Forecasting transaction rates: the Autoregressive Conditional Duration Model, *NBER Working Paper Series*, Working Paper 4966.

Engle R.F., Russel J.E. (1998) Autoregressive Conditional Duration: a new model for irregularly spaced transaction data, *Econometrica*, 66, 1127-1162.

Grammig J., Hujer R., Kokot S., Maurer K.O. (1998) Modeling the Deutsche Telekom Ipo using a new ACD specification - An application of the Burr-ACD model using High-frequency Ibis data, *manuscript*.

Golia S. (2001) Long memory effects in ultra-high frequency data, *Quaderni di Statistica*, 3.

Jasiak J. (1999) Persistence in Intertrade Durations, *manuscript*.

Lancaster T. (1990) *The econometric analysis of transition data*, Cambridge University Press.

Zuccolotto P. (2001) Autoregressive Conditional Duration models for ultra-high frequency data, *Rapporti di ricerca del Dipartimento Metodi Quantitativi, Università di Brescia*, 191.

Appendice

La funzione di rischio per le durate è ricavabile a partire da quella dell'errore attraverso la relazione

$$h(t|\psi_i; \boldsymbol{\eta}) = h_\epsilon \left(\frac{t - t_{i-1}}{\psi_i} \right) \frac{1}{\psi_i}.$$

Infatti, dette $f_\epsilon(\epsilon_i)$, $S_\epsilon(t)$ e $h_\epsilon(t) = \frac{f_\epsilon(t)}{S_\epsilon(t)}$ la funzione di densità, la funzione di sopravvivenza e la funzione di rischio del generico errore ϵ_i , e $f(x_i|\psi_i; \boldsymbol{\eta})$, $S(t|\psi_i; \boldsymbol{\eta})$ e $h(t|\psi_i; \boldsymbol{\eta}) = \frac{f(t - t_{i-1}|\psi_i; \boldsymbol{\eta})}{S(t|\psi_i; \boldsymbol{\eta})}$ le stesse funzioni per la durata i -esima x_i (condizionatamente a ψ_i), essendo $x_i = \psi_i \epsilon_i$, si ha

$$f(x_i|\psi_i; \boldsymbol{\eta}) = \left| \frac{d\epsilon_i}{dx_i} \right| f_\epsilon \left(\frac{x_i}{\psi_i} \right) = \frac{1}{\psi_i} f_\epsilon \left(\frac{x_i}{\psi_i} \right)$$

$$S(t|\psi_i; \boldsymbol{\eta}) = \int_{t-t_{i-1}}^{\infty} \frac{1}{\psi_i} f_\epsilon \left(\frac{x_i}{\psi_i} \right) dx_i = \int_{\frac{t-t_{i-1}}{\psi_i}}^{\infty} \frac{1}{\psi_i} f_\epsilon(\epsilon_i) \psi_i d\epsilon_i = S_\epsilon \left(\frac{t - t_{i-1}}{\psi_i} \right)$$

e quindi

$$h(t|\psi_i; \boldsymbol{\eta}) = \frac{f(t - t_{i-1}|\psi_i; \boldsymbol{\eta})}{S(t|\psi_i; \boldsymbol{\eta})} = \frac{\frac{1}{\psi_i} f_\epsilon \left(\frac{t - t_{i-1}}{\psi_i} \right)}{S_\epsilon \left(\frac{t - t_{i-1}}{\psi_i} \right)} = h_\epsilon \left(\frac{t - t_{i-1}}{\psi_i} \right) \frac{1}{\psi_i}.$$