Quaderni di Statistica
Vol. 3, 2001

Long memory effects in ultra-high frequency data

Silvia Golia
Dipartimento di Metodi Quantitativi, Università degli Studi di Brescia
E-mail: golia@eco.unibs.it

Summary: In the present paper, tick-by-tick, or ultra-high frequency data are analized. An Autoregressive Conditional Duration (ACD)-type process for the durations between consecutive events, the Fractional Integrated Autoregressive Conditional Duration process (Jasiak, 1999), is reviewed and discussed in order to admit the presence of long memory patterns. This process, as the ACD, is based on the assumption that the temporal dependence in the durations is captured by the mean function. The long term dependency is examined on the Italian stock Tiscali time series recorded from 5^{th} to 15^{th} of May 2000.

Key words: Durations, Long Memory, ACD, Fractional Integrated ACD.

1. Introduction

The analysis of financial time series traditionally considers weekly or daily data. However, in the recent years, intra-daily data have become available and new methodologies have been developed. Besides ultra-high frequency regularly spaced data, irregularly observed data have been used. In this scenario, the datum is stored only if there is an occurrence of the event one is interested to. From a statistical point of view, the occurrence of the event is the outcome of a random variable. This implies that the durations between consecutive events are themselves random and can be described through an ACD-type stochastic process. Nevertheless, the basic formulation of the ACD (Autoregressive Conditional Duration) process, proposed by Engle and Russel (1998), as well as several extentions like the Asymmetric ACD (Bauwens and Giot, 1998), the Burr ACD (Gramming and Maurer, 1999), the Logarithmic ACD (Bauwens and Giot, 1999), the Asymmetric Log-ACD (Bauwens and Giot, 2000 and De Luca, 2001) and the SCD (Stochastic Conditional Duration) (Bauwens and Veredas, 1999), don't allow the presence of persistence or long memory in the real data. Empirically, the evidence of long memory in the durations is revealed by a highly persistent pattern of the autocorrelations, displaying an hyperbolic rate of decay. This feature must be taken into account when a model is builded and used, for example, for forecasting purposes. In this work the *Fractional Integrated Autoregressive Conditional Duration process* (FIACD), firstly introduced by Jasiak in 1999, is presented and examined.

The paper is organized as follow. In section 2 the FIACD process is reviewed and discussed. In section 3 the results of simulation procedures are shown joined with the analysis of the Italian stock Tiscali time series with a sample period from 5^{th} to 15^{th} of May 2000, available from Borsa Italiana Spa. Conclusions are reported in section 4.

2. The FIACD process

Let n be the number of events observed at random times t_i, $i = 1, \ldots, n$, and let $x_i = t_i - t_{i-1}$ be the duration between the $(i-1)$-th and the i-th event. The main assumption of the ACD class of models is that all the temporal dependence in the durations is captured by the mean function. The expected i-th duration, which is the conditional mean of the i-th duration, is written as a function of the past durations:

$$\mathrm{E}(x_i | x_{i-1}, x_{i-2}, \ldots, x_1) = g(x_{i-1}, x_{i-2}, \ldots, x_1; \boldsymbol{\theta}), \qquad (1)$$

where $\boldsymbol{\theta}$ is the parameter vector characterizing g. The ACD class of models consists of the parameterization of these expected durations and can be expressed as:

$$x_i = \psi_i \epsilon_i, \quad \{\epsilon_i\} \sim \text{i.i.d.} \quad \text{with} \quad \mathrm{E}(\epsilon_i) = 1 \tag{2}$$
$$\psi_i = g(x_{i-1}, x_{i-2}, \ldots, x_1; \boldsymbol{\theta})$$

Equation (**??**) means that x_i given the past are independent and identically distributed and $f(x_i|x_{i-1}, x_{i-2}, \ldots, x_1; \boldsymbol{\theta}) = f(x_i|\psi_i; \boldsymbol{\theta})$. In the basic formulation of an ACD(p,q) process (Engle and Russell, 1998) the conditional expected durations are expressed as

$$\psi_i = \omega + \sum_{j=1}^{p} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j} = \omega + \alpha(L)\, x_i + \beta(L)\, \psi_i,$$

where $\omega > 0, \alpha(L) = \alpha_1 L + \alpha_2 L^2 + \ldots + \alpha_p L^p, \beta(L) = \beta_1 L + \beta_2 L^2 + \ldots + \beta_q L^q$, $\alpha_j, \beta_j \geq 0$ and $\sum_{j=1}^{p} \alpha_j + \sum_{j=1}^{q} \beta_j < 1$ in order to guarantee the positivity of the durations and the stationarity of the process respectively. The ACD(p,q) process can be rewritten as an ARMA(max(p,q),q) process in x_i as following:

$$[1 - \beta(L) - \alpha(L)]x_i = \omega + [1 - \beta(L)]\, v_i, \tag{3}$$

where $v_i = x_i - \psi_i$ is a martingale difference sequence. This model takes into account only the short dependence in the expected durations and imposes an exponential decline pattern in the autocorrelation function. In order to allow a longer dependence, the ψ_i must be expressed in a different manner.

One possible specification of ψ_i is given by the more flexible *Fractional Integrated Autoregressive Conditional Duration process* (FIACD) (Jasiak, 1999). Let us consider the ARMA rapresentation (**??**) for the ACD process, applying the fractional differencing operator $(1-L)^d$, with $0 \leq d \leq 1$, to the durations x_i one obtains the following expression for the FIACD(p,d,q) process:

$$[1 - \phi(L)](1 - L)^d x_i = \omega + [1 - \beta(L)]v_i, \tag{4}$$

where, again, $v_i = x_i - \psi_i$ is a martingale difference sequence by construction, $\beta(L) = \beta_1 L + \beta_2 L^2 + \ldots + \beta_p L^p$, $\phi(L) = \phi_1 L + \phi_2 L^2 +$

$\ldots + \phi_q L^q$ and all the roots of $1 - \beta(L)$ and $1 - \phi(L)$ lie outside the unit circle. From (**??**), substituting $x_i - \psi_i$ for v_i, an alternative representation of FIACD(p,d,q) is derived:

$$\begin{aligned} [1 - \beta(L)]\psi_i &= \omega + [1 - \beta(L) - [1 - \phi(L)](1 - L)^d]x_i \\ &= \omega + \Lambda(L)x_i, \end{aligned} \tag{5}$$

where $\Lambda(L) = \lambda_1 L + \lambda_2 L^2 + \lambda_3 L^3 + \ldots$ is a polynomial of infinite order, $\omega > 0$ and $0 \leq d \leq 1$. If $0 < d < 1/2$, the FIACD is a long memory process with an hyperbolic rate of decay in the autocorrelation function.

In order to guarantee the positivity of the expected durations ψ_i all the coefficients in (**??**) must be equal or bigger than zero, i.e. $\beta_i \geq 0$ for $i = 1, 2, \ldots, p$ and $\lambda_k \geq 0$ for $k = 1, 2, \ldots$ It is interesting to observe that for $d = 0$, the FIACD process is led back to an ACD; for $d = 1$ one obtains an integrated process. For $0 < d \leq 1$ the expansion of $(1 - L)^d$ evaluated in $L = 1$ is equal to zero, and the sum of all the coefficients is equal to 1; this means that the first unconditional moment of the duration is infinite and the FIACD process is not weakly stationary. Nevertheless, following the path outlined by Bougerol and Picard (1992) for the IGARCH processes, one shows that the FIACD(p,d,q) class of processes is strictly stationary [1] and ergodic for $0 \leq d \leq 1$ (Jasiak, 1999).

A particular specification of such a process is the FIACD(1,d,1) process. In this case it is easy to derive the parameters of the $\Lambda(L)$ polynomial as function of the parameters β, ϕ and d.

Let $\pi_k = (-1)^k \left[\frac{d(d-1)(d-2)\ldots(d-k+1)}{k!}\right]$, be the terms of the expansion of $(1 - L)^d$, it is possible to write:

$$\lambda_1 = \phi - \beta + d, \quad \lambda_k = \phi\pi_{k-1} - \pi_k \qquad k = 2, 3, \ldots. \tag{6}$$

The constraints on the parameters β and ϕ which guarantee the positivity of the durations are the following:

$$0 \leq \beta \leq \phi + d, \quad \phi \leq \frac{1 - d}{2}.$$

[1] Due to the definition of *strictly stationary process*, i.e. a process whose probability structure is invariant under time shift, a strictly stationary process may not be weakly stationary because it may not have finite first and second order moments, as in this case (Rosenblatt, 1985, pag 13).

Moreover, because $\sum_{k=0}^{\infty} \pi_k = 1$, one shows that the sum of the coefficients of the process is equal to one [2], no matter the specific values of d, β and ϕ.

In general the durations x_i are not independently distributed but the joint likelihood function can always be written as the product of the conditional density functions and the log-likelihood[3]:

$$l(\boldsymbol{\theta}; x) = \sum_{i=1}^{n} \ln f(x_i|\psi_{i-1}; \boldsymbol{\theta}),$$

where ψ_{i-1} contains all the past information, according to the assumption (**??**). The density function f is chosen from a parametric family, like Exponential or Weibull, leading to a fully parametric duration model and the parameters are estimated by maximum likelihood. The log-likelihood function becomes, in the case of an Exponential distribution with parameter $\lambda = 1$,

$$l(\boldsymbol{\theta}; x) = -\sum_{i=1}^{n} \ln \psi_i - \sum_{i=1}^{n} \frac{x_i}{\psi_i}$$

whereas in the case of a Weibull distribution with parameter $\gamma > 0$ and $\delta = 1/\Gamma(1 + 1/\gamma)$

$$l(\boldsymbol{\theta}; x) = \sum_{i=1}^{n} \left\{ \ln\left(\frac{\gamma}{x_i}\right) + \gamma \ln\left(\frac{x_i \Gamma(1 + 1/\gamma)}{\psi_i}\right) - \left(\frac{x_i \Gamma(1 + 1/\gamma)}{\psi_i}\right)^{\gamma} \right\}.$$

3. Application to simulated and real durations

In this section results obtained from simulation procedures for a FI-ACD(1,d,1) process are reported. Moreover, the trading times of the Italian stock Tiscali, recorded from 5^{th} through 15^{th} of May 2000, are analyzed.

---

[2] $\beta + \sum_{k=1}^{\infty} \lambda_k = \beta + \phi - \beta + d + \sum_{k=2}^{\infty} \lambda_k = \phi + d + \sum_{k=2}^{\infty} (\phi \pi_{k-1} - \pi_k) = \phi \sum_{k=0}^{\infty} \pi_k - \sum_{k=0}^{\infty} \pi_k + 1 = 1$.

[3] This log-likelihood function does not differ from the one of the ACD models except for a different specification of ψ_i. For a wide simulation study of the ACD estimation process, see Zuccolotto (2001).

Table 1. Descriptive statistics from the simulation of a FIACD(1,d,1) model

		ω	β	ϕ	d
a)	mean	0.0089	0.628	0.23	0.406
	median	0.0087	0.629	0.228	0.407
	std.	0.002	0.0341	0.0205	0.0313
	MSE	0.000004	0.0012	0.0005	0.0010
b)	mean	0.0094	0.4988	0.2449	0.2658
	median	0.0090	0.5087	0.2507	0.2670
	std.	0.0035	0.0777	0.0590	0.0297
	MSE	0.00001	0.0060	0.0035	0.0009

The generation and estimation of a time series from a FIACD process involve the approximation of the polynomial of infinite order $\Lambda(L)$ in (**??**); in this paper the chosen truncation point k of the expansion of $(1-L)^d$ is set equal to 1000. Moreover the estimation step requires conditioning of the model on pre-sample values of the durations so that the sample is augmented by a kx1 vector of durations all set equal to the unconditional sample mean.

Table 1 reports some descriptive statistics related to the estimated parameters of 500 time series of length 10000 generated by two exponential FIACD(1,d,1) processes with the following sets of parameters:

a) $\omega = 0.0084$, $\beta = 0.63$, $\phi = 0.22$, $d = 0.41$

b) $\omega = 0.0087$, $\beta = 0.51$, $\phi = 0.25$, $d = 0.27$.

In Figure 1a, one can see the plot of a simulated time series using the first set of coefficients, joined with its autocorrelation plot (Figure 1b).

The Berndt, Hall, Hall and Hausmann (BHHH) algorithm (Berndt, Hall, Hall and Hausmann, 1974) with numerical derivatives is used in the estimation procedure. This algorithm had no trouble converging for the sample and the results appeared robust to initial values imposed.

One can conclude that constrained maximum likelihood estimation provides good estimators of the parameters of a FIACD(1,d,1) model.

The last part of this section is left to the analysis of the trading times

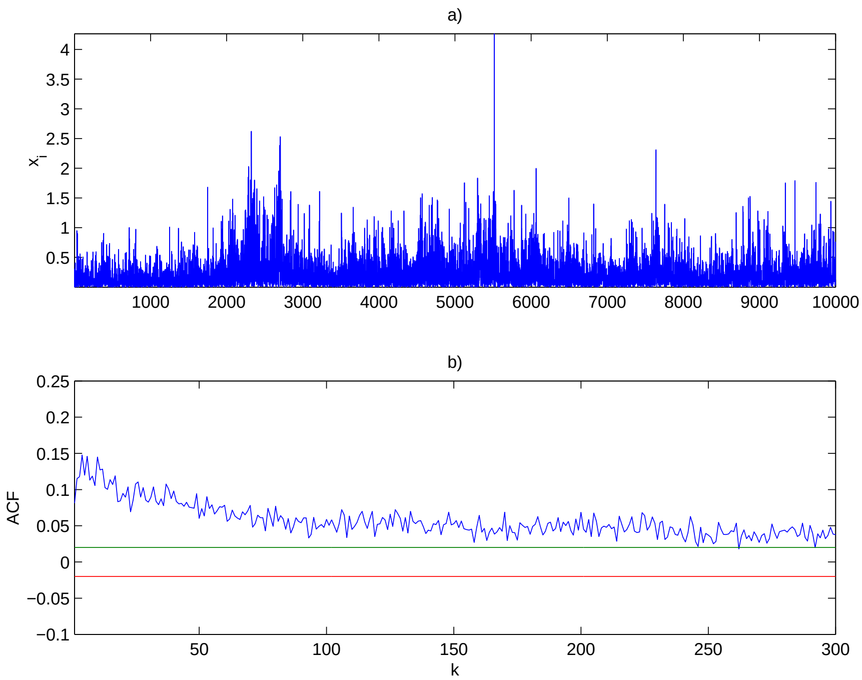

Figura 1. a) Simulated FIACD time series plot b) Autocorrelation plot

of the Italian stock Tiscali, available from Borsa Italiana Spa. The sample period goes from 5^{th} through 15^{th} of May 2000. Durations between successive changes in prices are examined, removing all the null durations due to multiple trading at the same time. Moreover it has to be pointed out that all the trades before 9:30 am, corresponding to the market opening, has been discarded. The final number of available durations is 12997. In Figure 2a the plot of these durations is drawn. The minimum duration is 1 second, the maximum 329 seconds, i.e. 5 minutes and 29 seconds. The average duration between successive changes in prices is 15.5 seconds with standard deviation 20.76.

The autocorrelation plot of the durations (Figure 2b) displays an hyperbolic decay which suggests the possible presence of long memory in the durations. For this reason an Exponential FIACD(1,d,1) model is applied to the series.

The estimated parameters with their standard errors are reported in Table 2.

The estimated parameter d is almost equal to 1, suggesting a stronger persistence than the one due to a long memory process. In order to evaluate the goodness of fit of the model, the standardized residuals, defined

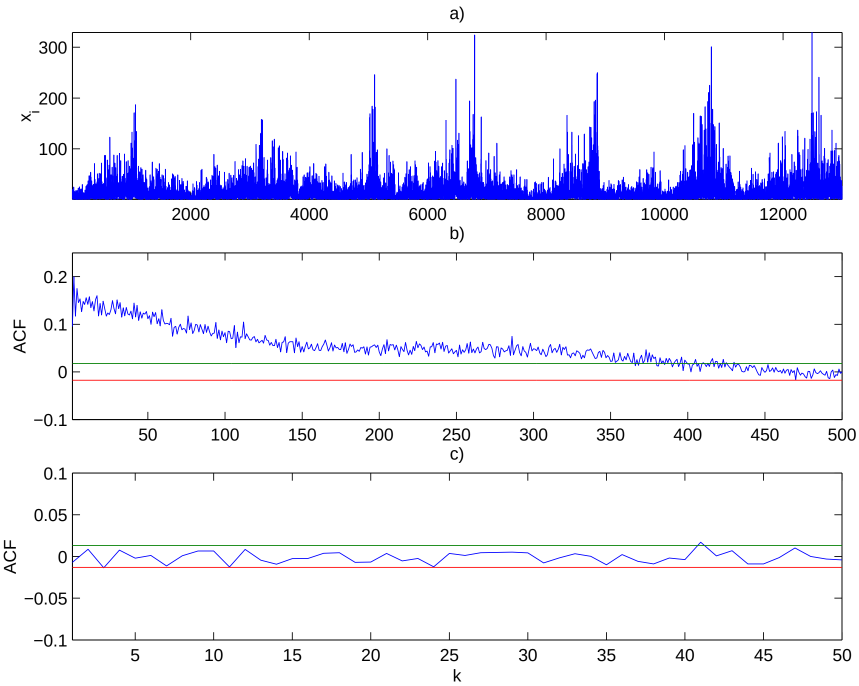

Figura 2. a) Tiscali time series (May 5^{th}-15^{th}) b) Tiscali duration Autocorrelation plot c) Residuals Autocorrelation plot

by

$$\hat{\epsilon}_i = \frac{x_i}{\hat{\psi}_i}, \qquad i = 1, 2, \ldots$$

must be analyzed. Under the hypothesis of correct specification of the model, the residuals $\hat{\epsilon}_i$ should be i.i.d. with mean and variance equal to 1. Figure 2c) shows the autocorrelation plot of the residuals and the Ljung-Box test applied to $\hat{\epsilon}_i$ and $\hat{\epsilon}_i^2$ for 20 and 50 lags gives the statistics $\text{LB}_{\hat{\epsilon}_i}(20)$=20.44, $\text{LB}_{\hat{\epsilon}_i^2}(20)$ = 18.80 (critical value=31.41) and $\text{LB}_{\hat{\epsilon}_i}(50)$ = 47.69, $\text{LB}_{\hat{\epsilon}_i^2}(50)$ = 37.15 (critical value=67.50) respectively. One can conclude that the residuals are uncorrelated. The mean and standard error of the residuals are 0.98 and 1.07 respectively.

Given that the variance of $\hat{\epsilon}_i$ is slightly bigger than one, one can assume the more general Weibull distribution. Under the assumption of a Weibull distribution with parameters $\delta = 1/\Gamma(1+1/\gamma)$ and $\gamma > 0$ for the durations, the estimations of the parameters of the Weibull FIACD(1,d,1) are reported in Table 2. Although the null hypothesis of $\gamma = 1$, which

Table 2. The Exponential FIACD(1,d,1) and Weibull FIACD(1,d,1) estimates for Tiscali; standard errors are reported in parenthesis.

	ω	β	ϕ	d	γ
EFIACD	0.0237	0.9503	-0.0393	0.9896	-
	(0.0062)	(0.0052)	(0.0110)	(0.0149)	-
WFIACD	0.0239	0.9503	-0.0395	0.9898	1.0157
	(0.0063)	(0.0059)	(0.0183)	(0.0230)	(0.0069)

drives back the model to the Exponencial FIACD, is rejected, given that the values of the estimated coefficients are almost the same for the EFIACD and the WFIACD models, one can conclude that the exponential distribution is a sufficiently good approximation of the distribution of the durations.

4. Conclusions

The paper presents and analyzes the FIACD class of models for high frequency duration data with highly persistent pattern in the autocorrelation function. A long memory feature is present if the fractional parameter d lies between 0 and 0.5. The application of the model to the Italian stock Tiscali provides an interesting result. Although it is not possible to claim that the Tiscali duration time series has been generated by a long memory process ($\hat{d}_{Tiscali} \approx 1$), the goodness of fit to the data of the FIACD(1,d,1) model shows that this model is a good one in the case of highest persistence too.

Acknowledgments: I wish to thank G. De Luca, M. Sandri and P. Zuccolotto for useful comments and suggestions.

References

Bauwens L., Giot P. (1998) Asymmetric ACD models: introducing price information in ACD models with a two state transition model, *Core Discussion Paper 9844*, Université Catholique de Louvain, Belgium.

Bauwens L., Giot P. (1999) The logarithimic ACD model : an application to the bid-ask quote process of three NYSE stocks, *Core Discussion Paper 9789*, Université Catholique de Louvain, Belgium.

Bauwens L., Giot P. (2000) Modelling and predicting intra-day price movements in stoch markets with autoregressive conditional duration models. *Core Working Paper*, Université Catholique de Louvain, Belgium.

Bauwens L., Veredas D. (1999) The stochastic conditional duration model: a latent factor model for the analysis of financial durations, *Core Discussion Paper 9958*, Université Catholique de Louvain, Belgium.

Bougerol P., Picard N. (1992) Stationarity of GARCH processes and some nonnegative time series, *Journal of Econometrics*, 52, 115-127.

Berndt E., Hall B., Hall R., Hausmann J. (1974) Estimation and inference in nonlinear structural models, *Annals of Economic and Social Measurement*, 3, 653-665.

De Luca G. (2001) La direzione dei prezzi di titoli azionari e i modelli ACD, *Quaderni di Statistica del Dipartimento di Economie, Società e Istituzioni, Università di Verona*, Quaderno n. 2.

Engle R., Russell J. (1998) Autoregressive conditional duration; a new model for irregularly spaced transaction data, *Econometrica*, 66, 1127-1162.

Gramming J., Maurer K.O. (1999) Non-monotonic hazard functions and the autoregressive conditional duration model, *Discussion Paper 50, SFB 373*, Humboldt University Berlin.

Jasiak, J. (1999) Persistence in intertrade durations, *Manuscript*, York University.

Rosenblatt, M. (1985) *Stationary sequences and random fields*, Birkhäuser, Basel.

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