

On the Detection of Long Memory in the Intradaily Pattern of Stock Returns

*Alla ricerca di memoria lunga
nell'andamento giornaliero dei rendimenti azionari*

Marco J. Lombardi

Dipartimento di Statistica “G. Parenti”

Università degli studi di Firenze

mjl@ds.unifi.it

Riassunto: Nell’analisi statistica dei mercati azionari, si usa a volte il campo di variazione giornaliero del prezzo come indice di volatilità, al posto della più diffusa varianza condizionata. Alcuni autori hanno di recente identificato la presenza di memoria lunga nella volatilità dei mercati finanziari e hanno sviluppato nuovi modelli che tengono conto di maggiore persistenza della varianza condizionata. Il nostro lavoro ha l’obiettivo di identificare e stimare la memoria lunga nell’andamento di un indice alternativo di volatilità, ovvero la differenza relativa giornaliera di prezzo.

Keywords: Volatility, Long memory, Fractional integration, GARCH processes.

1. Introduction

The concept of long memory originated in engineering and physics during the fifties, and was first used for statistical applications by Granger (1980) and Hosking (1981). It is well known¹ that, for standard ARMA processes, the autocorrelation function decreases exponentially. Intuitively, we may argue that a process possesses long memory if its autocorrelation function exhibits a “slower” pattern of decay. The most simple way to obtain a process with long memory features is to incorporate in the standard ARMA formulation the *fractional difference operator* $(1 - L)^d$, with $0 \leq d \leq 1$, and L the lag operator. The resulting process is known as ARFIMA, that is AutoRegressive Fractionally Integrated Moving Average:

$$\eta(L)(1 - L)^d(y_t - \mu) = \zeta(L)\epsilon_t. \quad (1)$$

The fractional difference operator may be expanded in a McLaurin series to produce an infinite polynomial in L :

$$(1 - L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k - d)}{\Gamma(k + 1)\Gamma(-d)} L^k = \sum_{k=0}^{\infty} \varphi_k(d) L^k, \quad (2)$$

which is more suitable for estimation purposes, since the φ_k coefficients have a known form ($\Gamma(\cdot)$ is the gamma function) depending on d . It is obvious that, in order to obtain a manageable expression to be used for estimation, (2) must be truncated for some sufficiently high number M (typically 1000). It may be shown (Hosking (1981)) that the autocorrelation function of ARFIMA processes decreases hyperbolically for $0 < d < 1$,

¹See, for example, Hamilton (1994) – page 50.

thus displaying a sensibly slower pattern of decay with respect to the standard ARMA case. The meaning of the term *fractional integration* is that, when $d = 0$, we obtain a standard ARMA process, whereas, for $d = 1$, the result is an ARIMA process of order 1. So, the use of the fractional difference operator allows us to obtain a continuum of possibilities between the polar cases of unit roots processes and of integrated processes of order 0. One drawback with ARFIMA processes is that the inclusion of μ in the parameters to be estimated with maximum likelihood may induce distortion in the estimates of the other parameters; it is thus preferable to set μ equal to the mean of the series for the ML estimation.

ARFIMA processes have been applied to several economic variables, such as inflation, unemployment or interest rates². Following a seminal paper by Ding *et al.* (1993), several authors focused on the detection and the estimation of long memory in the volatility of the returns of financial markets. In particular, Baillie *et al.* (1996), introduced a new model for the conditional variance by inserting the fractional difference operator into a standard GARCH (Bollerslev (1986)) formulation:

$$\psi(L)(1-L)^{d^*}\epsilon_t^2 = \omega + [1 - \beta(L)]w_t, \quad (3)$$

This model, known as FIGARCH, has been applied to several real-world time series, both from stock and from exchange rate markets, highlighting the presence of long memory in the conditional variance of the returns. There are obvious advantages in modelling the volatility process with a FIGARCH model; above all, that we can model and estimate simultaneously the mean and the variance equation. Yet, there are still some troublesome drawbacks: the asymptotic theory for FIGARCH processes is not well established yet, and moreover it is not clear whether the process is or is not stationary, nor whether the FIGARCH model is indeed a generalization of a standard GARCH – IGARCH framework.

No attention has been devoted, up to now, to the estimation of long memory in alternative indices of volatility. Among those alternative indices, one of the most widely used, because of its computational straightforwardness, is the technical indicator known as *Infra-Daily Volatility (IDV)*, considered, among the others, in Gallo and Pacini (1998). Let us denote with P_t^H and P_t^L , respectively, the highest and the lowest price observed during the trading day t . The *IDV* index is:

$$IDV_t \equiv \frac{P_t^H - P_t^L}{P_t} \approx \ln P_t^H - \ln P_t^L, \quad (4)$$

where P_t denotes the closing price on day t .

2. An empirical study

Our work is intended to estimate long memory in the Infra-Daily Volatility index for the daily returns of three different large-capital stocks traded on the U.S. Market (in parentheses we report the corresponding ticker): General Electric (GE), IBM (IBM) and Procter & Gamble (PG). The sample period we will work on is from January 2, 1975 to June 29, 2001 (6696 observations). The presence of long memory in the series considered may be argued by observing the correlograms, reported in Figure 1: the pattern of decay seems to be sensibly slower than exponential, with positive autocorrelations persisting well beyond the thousandth lag.

²See, for a complete survey of those applications, Baillie (1996).

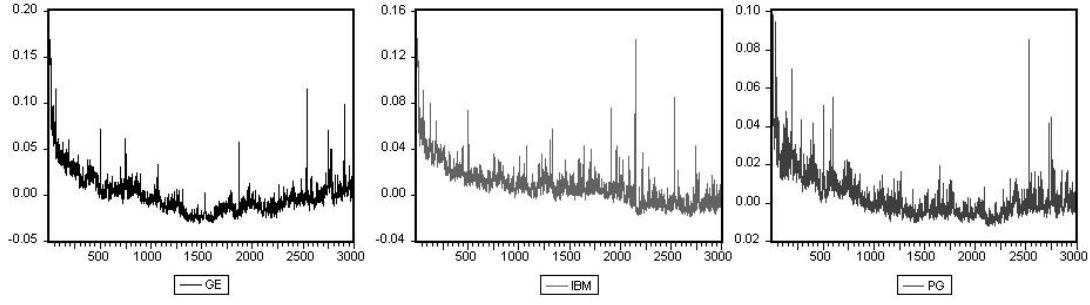


Figure 1: First 3000 autocorrelations for GE, IBM and PG in the above sample.

We have tried to fit the series with both an ARMA (1, 1) model and an ARFIMA (1, d , 1); the results are presented in Table 1.

Table 1: Estimation of ARMA (1, 1) model and ARFIMA (1, d , 1) models for the series presented (6696 observations). P-values for the coefficients are reported in parentheses.

	GE		IBM		PG	
	(1, 0, 1)	(1, d , 1)	(1, 0, 1)	(1, d , 1)	(1, 0, 1)	(1, d , 1)
μ	0.019692 (0.000)	0.019707 –	0.020197 (0.000)	0.020200 –	0.018231 (0.000)	0.018229 –
η	0.962455 (0.000)	0.338571 (0.000)	0.963214 (0.000)	0.484774 (0.000)	0.945134 (0.000)	0.601611 (0.000)
ζ	-0.791086 (0.000)	-0.520786 (0.000)	-0.783773 (0.000)	-0.634842 (0.001)	-0.736168 (0.000)	-0.738455 (0.000)
d	– –	0.410284 (0.000)	– –	0.404289 (0.000)	– –	0.427086 (0.000)
Loglik	21653.84	21707.16	21024.01	21113.37	21420.22	21534.93
LR test	106.64		178.72		229.42	
P-values	(0.000)		(0.000)		(0.000)	

The ARFIMA model performs clearly better, capturing the long-run dependency with the d parameter and thus displaying a reduction in the AR coefficient and a sensible increase in likelihood, as the LR tests show. We have tried to confirm our results by performing a KPSS test for long memory on the residuals of the ARMA model. This test, originally developed for testing stationarity, has been analyzed by Lee and Schmidt (1996) in order to test for long memory. Since we work with very long series, we have chosen, as suggested by Lee and Schmidt (1996), a bandwidth of $12 (T/100)^{1/4}$. We have tested for long memory the residuals of the original ARMA (1, 1) models an augmented-lag ARMA (10, 10). As Table 2 shows, the results are overwhelmingly in favor of the alternative hypothesis of long memory in the residuals and thus confirm our previous analysis.

3. Concluding remarks

We have shown evidence of long memory in the IDV index of three stocks traded on the U.S. market, fitting the series with both ARMA and ARFIMA models and showing the superiority of the latter.

Table 2: Long memory tests on the residuals of ARMA (10, 10) model and ARMA (1, 1). Bandwidth was set to $12 (T/100)^{1/4}$.

	GE		IBM		PG	
	(1, 1)	(10, 10)	(1, 1)	(10, 10)	(1, 1)	(10, 10)
KPSS test	1821.6	701.4	4283.537	687.751	4433.484	1118.752
P-values	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

One of the main advantages in modelling volatility with the *IDV* index is that, contrary to what happens for FIGARCH processes, both the asymptotic theory and the stationarity results for ARFIMA processes are clearly established. We indeed loose the possibility to estimate simultaneously the mean equation, but in the context of financial forecasts, this is a minor drawback: reliable forecasts on the sign of the returns are nearly impossible, so the operators often focus, for risk management and option pricing purposes, on the estimation of volatility.

References

- Baillie R. (1996) Long memory processes and fractional integration in econometrics, *Journal of Econometrics*, 73, 5–59.
- Baillie R., Bollerslev T. and Mikkelsen H. (1996) Fractionally integrated generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 74, 3–30.
- Bollerslev T. (1986) Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 31, 307–327.
- Ding Z.X., Granger C. and Engle R. (1993) A long memory property of stock market returns and a new approach, *Journal of Empirical Finance*, 1, 83–106.
- Gallo G. and Pacini B. (1998) Indicatori tecnici e volatilità di serie storiche finanziarie, *Statistica*, 58, 456–480.
- Granger C. (1980) Long-memory relationships and the aggregation of dynamic models, *Journal of Econometrics*, 14, 228–238.
- Hamilton J. (1994) *Time Series Analysis*, Princeton University Press, Princeton.
- Hosking J. (1981) Fractional differencing, *Biometrika*, 68, 165–176.
- Lee D. and Schmidt P. (1996) On the power of the KPSS test of stationarity against fractionally-integrated alternatives, *Journal of Econometrics*, 73, 285–302.