WORKING PAPER n.03.12

October 2003

Seasonality, memory and causality in the FIB30 market: a return-volume study

M. Caporin[1]

[1] GRETA, Venice

Seasonality, memory and causality in the FIB30 market: a volume-return study

Massimiliano Caporin
Unviersità di Venezia and GRETA Associati
mcaporin@unive.it

October 8, 2003

Abstract

In this paper I analyse the characteristics of the returns and of the volume of the FIB30 market. I focus at first on the identification of seasonal patterns, using the Flexible Fourier Regressions techique applied by Andersen and Bollerslev. Then, on the filtered data, I verify the long memory properties of the volume mean and of the volatilites of both the log-returns and the volume series. Finally, I move to a multivariate setting fitting a bivariate GARCH-type model in order to verify the causal relations among the mean and the variances of the considered variables. In this stage I fitted both traditional models and a new parameterisation, the EC-GARCH, I suggested in Caporin (2003). This new model is clearly preferred by standard information criteria and provides evidence of a causal relation among the variances of the volume on the log-returns.

In the last years there has been a growing interest in the study of the relation among prices and volumes, both from a theoretical point of view (as an example Blume, Easley and O'Hara, 1994) and from the empirical one, see among others Karpoff (1987). Most of the current empirical analysis considers different linear and non-linear specifications in order to verify and test the causal relations between prices (or returns) and volumes. However, most of them consider only the mean, restricting their attention on Granger's causality definition, alternatively they foc on the study of a simultaneous relation. In the last two decades, with the emerging ARCH literature, different specifications of conditional heteroskedasticity have been taken into considerations; they allow for a deeper analysis in applied studies on the causality topic, an effort that allows to adequately model the relation between returns, volumes and their volatility. These extensions can be thought both on a signle-asset case that in a much more general multivariate framework. This interest on multivariate heteroskedastic models maybe also coupled with the necessity of an extension of the causality concept, which must consider the spillover effect among variances, and the in-mean reaction of GARCH model components. A discussion on this topic was the object of a previous work (see Caporin (2003a and 2003c)) where

I considered the theoretical aspects of the second order causality and of the variance causality together with their link with the well known concept of first order causality, i.e. in Granger's sense. In that studies I focus on the present definitions and relations providing the theoretical conditions for non-causality in mean and in variance in a general VARMA-GARCH-M model. Moreover, I introduced a new GARCH-type model, the Exponential Causality GARCH (EC-GARCH) which allows for the detection of variance causality without any restriction on the parameters driving the causal relation between variables.

The aim of this work is twofold: on the one side, I analyse the seasonal components which affect the returns and the volumes of the FIB30 market; on the other side, I consider their memory and causality structure. Consequently, this paper deals almost exclusively with empirical analysis using the available techniques for the first part, while in the second it concentrates on long memory models belonging to the ARFIMA and/or FIGARCH class, together with the EC-GARCH to verify the causal relation among variances.

Given the empirical purposes of this work, a brief theoretical introduction on the models used will be included in the appendix, while the definition on the second order causality topic can be found in Caporin (2003c). The plan of the paper is as follow: in section 1 I synthetically describe the FIB30 market while in section 2 I analyse the dataset and I indetify and delete the outliers; section 3 is devoted to the estimation of the cyclical patterns in both the returns and volume; the following two sections, 4 and 5, consider the univariate and multivariate estimation of returns and volume, section 5 deals also with the variance causality; section 6 will conclude.

1 A (brief) description of the FIB30 market

In 1994 the Italian Stock Exchange moved to an automatic transaction system, by the end of that year a new market segment was created, to allow trading on derivatives and among these on the future on the stock market indices. In the Italian Exchange there exist four indices: the MIBTEL built on all the traded stocks, the MIB30 that consider only the thirty firms with higher capitalization and trading, the MIBR that consider the twenty stocks with higher capitalization among the ones excluded from the MIB30 and finally the NUMTEL the Italian correspondent of the NASDAQ, which group information, technology and communications firm recently entered into the market. In 1994 a future contract on the MIB30 index was launched; the objective of this operation was to obtain an increase in the transactions and in the efficiency of the stock market, both objectives were reached. Contract characteristics are the followings (valid up to the end of 2001): there are four contract maturities in March, June, September and December; the contract clear the third Friday of the maturity month at 9:30 AM or next day of open market if Friday is holiday; there are four traded maturities all over the year; last day of negotiation is the maturity day; the closing price is fixed by the Clearing House with respect to the MIB30 index at 9:30 AM (opening price); regulation of the contract by cash the first open market day after maturity; contract nominal value is determined with respect to the MIB30 index, each point of the index is worth 5 euros and the minimum price movement is of 5 index points; transactions last from 9:15 AM up to 17:30 PM (this up to the end of 2001, from January 2002 the market closes at 17:40 PM). It is worth to note that the main market open at 9:30, this mean that the future prices between 9:15 and 9:30 anticipate market movements since they respond to the information released during closed market periods (weekend or just the night), therefore we have could an increase in volumes and high variations in returns in this limited range. Alternatively, the operators could wait the opening of the main market before trading in the future to observe how the stocks react to information shocks. In both cases volume and prices (and therefore returns) of the future may be biased and may present abnormal movements. The selection of outliers and the filtration of the cyclical components will be analyzed in the following sections.

2 Data description and outliers deletion

The database used in this study was provided by the Italian stock exchange (Borsa Italiana S.p.A.). It contains approximately one year of transactions concluded on the FIB30 contract, the future on the Italian stock exchange index (MIB30), see previous section for details about the market and the index. The supplied data ranges from the 13th march 2000 to the 20th march 2001 for a total of 260 open market days. As specified in the previous section, contracts are traded with four different fixed maturities (mid March, June, September and December), with the possibility to trade on the next four maturity dates. This mean that at any given time there are, possibly, four contracts traded. In this analysis I concentrate the attention on the contract with the closest maturity, that is also the most traded one. It covers about the 95% of the market, apart in the last week of its life, when the trading in the next to maturity contract increase. This fact represents a problem in dealing with such a database, I have to check if it is necessary to exclude the last days of a contract life to avoid noisy prices, that may be biased by the roll-over process. I will return on this problem in a few steps. The data provided are "transaction" data, that is they record all contracts concluded, i.e.for each concluded contract the database contains: the current date; an identifier of the contract and its ISIN code, that are different for each maturity and instrument traded; the trading and clearing time, that is when contract is entered in the system and when it is matched with a counterpart; the contract number, price and finally the volume, the number of futures exchanged (see figure 1 for a snapshot of the row data).

Even if the Italian stock exchange market is not one of the biggest european ones (it represents only 2% of worldwide stock markets capitalization), this boils down in an enormous amount of data, more than 20 millions of informations, with peaks of more than 15000 contracts signed in a trading day, see table 2 for a summary of descriptive statistics. However this dataset cannot be used for all market microstructure analysis since bid/ask spread, the sign of the contract and the identifier of the subject that trades the contract are missing. Even with this limitation the amount of information contained is still considerable, and useful for the purposes of the present work. Given that I concentrate on the prices of the closest maturity contract, I have to solve another problem: I must specify how to treat the rollover days, where the most traded contract expires and there is a change in the average price of the futures. It is well known that a future price get closer to the price of the underlying as it get closer to the maturity, while the price of the next-to-maturity contract include also expectations about the underlying and the interest rates movements. Different solutions for obtaining return series from future prices deal with this problem and suggest alternative strategies, from shifting the prices series of the delta between the two contract at the maturity date, or just adjusting return with a factor. I will not follow any of these suggestions but I will statistically test the relevance of the returns in the maturity date: I will add an impulse dummy variable for each maturity date (four in the considered sample) and then test its significativity. This additional variable will, eventually, exclude the return across the maturity date without introducing a bias on the structure of the underlying process.

Date yyyymmdd	ISIN code and identifier of the contract	Contract trading and clearing time	Contract number	Contract price	Volume
⋮					
20010201	IT0002083829 MIB30C1	115245 115245	105106	44720.0000	4
20010201	IT0002083829 MIB30C1	115245 115245	105107	44720.0000	1
20010201	IT0002083829 MIB30C1	115245 115245	105108	44730.0000	1
20010201	IT0002083829 MIB30C1	115252 115252	105113	44710.0000	1
20010201	IT0002083829 MIB30C1	115254 115254	105114	44730.0000	1
20010201	IT0002083829 MIB30C1	115254 115254	105115	44735.0000	1
20010201	IT0002083829 MIB30C1	115254 115254	105116	44735.0000	3
20010201	IT0002083829 MIB30C1	115254 115254	105117	44735.0000	3
20010201	IT0002083829 MIB30C1	115254 115254	105118	44740.0000	1
20010201	IT0002083829 MIB30C1	115254 115254	105119	44740.0000	2
20010201	IT0002083829 MIB30C1	115255 115255	105120	44740.0000	1
20010201	IT0002083829 MIB30C1	115255 115255	105122	44740.0000	1
⋮					

Figure 1: extract from the dataset provided by the Italian Exchange

The database records transaction series, for the analysis I need to transform them into equally spaced non-overlapping series. I decided to run the analysis on the 5 minute interval, a smaller tick frequency may not show the causality relation, while a longer one may destroy the microstructure and the correlation and causality between volume, returns and their volatility, resulting in a contemporaneous relation. Converting transaction data into five-minute observations is not a problem for the volume series: I have just to sum over the contract traded every 5 minutes. For the prices this is not so simple, I could in principle take the last price of all 5 minute intervals, take a weighted average over the 5 minutes, take the median, the simple mean and other combinations or means. However, the main problem is to avoid biases due to the averaging therefore, I chose to consider the last price of the interval, since it will include all the relevant information impacts, then I price all the volume traded in the interval at this last observed price.

Day	Motivation of exclusion
26 April 2000	Market was open up to 18 PM due to technical problems during the day
5 July 2000	Market was open up to 19 PM due to technical problems during the day
14 August 2000	Abnormal movements due to very low trading
28 August 2000	Market opened at 10 due to technical problems
19 February 2001	Market opened at 12:30 due to technical problems

Table 1: days removed from the sample and motivation

Another point concerns the first 15 minutes of open market, from 9:15 to 9:30: I decided to compare the inclusion and exclusion of the first 15 minutes observing the estimated patterns and a set of descriptive statistics. This paper presents some of the graphs and analysis. The full comparion among the two periods is included in Caporin (2002a). The first 15 minutes, in particular the range 9:15-9:20 concentrate abnormal volume and price movements due to the impact of news released during closed market hours. In this view I introduce another dummy variable that excludes the first return of each day in order to test the close-to-open return nad volume relevance. A significant value in this dummy is expected since most of announcements of the central banks and of bigger firms are released when markets are closed. Moreover, it will be significant for the relevance of the price movements in the range 9:15-9:30. These announcements affect the underlying stock market index and the term structure of interest rates, resulting therefore in movements also in the future prices. A non-significant dummy will reveal that the news released do not heavily affect market prices probably because they are discounted well in advance by the whole market. In the following I will discuss some descriptive statistics computed with or without the range 9:15-9:30.

A further problem is considered in the preliminar analisys of the dataset. Some days in the sample are completely deleted because the market showed anomalous trading or anomalous opening due to technical problems (for details see table 1).

Particular attention has then been given to outlier detection and exclusion; I presume that they are mainly due to operational errors. Part of them are identified observing clearing time, abnormal high price sell contracts and low price buy contracts are generally not executed, this mean a clearing time of zero, but the problem is with abnormal low price sell contracts and high price buy contracts that may be automatically matched with counterparts once entered into the system. In this last case the outliers are detected by a procedure that tests the presence of an operation of opposite sign in the next few seconds, and in that case delete both. However, the correction may not be the next operation, alternatively there cannot be a correction at all. Therefore, there remain other outliers. Normally, these are identified and deleted by a procedure that is run on a daily basis: at first the (daily) standard deviation of the log-returns is computed, then outliers are defined as the returns outside two bounds determined as three times the standard deviations (positive and negative), and are then removed; the procedure should also check for jumps in the series, that could be detected as outliers. Neverthless, I follow this alternative procedure: assume that we have a presumed outlier p_t, then we take the 10 precedent contract prices and compute their mean $\mu_{t-1,t-10} = (10)^{-1} \sum_{i=1}^{i=10} p_{t-i}$; we remove this mean from the 10 contracts and the suspect outlier price we are analyzing and consider the absolute value of these differences $d_{t-i} = \left|p_{t-i} - \mu_{t-1,t-10}\right|$ $i = 0, 1...10$; we compare the presumed outlier difference d_t with the biggest difference among the 10 precedent prices $d_{10} = \max\left(d_{t-i} \quad i = 1...10\right)$, the contract is then deleted if $d_t > 3d_{10}$. The procedure controls also for jumps in the price series that, following the previous methodology, may be detected as outliers in the jumping contract price. This methodology is based on the prices preceding any possible outlier, however to allow for an easier identification of jumps the same reasoning can be based on the 10 successive prices. I compared these two alternative methodologies in a limited sample and I get absolutely no difference on the identified outliers, therefore, given the gains that could be attained in the software implementation (these procedures are really time consuming in a database consisting of millions of points), I preferred basing the algorithm on the 10 successive prices.

The choice of this approach was motivated by the interest in preserving extreme events, an example can be seen in Figure 2 where I report the observed log-returns for one day in the sample, the 20 February 2001. In the graph I report also the bounds computed as $3x$standard deviations. Differently, in Figure 3, I report the prices with the outliers evidenced. Using the suggested procedure, only four outliers (the first on the left is a group of two outliers) are detected and deleted, while using the "standard" methodology much more point should have been deleted, incurring in the risk of eliminating extreme events. The difference among the two procedure is, however, really minimal, in the sense that I identified outliers in transaction data which in the following step will be aggregated and averaged over a 5 minute interval, reducing therefore the possible effects of maintaining in the dataset an outlier or of excluding an extreme event.

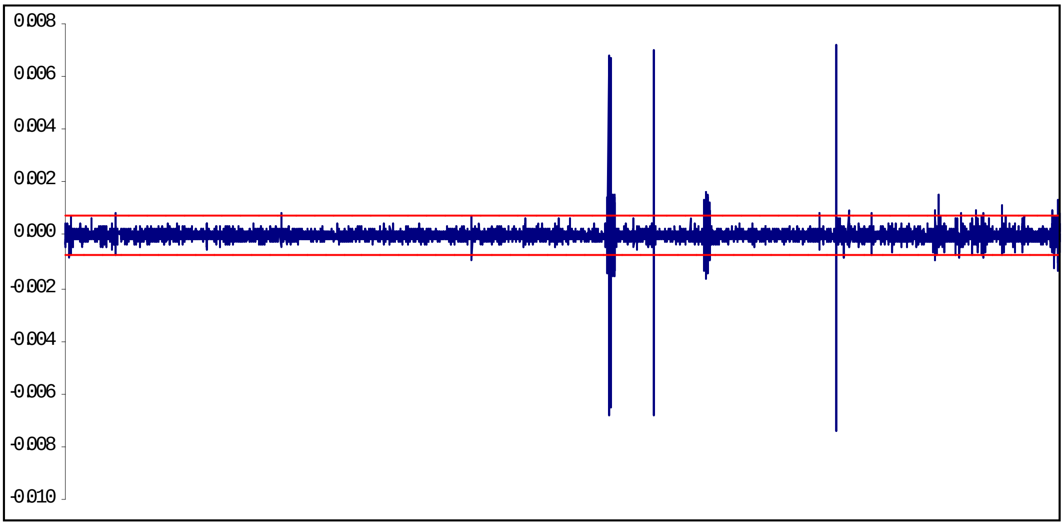


Figure 2: Log-returns and daily bounds for outlier detection

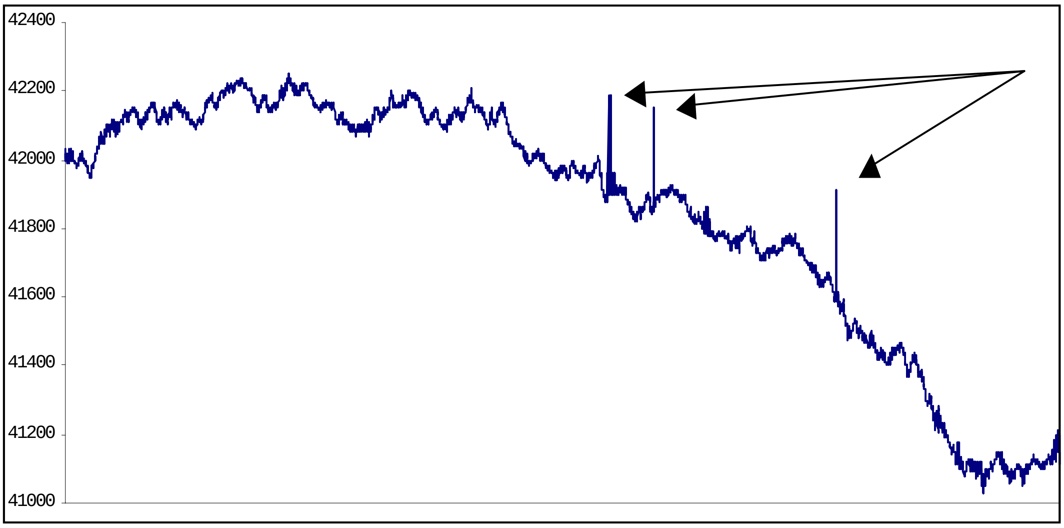


Figure 3: Prices with the outlier evidenced by our procedure

Tables from 2 to 4 show a group of descriptive analysis that compares returns and volume across the quarters. For the returns, I concentrate directly on the 5-minute series, derived from prices observed at the end of each of the 5-minute intervals. Table 3 compares a set of moments and extreme values between the return series that include or exclude the period 9:15-9:30. As we can see, the main changes appear on the sample skewness and kurtosis because we are including returns belonging to the queues of the distribution: deleting the first 15 minutes of open market the first return of each day will be computed with respect to the last price of the previous day, resulting most of the times in a bigger value (in absolute terms) since the news released when markets were already closed will affect market specially between 9:15 and 9:30.

In tables 2 and 4 I concentrated the attention on volume to justify the choice of the contract with closest maturity as the reference contract. We can see that the number of contracts traded in the future with next to closest maturity is around 5-8% of the most traded future. Moreover this behavior is stable during all the year considered. Table 2 allows to make a set of considerations on the volume traded in this derivative market, more that 3,8 millions contract were signed, with peaks of more than 37 thousand per day. The period with lower activity is the one that range from 16/06/00 to 14/09/00, not a surprising behavior. This pattern becomes much more clearly in figure 4, where the traded volume of the two contracts of table 2 is shown. In this same table I compare also the effect of the exclusion of the first 15 minutes and of the days listed in table 1. Table 4 is the volume descriptive analysis in the five minute intervals, showing the effects of removing the first 15 minutes of open market: as expected, skewness and kurtosis goes in the direction of normality since we are deleting values that are in the queues of the distributions.

The Figures 5 and 6 report the whole sample data after the outlier deletion.

.

Quarters	17/03-15/06/2000			16/06-14/09/2000		
Contract	15/06	15/06 adj.	14/09	14/09	14/09 adj.	14/12
Number of days	62	60	62	64	61	64
Mean volume	17362	16171	858	11321	10554	799
Volume s.d.	3781	3594	2117	3161	2759	2548
Minimum	6586	5662	15	2624	6354	10
Maximum	27323	26016	12190	24801	21956	17631
Volume	1059058	970249	52335	724551	643765	51133
Quarters	15/09-14/12/2000			15/12/00-15/03/01		
Contract	14/12	14/12 adj.	15/03	15/03	15/03 adj.	14/06
Number of days	65	65	65	62	61	62
Mean volume	15728	14664	667	15902	14900	860
Volume s.d.	3540	3433	2070	5706	5310	2349
Minimum	6475	5777	5	4451	4088	8
Maximum	24514	23362	11196	37168	34698	13526
Volume	1022333	953163	43380	985907	908921	52455

Table 2: descriptive statistics of the daily volume across the quarters comparing the two most traded contracts and the effects of data cleaning, i.e. identification and deletion of outliers and days wih abnormal movements (contracts recorder between 9:15 and 9:30 are excluded)

Quarters	17/03-15/06/2000		16/06-14/09/2000	
9:15-9:30	with	without	with	without
Mean	-4,62E-06	-4,45E-06	9,20E-07	9,49E-07
Standard dev.	0,00168	0,00173	0,00084	0,00086
Kurtosis	40,11631	50,92920	12,97777	15,05628
Skewness	-1,95357	-2,14094	0,58585	0,78776
Minimum	-0,03441	-0,03713	-0,00621	-0,00621
Maximum	0,01215	0,01828	0,01263	0,01263
Quarters	15/09-14/12/2000		15/12/00-15/03/01	
9:15-9:30	with	without	with	without
Mean	-1,13E-05	-1,16E-05	-2,61E-05	-2,69E-05
Standard dev.	0,00118	0,00119	0,00134	0,00135
Kurtosis	12,48523	14,27727	24,84416	24,67570
Skewness	0,01205	-0,23917	0,30956	-0,07730
Minimum	-0,01465	-0,01532	-0,01387	-0,01733
Maximum	0,01126	0,01256	0,02349	0,02095

Table 3: descriptive statistics of the 5 minute returns series across quarters comparing the effects of excluding or not the contracts concluded between 9:15 and 9:30

Quarters	17/03-15/06/2000		16/06-14/09/2000	
9:15-9:30	with	without	with	without
Mean	174,222	168,446	114,091	109,933
Standard dev.	132,624	125,727	104,451	99,130
Kurtosis	3,346	2,631	12,546	11,587
Skewness	1,528	1,411	2,531	2,427
Minimum	1	1	0	0
Maximum	1073	948	1324	1324
Sum	1034879	970249	688997	643765
Quarters	15/09-14/12/2000		15/12/00-15/03/01	
9:15-9:30	with	without	with	without
Mean	157,460	152,750	160,315	155,212
Standard dev.	130,888	126,293	141,693	135,341
Kurtosis	5,204	4,243	13,808	14,023
Skewness	1,776	1,702	2,386	2,306
Minimum	1	1	1	1
Maximum	1431	1147	2131	2131
Sum	1013256	953163	968141	908921

Table 4: descriptive statistics of the 5 minute volume series across the quarters comparing the effects of excluding or not the contracts concluded between 9:15 and 9:30

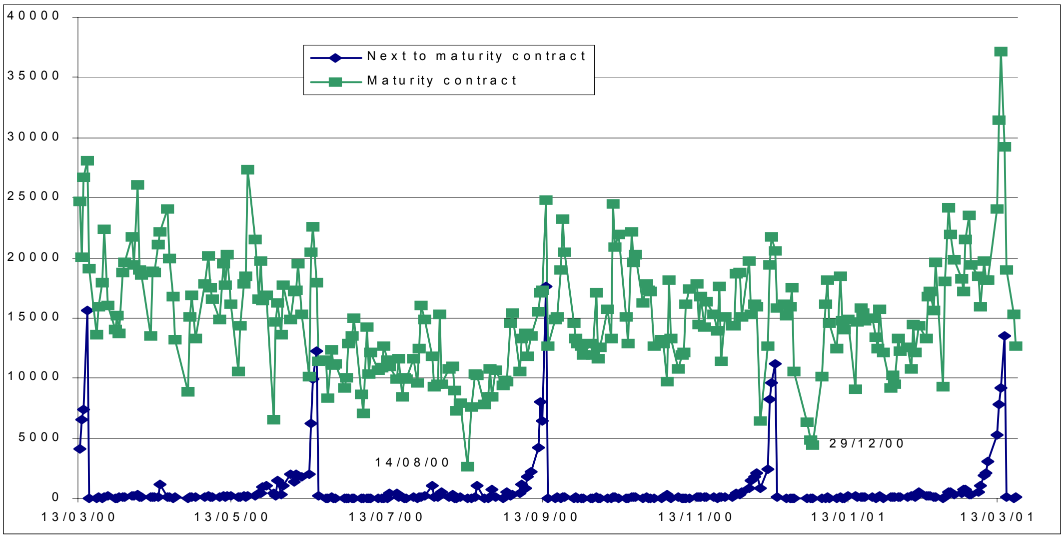


Figure 4: volume in the two most traded contracts

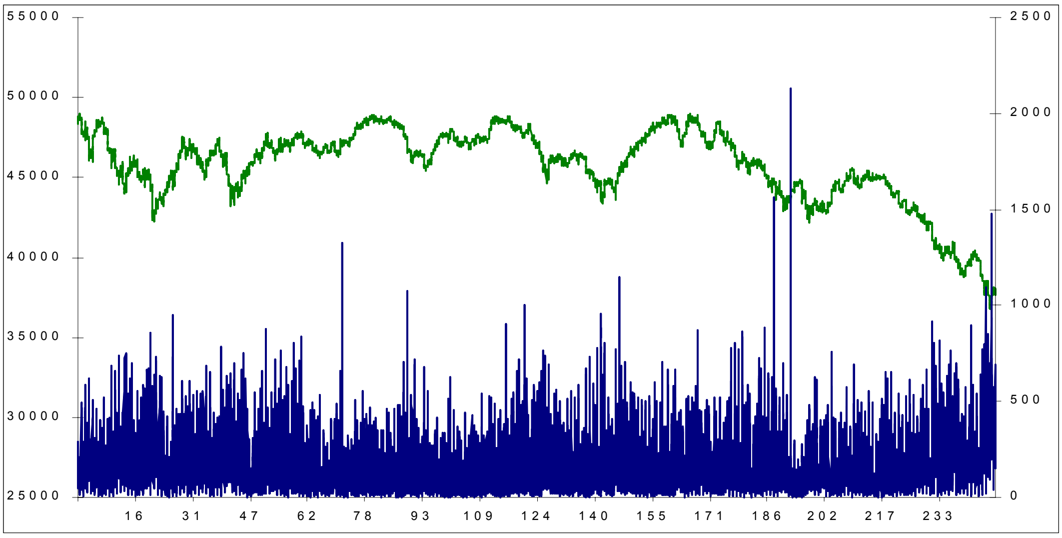


Figure 5: 5 minute prices and volumes

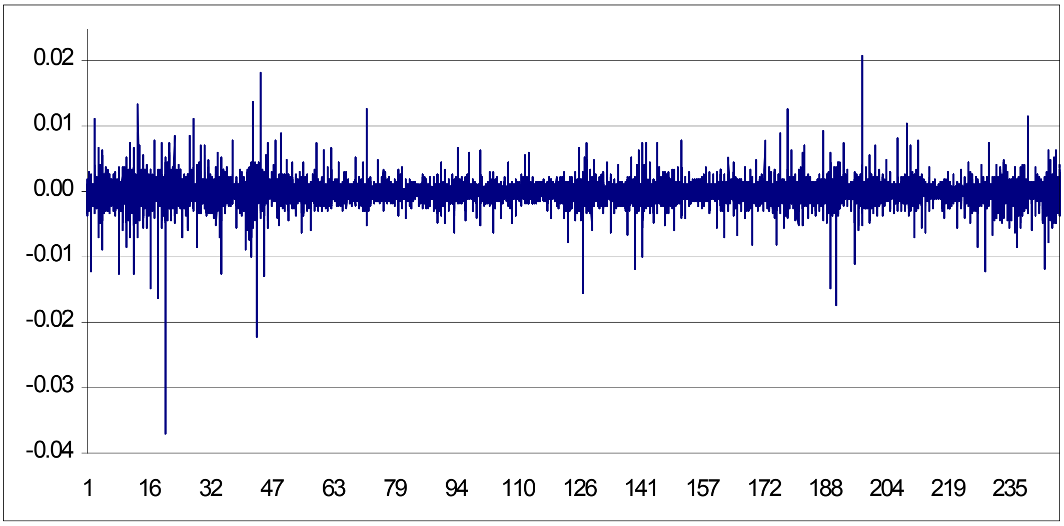
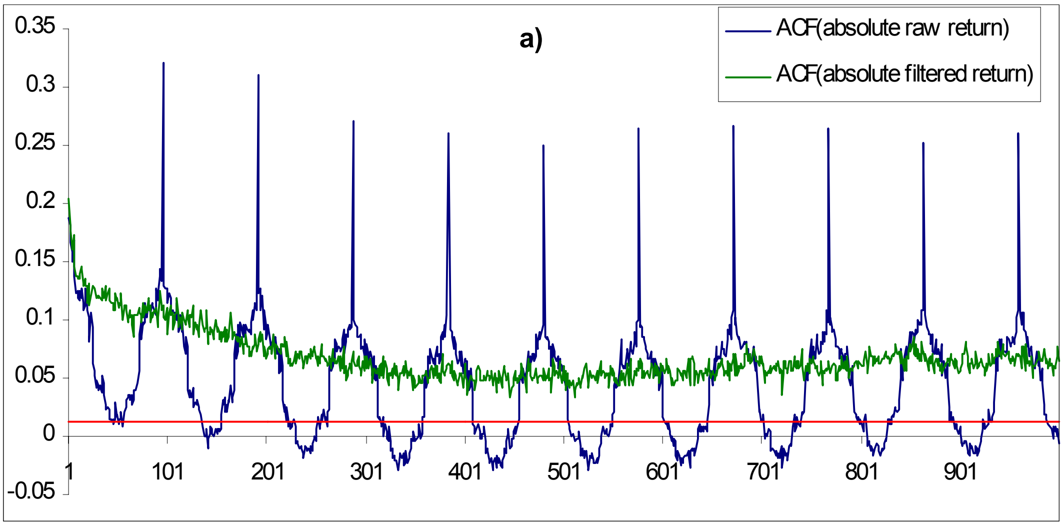


Figure 6: 5 minute returns

3 Filtering cyclical patterns

This section focuses on the identification of cyclical patterns in the mean and in the volatility of both returns and volume, following the approach of Bollerslev and Andersen (1997). I will study in detail at first the return series (contracts recorder from 9:30 up to 17:30) and then the volume.



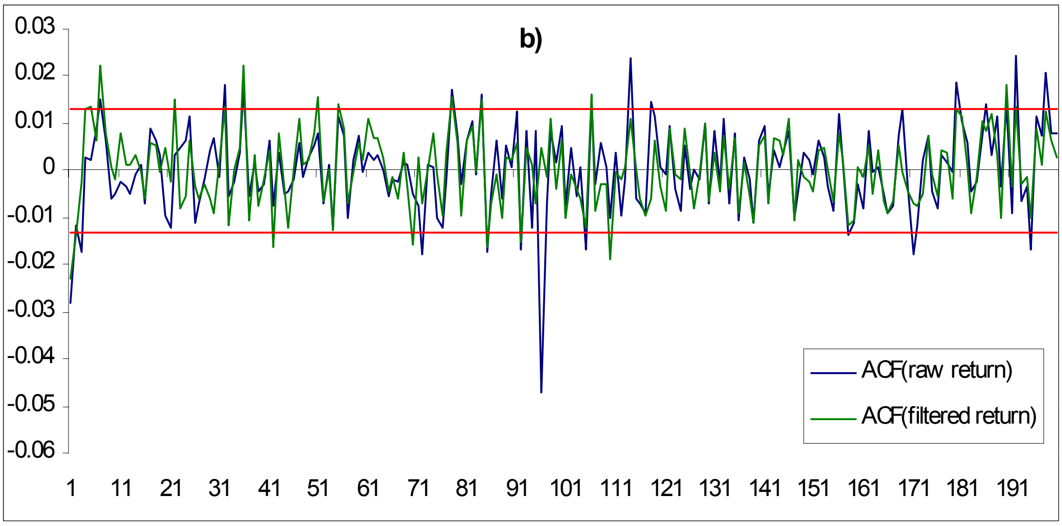


Figure 7: autocorrelation function of the 5 minute returns

Figure 7 shows the sample autocorrelations computed on 5-minute returns (200 lags considered) and on their absolute value (1000 lags considered). As expected, the mean dynamic can be explained by a small order ARMA model, given the limited number of significant sample autocorrelations. The first panel clearly evidence a cyclical pattern, that is responsible also for the peak in panel b) at the 96th lag. The structure of the oscillation indicates a combination of different cyclical components acting at different frequencies (note the peaks at the top of the sinusoidal pattern). The very same observations can be derived for the series that include the contracts concluded between 9:15 and 9:30, the only difference is that the peaks are at the 99th lag and multiple. Following Andersen and Bollerslev (1997) in figure 8 I verify the presence of this periodical behavior. Note that the first 5-minute interval is not represented in both graphs given its very high value; plotting it the cyclical patter will not have been so evident.

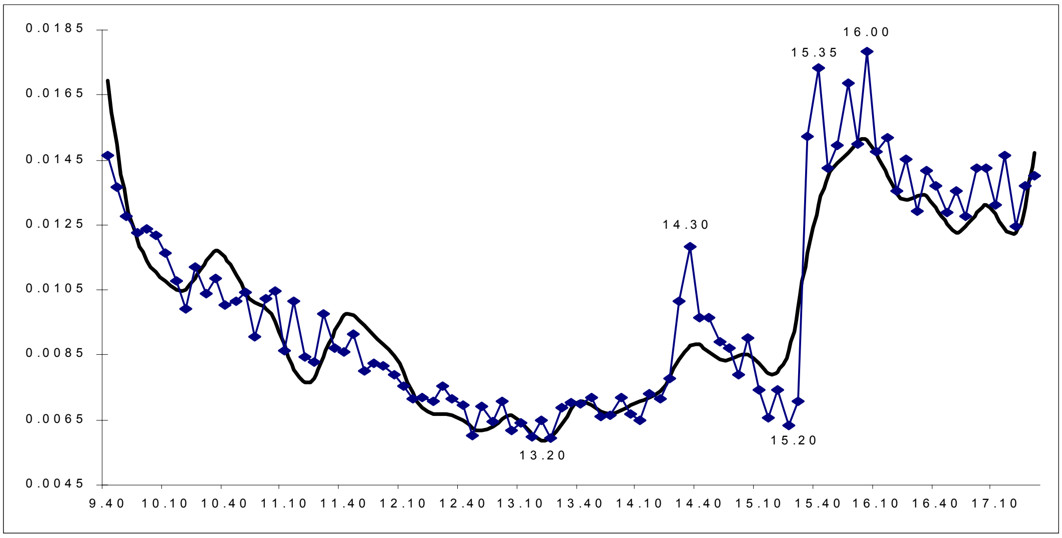


Figure 8: cyclical pattern in the absolute returns 9:35 - 17:30

The graphs represents the average absolute return across the five minute intervals in the sample: the open market day last for 8 hours (in figure 8), so 96 (let me defite it as the N index) ordered five minute intervals are considered; for each of these intervals I computed the mean across the days in the sample (247=T), and graphed them:

$$\hat{\mu}_n = \frac{1}{T}\sum_{t=1}^{T} x_{t,n} \qquad n = 1, ...N \tag{1}$$

The graphs clearly show a cyclical pattern, the well known U effect, returns are high when market opens and closes, they decrease during the day up to 2-3 PM and then increase. This effect is well documented for stock market returns and volatility. Two additional effects are noted: there is a peak after lunch, due to an increase in the trading after the break, that last for about an hour, than the trading decreases as quickly as it increased, until the first signals come from the American markets: in fact we can observe (note the timing in the graphs) that returns and volume react half an hour before the opening of the New York stock exchange. Graph 8 shows the pattern computed with (1), while the bold line is the estimated pattern.

The seasonal component can be filtered by the Flexible Fourier Form regression, see the appendix for a brief description of this technique. Flexible Fourier form is a deterministic filtration procedure, I do not consider any stochastic cyclical specification since the behavior of this periodic patter is very evident from data, moreover in a recent paper Beltratti and Morana (2000) compared the two methodologies and their choice was in the direction of the FFF, given the very close results. As in the cited papers the returns filtration is based on the following representation

$$r_{t,n} = E\left[r_{t,n}\right] + \frac{\sigma_{t(,n)} s_{t,n} Z_{t,n}}{\sqrt{N}} \tag{2}$$

where the return $r_{t,n}$ at day t and interval (5 minute) n is the sum of its expected value plus a term that depend on a daily (or intra-daily) volatility $\sigma_{t(,n)}$, on an error term $Z_{t,n}$ and on a function $s_{t,n}$ that explain the cyclical behavior. After a transformation on the data a Fourier regression is run, considering the harmonics reported in table 5.

Parameter	Component correspondence	Time correspondence
α_1, α_2, α_3	Quadratic component	---
γ_1	Dummy for opening	---
γ_2, γ_3, γ_4	Dummies for maturity	---
δ_1, ϕ_1	Harmonic of period 1	8 hours (1 day - 96 5 minute intervals)
δ_2, ϕ_2	Harmonic of period 2	4 hours (48 intervals)
δ_3, ϕ_3	Harmonic of period 3	2 hours and 40 minutes (32 intervals)
δ_4, ϕ_4	Harmonic of period 4	2 hours (24 intervals)
δ_6, ϕ_6	Harmonic of period 6	1 hour and 20 minutes (16 intervals)
δ_8, ϕ_8	Harmonic of period 8	1 hour (12 intervals)
δ_{12}, ϕ_{12}	Harmonic of period 12	45 minutes (9 intervals)
δ_{16}, ϕ_{16}	Harmonic of period 16	30 minutes (6 intervals)

Table 5: harmonics and correspondence with time, 96 interval per day, 9:30-17:30

The filtered returns are obtained by the following expression:

$$r^{f}_{t,n} = \frac{r_{t,n} - E\left[r_{t,n}\right]}{s_{t,n}} \tag{3}$$

Given the flexible representation of the FFF I added also a group of dummy variables to take into account and test the effect due to the close-to-open movements (the effects of the first 15 minutes of open market will also be included) and to the maturity dates: I added 3 dummies to consider the three maturity dates included in the sample (impulse dummy) and an additional one to consider the daily opening effect, in this last case the dummy get a value 1 for the first return of each day, a significant value of its associated coefficient implies that the information flow during closed market period impacts on prices on the first return of the day. The estimation results are summarized in table 6, where the parameters, standard deviation and significativity tests are considered. A graph of the estimated cyclical component is reported in figure 8, while a graph of the autocorrelation function of filtered data in included in figure 7. The harmonics listed in table 5 have been chosen with a general to particular selection procedure, starting form a FFF regression with harmonics up to a period of 24 and iteratively deleting the less significant harmonic. Using the whole sample the harmonics change because there will be 99 intervals per day. However, the cyclical patterns are not heavily influenced by this fact.

Parameter	Estimate	Std. Err.	t-value	Parameter	Estimate	Std. Err.	t-value
α_1	-7,501	0,938	-7,994	δ_3	-0,230	0,061	-3,774
α_2	-10,531	2,712	-3,884	ϕ_3	-0,086	0,034	-2,526
α_3	3,478	0,903	3,851	δ_4	-0,111	0,039	-2,871
γ_1	1,430	2,550	0,561	ϕ_4	0,101	0,030	3,417
γ_2	2,267	2,550	0,889	δ_6	0,069	0,027	2,575
γ_3	2,415	2,550	0,947	ϕ_6	-0,128	0,026	-4,911
γ_4	2,462	0,234	10,533	δ_8	-0,031	0,025	-1,282
δ_1	-1,309	0,530	-2,470	ϕ_8	0,110	0,025	4,455
ϕ_1	-0,405	0,079	-5,134	δ_{12}	-0,030	0,024	-1,241
δ_2	-0,569	0,132	-4,309	ϕ_{12}	0,046	0,024	1,922
ϕ_2	-0,171	0,044	-3,882	δ_{16}	0,014	0,024	0,576
				ϕ_{16}	0,067	0,024	2,826

Table 6: estimation results for the return cycle, period considered 9:30-17:30

Statistics	9:30-17:30		9:15-17:30	
	Raw	Filtered	Raw	Filtered
Mean	-1,056e-5	-9,661e-4	-1,032e-5	-0,00081
St. Dev.	0,00131	0,11353	0,00129	0,11756
Kurtosis	-1,223	-0,165	-0,938	-0,163
Skewness	50,772	6,831	41,303	6,981
Q(5)	29,342	25,178	10,372	24,325
Q(10)	38,923	39,770	18,523	39,721
Q(20)	46,335	44,511	25,837	45,740
Q(50)	84,085	98,332	63,290	103,646
Q(100)	219,632	176,500	152,049	178,770
$Q^2(5)$	57,562	2553,908	82,956	2624,742
Q^2 (10)	92,361	3848,511	154,235	3926,159
Q^2 (20)	152,718	5887,090	237,537	5872,836
Q^2 (50)	190,973	10751,499	302,345	10837,267
Q^2 (100)	1441,911	17024,438	1255,376	17018,321
JB	2266576,7	14713,698	1501913	16367
LM(2)	22,291	17,141	3,769	17,246
LM(5)	30,194	25,759	10,538	24,731
LM(10)	40,042	42,102	18,696	41,764
LM(20)	47,754	46,477	26,272	47,294

Table 7: moments and test of 5 minute returns

Finally, in Table 7 I compare the moments and the tests for correlation and ARCH effects computed on the raw and on the filtered return series (the table shows considers also a filtration on the whole sample to compare the results with the strategy I chose). After removing the periodic component, a long memory effect appears in the volatility of the 5 minute returns: note the smooth convergence toward zero of the autocorrelation of absolute returns, they are significant for all the lags graphed in Figure 7. The Portmanteau test computed on squared observations reacts to this emerging pattern while Engle LM test is not influenced. The filtration of periodic components shifts also the distribution of the returns toward normality, see Figure 10 which reports a kernel density estimate of 5 minute returns distribution before and after the filtration, compared with a normal distribution. The graph plots standardized returns (filtered and not) to avoid scale problems. A final comment on table 7 concerns the long run correlation evidenced by the Lijung-Box statistic: it seems that the exclusion of the observations recorded in the range 9:15-9:30 create additional correlation in the filtered series. This fact can be interpreted as an evidence of a true long memory pattern biased by abnormal movements or by a direct effect of including jumps and breaks in an underlying autocorrelated process. My preference is for the second hypothesis and for this reason the model estimated of the following section will be based on the whole sample.

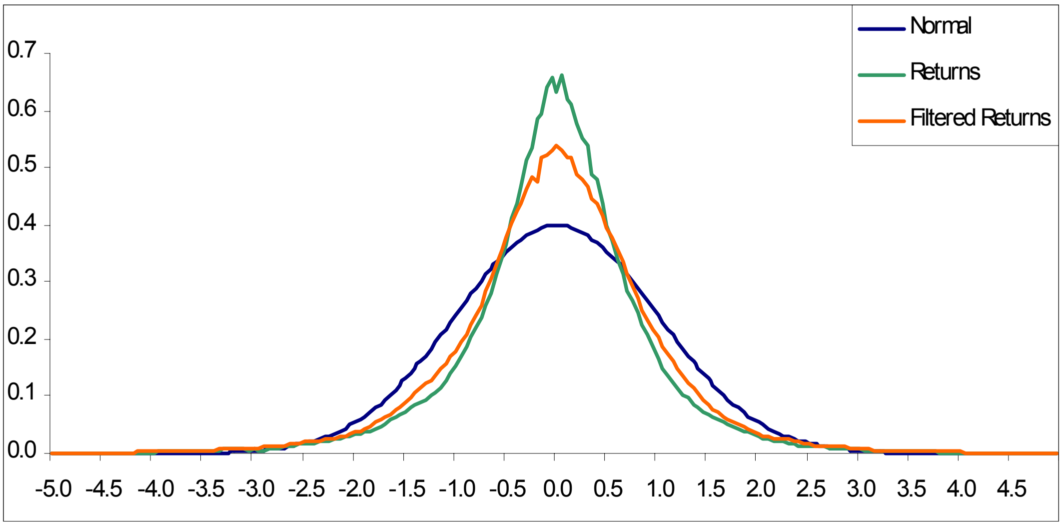


Figure 9: Kernel density estimate of 5 minute returns, period 9:30-17:30

Before turning to the analysis of the volume series a couple of comments on the estimated parameters of Table 6: the dummies reflecting the maturity date effect are not significant, probably they will have a greater impact in the mean than in the variance; as expected the dummy introduced to take into account the correction for contracts concluded between 9:15 and 9:30 is highly significative, therefore news impact affect market movement at the opening; all the periodic components included in the FFF regression resulted to be significant; moreover thay remove the greater part of the periodic pattern from the data. This indicates that the choice of a deterministic procedure was correct.

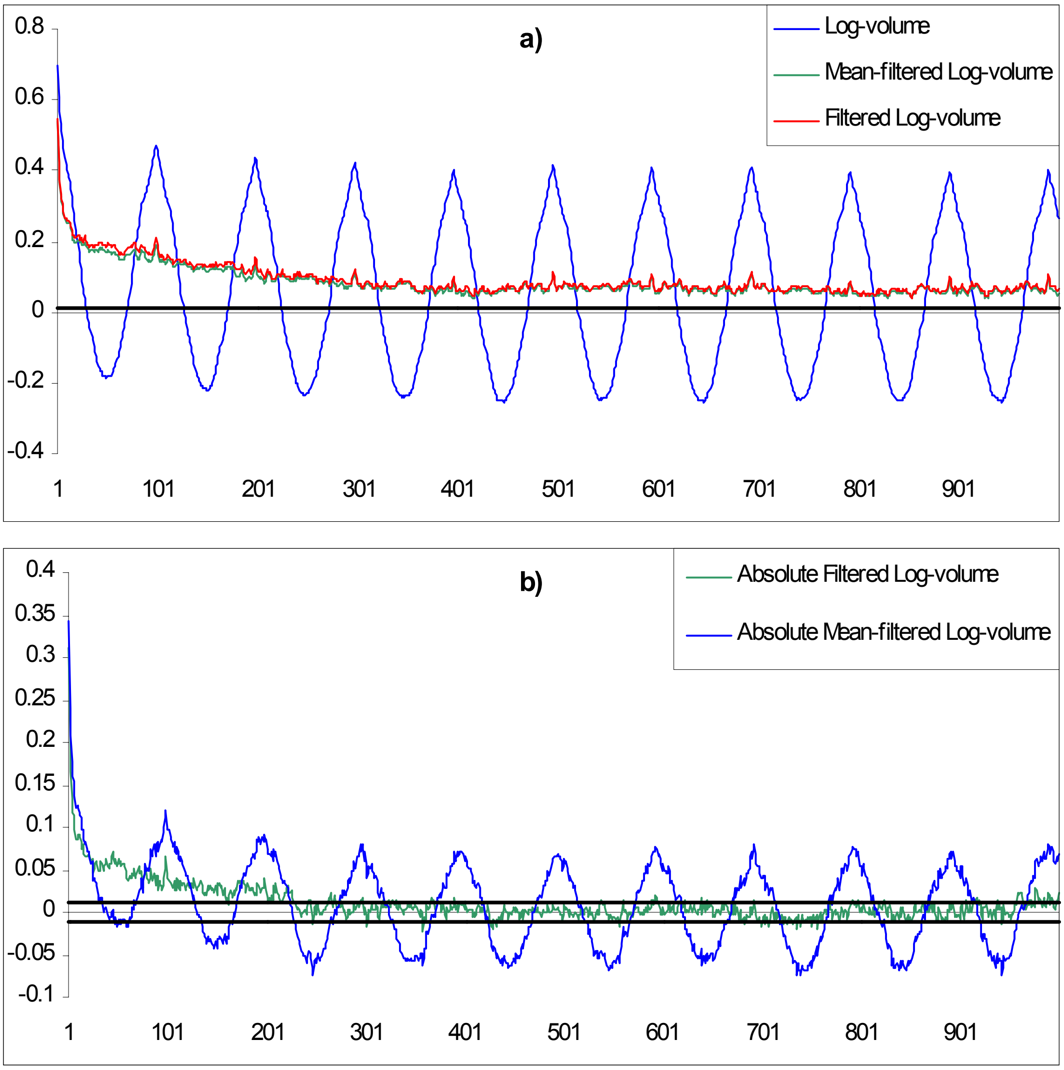


Figure 10: Autocorrelations of Volume

Let me now consider the volume series. Before all analysis the volume is rescaled with a log transformation, this operation will be necessary in a bivariate modelling view, to avoid any scale problem in parameter estimation, moreover this will change the distribution of volumes from a lognormal (volume are positive by construction) to a normal setting. As for the returns we start considering autocorrelations depicted in Figure 10. Clearly the autocorrelations computed on volume and on its absolute value are identical.

The seasonal pattern is found here in the mean, as Figure 10 (panel a) shows, but also in the variance: this second pattern emerges once the cyclical component in the mean has been removed, as we can see from Figure 10 (panel b). Therefore, for the volume the flexible Fourier form is used both on the mean and on the variance, the representation used is the following

$$\begin{aligned} v_{t,n} &= E\left[v_{t,n}\right] + \frac{\sigma_{t(,n)} s_{t,n} Z_{t,n}}{\sqrt{N}} \\ E\left[v_{t,n}\right] &= \bar{s}_{t,n} + E\left[\tilde{v}_{t,n}\right] \end{aligned} \tag{4}$$

with two distinct seasonal components. The Harmonics used both in the mean and in the variance are the same used for the returns (see table 3) as well as the dummies. I tried to estimate a complete model in one step, but the elevate number of parameters gave problems in the convergence of the optimization algorithm, therefore I adopted a two step procedure: in the first stage I have estimated and remove the cyclical component in the mean, while in the second step I dealt with the seasonal pattern in the variance. The opposite filtration scheme, at first the seasonal pattern in the variance and then the one in the mean in not efficient. I considered this alternative approach and applied the FFF-technique on the variances, it turned out that the cyclical component were not completely filtered and moreover the one on the mean resulted much more noisy. I presume that this is due both to the structure of the FFF technique and to the volume series values, all positive. Figures 11 and 12 show the seasonal patterns as they are found in the data and estimated by the FFF regression on data that includes the range 9:15-9:30. This inconsistency with the figures of the returns is introduces to verify that the cyclical patterns are relatively influenced by the number of daily intervals.

The periodic pattern in the volume mean is really similar to the one found in the variance of the returns: we can easily identify the peak in the first afternoon and the abrupt increase in volumes in coincidence with New York opening. Interestingly, the patter of volume volatility is of a very different shape, it peaks at noon and shows the widest variations in correspondence to New York opening and after 17 PM. This may not be so surprising: market activity is high early in the morning and near the end of the day, with an elevate number of contracts traded with high variation in prices, conversely at noon market activity is very low, prices do not move since trader wait for signals from the American market, therefore the number of contracts traded (the volume) per 5 minute intervals vary much more than in the morning or in the late afternoon. This particular behavior can explain the deterministic patterns of returns and volume.

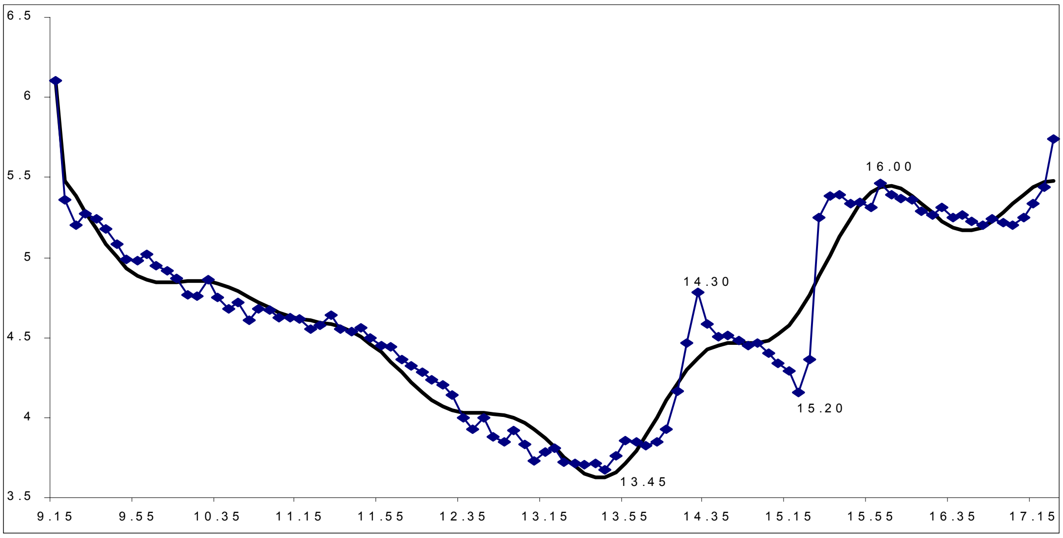


Figure 11: cyclical patter in the mean of the volume 9:20-17:30

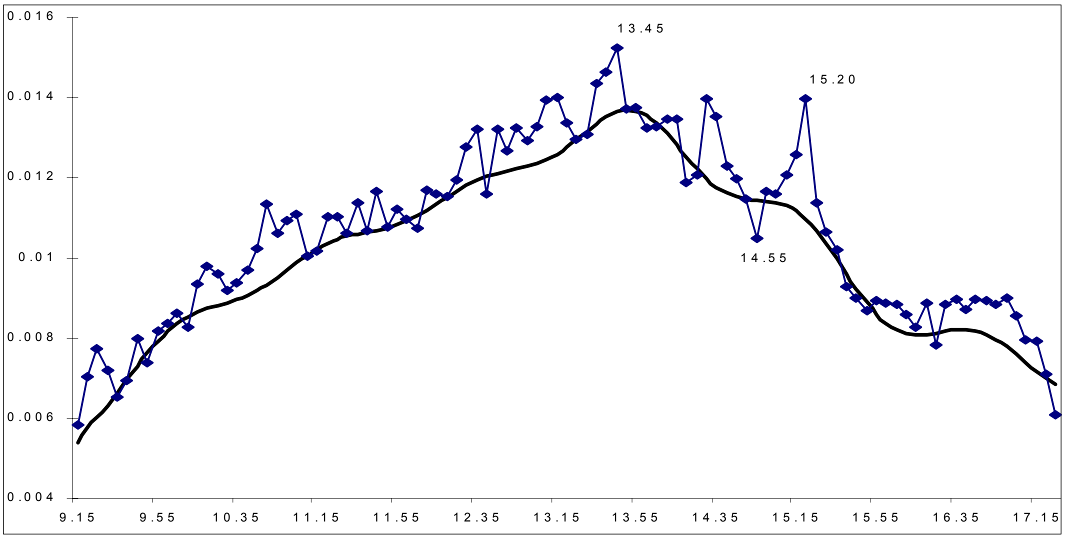


Figure 12: cyclical pattern in the variance of the volume 9:20-17:30

The estimated parameters of the two FFF regressions are reported in Table 8. As for the returns most of the harmonics used in the regression resulted to be significant, both in the mean and in the variance. The dummy for beginning of the day is significant in both regressions while the impulse dummies for the maturity dates are significant only for the variance, as if they affect only volatility and not the level of the traded contracts. In this table I report the estimation for the range 9:30-17:30 to compare the results with the one of table 6.

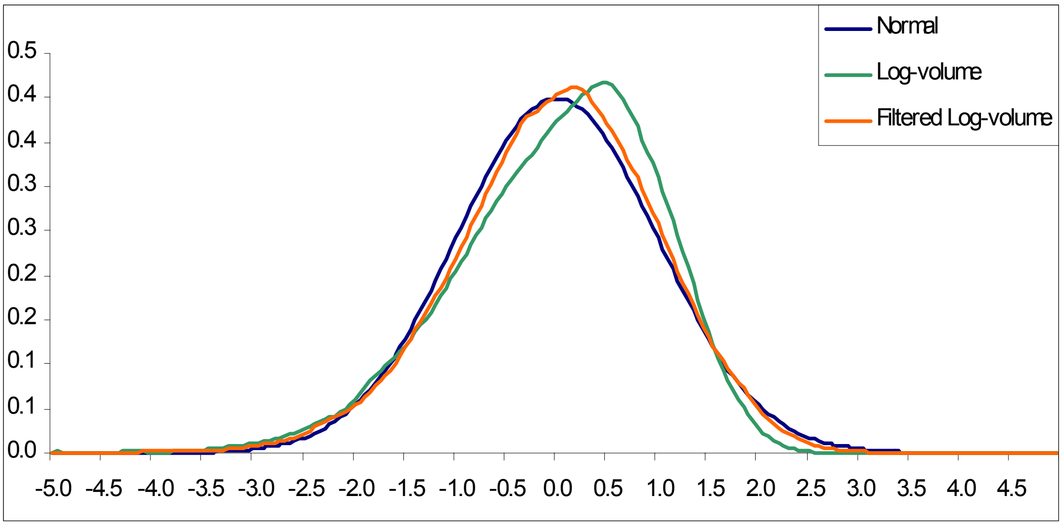


Figure 13: kernel density estimate of the log-volume

Similarly to the Returns, filtered volume is obtained by:

$$v_{t,n}^{f} = \frac{v_{t,n} - \bar{s}_{t,n} - E\left[\tilde{v}_{t,n}\right]}{s_{t,n}} \tag{5}$$

Consider now Table 9, where I report moments and tests for correlation and ARCH effects computed on the raw volume and after removing each of the two periodic component.

	Log-Volume - mean			Log-Volume - variance		
Paameter	Estimate	Std. Err.	t-value	Estimate	Std. Err.	t-value
α_1	7,807	0,281	27,738	-10,885	0,819	-13,290
α_2	-9,474	0,813	-11,647	14,721	2,367	6,219
α_3	3,193	0,271	11,786	-4,972	0,788	-6,306
γ_1	0,500	0,765	0,653	-53,645	2,226	-24,096
γ_2	-0,305	0,765	-0,399	-53,645	2,226	-24,096
γ_3	0,100	0,765	0,131	-53,645	2,226	-24,096
γ_4	-0,351	0,070	-5,007	0,561	0,204	2,751
δ_1	-1,172	0,159	-7,377	2,346	0,462	5,073
ϕ_1	-0,222	0,024	-9,404	0,071	0,069	1,032
δ_2	-0,585	0,040	-14,758	0,642	0,115	5,561
ϕ_2	-0,166	0,013	-12,541	0,116	0,038	3,024
δ_3	-0,148	0,018	-8,086	0,228	0,053	4,290
ϕ_3	-0,040	0,010	-3,926	-0,030	0,030	-1,007
δ_4	-0,136	0,012	-11,776	0,144	0,034	4,272
ϕ_4	0,049	0,009	5,557	-0,095	0,026	-3,705
δ_6	0,041	0,008	5,140	-0,032	0,023	-1,387
ϕ_6	-0,090	0,008	-11,522	0,007	0,023	0,292
δ_8	-0,004	0,007	-0,477	0,025	0,021	1,188
ϕ_8	0,055	0,007	7,405	-0,010	0,022	-0,462
δ_{12}	-0,016	0,007	-2,243	-0,004	0,021	-0,196
ϕ_{12}	0,012	0,007	1,711	-0,033	0,021	-1,582
δ_{16}	0,010	0,007	1,402	-0,029	0,021	-1,391
ϕ_{16}	0,045	0,007	6,374	0,003	0,021	0,160

Table 8: FFF regressions on log-volume, 9:30-17:30

Observing the table and the autocorrelations reported in Figure 10 we can note that a long memory effect is present both on the mean and on the variance of the volume. This effect is evident in the Portmanteau tests and in the Engle ARCH test. Interestingly, Skewness and Kurtosis evidence a distribution very close to the normal a fact that is much more evident in Figure 13. This particular effect of the filtration process becomes here much more evident than in the return case. Moreover, the estimated seasonal component in the variance is removed by a division and, since the values of the pattern are small compared to log-volume, the scale of the process is affected (note the increase in the standard deviation). In the following analysis the resulting filtered volume is rescaled one more time, dividing it by 100.

The exclusion of the contracts traded in the range 9:15-9:30 has here lower effects, the tests and the moment and almost equivalent.

	9:30-17:30			9:15-17:30		
Statistic	Raw	Mean filt.	Var. Filt.	Raw	Mean Filt.	Var. Filt.
Mean	4,618	-1,039e-13	0,018	4,647	-4,553e-14	-0,052
St. Dev.	0,936	0,763	71,988	0,942	0,758	73,603
Skewness	1,053	-0,285	-0,301	1,053	-0,289	-0,310
Kurtosis	1,137	3,853	3,511	1,137	3,866	3,439
Q(5)	39820,279	19185,399	19610,140	42157,828	19734,710	20377,917
Q(10)	62150,820	27802,570	28530,604	66140,347	28685,937	29768,113
Q(20)	83851,440	39924,860	41368,364	90418,897	40847,234	42945,089
Q(50)	93767,125	62361,783	66613,632	102471,400	65127,335	70888,597
Q(100)	174183,170	92785,812	103294,340	180922,470	97031,012	110257,680
$Q^2(5)$	40061,678	5418,088	5562,467	42274,666	5894,600	6000,630
Q^2 (10)	62074,589	7039,441	6849,840	65807,611	7706,708	7551,529
Q^2 (20)	83468,164	8718,860	8546,765	89803,130	9596,364	9627,444
Q^2 (50)	93201,321	9195,682	11105,105	101891,920	10188,584	12831,137
Q^2 (100)	174763,220	12307,430	14267,603	181726,110	13240,205	16805,901
JB	12188,761	1362,619	973,746	12569,571	1443,666	979,909
LM(2)	12020,309	7437,703	7582,311	12616,673	7694,255	7889,679
LM(5)	12492,956	7892,492	8042,759	13107,417	8157,775	8362,377
LM(10)	12572,968	8070,447	8227,787	13189,415	8345,803	8552,206
LM(20)	12622,467	8186,349	8351,431	13234,411	8461,731	8677,751

Table 9: Moments and tests of log-volume series

4 Univariate analysis

Consider now the two different series of filtered returns ($r^f_{t,n}$, in the following R_t) and volume ($v^f_{t,n}$, in the following V_t), as we can see from the ACF in figure 4 and 5, they show a different behavior. The returns clearly have a limited ARMA structure in the mean together with a long memory behavior in the variance, while volumes show long memory both in mean and in variance. In this section I try to fit a univariate model on these two series. The scope is to provide an initial base on which I can build up a multivariate model, as well as to have a set of starting value for its parameters. The following table reports the best parameterisations obtained for the two filtered series. I considered model combinations up to the order 2 for the ARMA oders p, q and for the GARCH order m, while only up to 1 for the ARCH order l. An increase in any the orders resulted in non significative parameter estimates. In the first column of the table I report the parameters used in the different specifications - inside the tables is reported the estimate and below it the correspondent standard error - and then, in the order: the log-likelihood; the informations criteria of Akaike (AIC), Hannan-Quinn (HQ), Schwarz (BIC) and Shibata (SH); the Box-Pierce - Portmanteau - test for residual $Q(k)$ and squared residual correlation $Q^2(k)$, computed up to k lags; the sample Skewness and Kurtosis together with the Jarque-Bera normality test; the Engle Lagrange Multiplier test $LM(k)$ for residual ARCH effects computed up to lag k. The preferred model is the one that minimises the maximum number of information criteria.

	Returns	Volume
	9:15-17:30	9:15-17:30
μ	0.00015	0.11343
	0.21818	0.05424
d_l		0.32058
		0.00723
ϕ	0.40289	
	0.04587	
θ	0.43134	-0.06995
	0.04682	0.00940
ω	0.30999	0.14505
	0.01327	0.02151
d	0.00074	0.06521
	2.18177	0.02537
β_1	0.53403	0.51301
	0.01368	0.15082
ψ	0.31495	0.42748
	0.01122	0.13835
LL	20097.919	-21522.460
AIC	-1.64329	1.76088
HQ	-1.64254	1.76164
BIC	-1.64371	1.76047
SH	-1.64329	1.76088
Q(5)	12.62629	3.95188
Q(10)	32.23680	14.72679
Q(20)	37.20633	42.03025
Q(50)	58.93983	85.27074
Q(100)	119.14041	269.35450
$Q^2(5)$	9.03280	3.57777
$Q^2(10)$	22.57878	14.05195
$Q^2(20)$	29.76496	22.06650
$Q^2(50)$	62.87172	58.11797
$Q^2(100)$	94.36597	128.33691
Sk	-0.12965	0.00261
K	5.51105	3.46881
JB	6561.12520	223.98374
LM(2)	3.11289	0.78289
LM(5)	12.50059	3.96542
LM(10)	31.35756	14.81023
LM(20)	36.77267	42.52746

Table 10 - Univariate estimation of filtered returns and volume series.

Analyzing the result of the univariate estimates and the corresponding information criteria and tests we choose the ARFIMA(1,d,1)-FIGARCH(1,d,1) for the returns and the ARFIMA(0,d,1)-FIGARCH(1,d,1) for the volume. These specifications were chosen both including and including the contracts concluded between 9:15 and 9:30. Our choice, given the results of the first chapter, was done only on the basis of the information criteria. The whole set of estimation is not reported for saving space but it can be requested directly to the author. Moreover, it is included in Caporin (2003b).

5 Multivariate analysis and causality

In this section I apply the multivariate techniques to infer the causal relations between returns, volume and their variances. A first analysis considered the approach of Cheung and Ng (1996): they suggest to fit univariate models that include lagged variables in the mean and lagged squared residuals in the variance equation as suggested by an analysis of cross correlations on the variables and on the squared variables. In this case there is evidence the cross-correlations evidence the existence of mean and variance causality, however the inclusion of the additional terms suggested by Cheung and Ng provides evidence of causality only from the returns to the volume. This analysis is not reported here and can be found in Caporin (2003b).

The previous estimations was a first attempt to causality testing, now I turn to a pure multivariate setting. At this point I estimate a VARFIMA(1,d,0) with residuals following a multivariate constant conditional correlation FIGARCH, that will be used as a benchmark model. The representation I use is the following, where r_t is the log-return series and v_t is the volume series; both variables are filtered from deterministic components:

$$
\begin{aligned}
\begin{bmatrix} R_t \\ \tilde{V}_t \end{bmatrix} &= \begin{bmatrix} \mu_R \\ \mu_V \end{bmatrix} + \begin{bmatrix} \phi_{1,1} & \phi_{1,2} \\ \phi_{2,1} & \phi_{2,2} \end{bmatrix} \begin{bmatrix} R_{t-1} \\ \tilde{V}_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{R,t} \\ \varepsilon_{V,t} \end{bmatrix} \\
\tilde{V}_t &= (1-L)^{d_{1,2}} V_t \\
\begin{pmatrix} \varepsilon_{R,t} \\ \varepsilon_{V,t} \end{pmatrix} &\sim F\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{R,t} & \sigma_{RV,t} \\ \sigma_{RV,t} & \sigma^2_{V,t} \end{bmatrix} \right) \\
\sigma^2_{R,t} &= \omega_1 + \beta_1 \sigma^2_{R,t-1} + \left[1 - \beta_1 L - (1-\psi_1 L)(1-L)^{d_{2,1}} \right] \varepsilon^2_{R,t} \\
\sigma^2_{V,t} &= \omega_2 + \beta_2 \sigma^2_{V,t-1} + \left[1 - \beta_2 L - (1-\psi_2 L)(1-L)^{d_{2,2}} \right] \varepsilon^2_{V,t} \\
\sigma_{RV,t} &= \rho \sigma_{R,t} \sigma_{V,t}
\end{aligned}
\tag{6}
$$

The model is estimated including data recorded from 9:15 and 9:30. Results are reportedin Table 11. As we can observe, the result contradicts the theory: there is evidence of the return causing the volume in the mean but not the opposite. However, we can explain this behavior reasoning on market structure. In the FIB30 market, the market makers quote prices for different maturities and we assume that they are as well as dealers, informed traders. Therefore as soon as a news impact on the market they react immediately changing prices. Differently, brokers and other traders, less informed that market makers, react observing price changes quoted by the firsts and adjust consequently then their positions. Using this temporal sequence, we can expect that changes in prices, responsible of changes in returns, will have a significant impact on the market volume.

Parameters i=1,2	Whole sample		02/01/02-15/03/02	
	Returns	Volume	Returns	Volume
μ_i	-0.00006 0.00060	0.10390 0.04851**	-0.00098 0.00317	0.25761 0.08650*
$d_{1,i}$		0.31751 0.00808*		0.31944 0.01752*
$\phi_{i,1}$	-0.02714 0.00699*	-0.11804 0.03113*	-0.01533 0.02742	-0.12402 0.05953*
$\phi_{i,2}$	0.00073 0.00101	0.07261 0.01079*	-0.00110 0.00409	0.10186 0.02426*
ω_i	0.00075 0.00008*	0.03912 0.00646*	0.00012 0.00148	0.14921 0.02351*
$d_{2,i}$	0.31319 0.01433*	0.16887 0.01549*	0.89112 0.24031*	0.08892 0.02126*
β_i	0.53139 0.02948*	0.68221 0.04202*	0.89312 0.09565*	0.01883 0.02839
ψ_i	0.30858 0.02589	0.58549 0.04396*	0.09571 0.15126	
ρ	-0.03625 0.00635*		-0.08009 0.03043	
AIC	-0.11433		-0.02897	
HQ	-0.11261		-0.02211	
BIC	-0.11541		-0.03388	
SH	-0.11433		-0.02898	
Q(5)	8.932	4.213	8.524	1.629
Q(10)	28.303	15.330	22.868	8.251
Q(20)	33.365	42.602	29.862	23.235
Q(50)	55.497	86.060	51.983	48.294
Q(100)	115.731	270.169	100.756	130.971
$Q^2(5)$	9.411	6.543	4.698	1.484
$Q^2(10)$	21.794	15.599	11.810	5.625
$Q^2(20)$	28.645	23.804	20.890	24.365
$Q^2(50)$	61.259	58.715	46.827	61.304
$Q^2(100)$	93.059	128.313	105.199	116.434

Table 11: bivariate model for returns and volume

The previous estimatio took into consideration only causality in mean, I have not yet considered second order causality within a multivariate framework. Given the complexity of the involved models and the CPU time needed to perform the estimation, from now on the results are based on a restricted sample. I consider the whole records of data (from 9:15 to 17:30) for a limited period of time, ranging from 2 January 2001 up to 15 March 2001, a total of 53 days, 5247 observations. To get a benchmark model for test on residual I re-estimated the baseline model, VARFIMA(1,d,0) with a CCC-FIGARCH, this is again reported in table 11. In this case the Volume GARCH structure has been modified into a FIGARCH(1,d,0), since the FIGARCH(1,d,1) performed very poorly. Although the β in the estimated FIGARCH(1,d,0) for the volume resulted non significant, I decided to maintain it in the model, this to preserve a limited short memory pattern within the volume series.

For the estimation of the model, I applied numerical optimizations procedures that allow for nonlinear constraints on parameters. I used the BFGS optimization algorithm for the first up to the third iteration, then the procedure switch to the Newton method. A full set of estimation for the multivariate benchmark model can be found in Caporin (2003b).

Given the benchmark estimates, the following step is the analysis of multivariate GARCH models that consider also the causal relation among variances. A first approact simply adds in the return (or volume) GARCH equation the lagged variance or squared residuals of the volume (return). The second model I consider in a new type of GARCH parameterisation that includes a causal relation, the Exponential Causality-GARCH model. Given that the model structure changes, I tried a wide range of possible specifications for the GARCH terms, starting form the GARCH(1,1) up to the FIGARCH(1,d,1). All the results are reported in Caporin (2003b), while here we focus on the most promising ones.

Specification	No causality		Exogenous Residuals		Exogenous Variance	
	Return	Volume	Return	Volume	Return	Volume
	Figarch(1,d,1)	Figarch(1,d,0)	Figarch(1,d,1)	Figarch(0,d,0)	Figarch(1,d,1)	Figarch(0,d,0)
Parameters	Estimate	Sd. Err.	Estimate	Sd. Err.	Estimate	Sd. Err.
VARFI(1,1)	-0.00098	0.00317	-0.00108	0.00138	-0.00105	0.00137
	0.25761	0.08650	0.25052	0.08058	0.25244	0.08024
	0.31944	0.01752	0.31993	0.01804	0.31974	0.01768
	-0.01533	0.02742	-0.01538	0.01869	-0.01524	0.02087
	-0.00110	0.00409	-0.00128	0.00281	-0.00124	0.00253
	-0.12402	0.05953	-0.12225	0.06054	-0.12202	0.05830
	0.10186	0.02426	0.10055	0.02383	0.10171	0.02364
Return	0.00012	0.00148	0.00002	0.00007	0.00000	0.00017
	0.89112	2.71137	0.96960	0.17348	0.95616	0.28148
	0.89312	1.31620	0.92879	0.06539	0.92281	0.11376
	0.09571	1.46300	0.04987	0.11706	0.05800	0.17707
Volume	0.14921	0.02351	0.16124	0.01650	0.16045	0.01643
	0.08892	0.02126	0.07877	0.01401	0.07946	0.01407
	0.01883	0.02839				
Causality			0.00016	0.00016	0.00027	0.00050
			0.00000	0.18917	0.00000	0.01894
Correlation	-0.08008	0.03044	-0.08077	0.02003	-0.08073	0.02031
Log-likelihood	87.414		88.618		87.940	
AIC	-0.02897		-0.02905		-0.02878	
HQ	-0.02211		-0.02174		-0.02147	
BIC	-0.03388		-0.03434		-0.03407	
SH	-0.02898		-0.02907		-0.02880	
	Return	Volume	Return	Volume	Return	Volume
Q(5)	8.524	1.629	8.678	1.647	8.729	1.676
Q(10)	22.868	8.251	23.356	8.254	23.349	8.286
Q(20)	29.862	23.235	30.680	23.184	30.552	23.184
Q(50)	51.983	48.294	52.214	48.128	52.391	48.148
Q(100)	100.756	130.971	100.378	131.217	100.748	131.117
$Q^2(5)$	4.698	1.484	4.640	1.724	4.557	1.728
$Q^2(10)$	11.810	5.625	11.079	5.860	11.239	5.869
$Q^2(20)$	20.890	24.365	19.142	24.705	19.562	24.724
$Q^2(50)$	46.827	61.304	44.689	62.826	45.284	62.835
$Q^2(100)$	105.199	116.434	102.225	118.437	102.944	118.432

Table 12.a: bivariate model estimates with causality components. Coefficient are in the order: VARFI μ_R, μ_V, d_V, $\phi_{1,1}$, $\phi_{1,2}$, $\phi_{2,1}$, $\phi_{2,2}$; GARCH part ω, d, β, ϕ; causality γ_R, γ_V.

Specification	No causality		Exponential Causality	
	Return	Volume	Return	Volume
	Figarch(1,d,1)	Figarch(1,d,0)	Figarch(1,d,1)	Garch(1,1)
Parameters	Estimate	Sd. Err.	Estimate	Sd. Err.
VARFI(1,1)	-0.00098	0.00317	-0.00122	0.00134
	0.25761	0.08650	0.33331	0.10461
	0.31944	0.01752	0.33207	0.01845
	-0.01533	0.02742	-0.02107	0.01599
	-0.00110	0.00409	-0.00156	0.00249
	-0.12402	0.05953	-0.13724	0.05985
	0.10186	0.02426	0.10747	0.02427
Return	0.00012	0.00148	0.00017	0.00006
	0.89112	2.71137	0.07249	0.01404
	0.89312	1.31620	0.92341	0.01481
	0.09571	1.46300		
Volume	0.14921	0.02351	0.01391	0.01044
	0.08892	0.02126	0.03002	0.01749
	0.01883	0.02839	0.95826	0.04313
Causality			-0.00611	0.00706
			-0.03655	0.01096
Correlation	-0.08008	0.03044	-0.08374	0.01944
Log-likelihood	87.414		105.821	
AIC	-0.02897		-0.03593	
HQ	-0.02211		-0.02862	
BIC	-0.03388		-0.04122	
SH	-0.02898		-0.03595	
	Return	Volume	Return	Volume
Q(5)	8.524	1.629	10.502	2.192
Q(10)	22.868	8.251	26.422	8.101
Q(20)	29.862	23.235	32.867	24.111
Q(50)	51.983	48.294	55.710	47.222
Q(100)	100.756	130.971	100.711	128.061
Q²(5)	4.698	1.484	6.743	10.741
Q²(10)	11.810	5.625	12.330	16.797
Q²(20)	20.890	24.365	18.877	33.713
Q²(50)	46.827	61.304	42.428	63.528
Q²(100)	105.199	116.434	97.129	119.462

Table 12.b: bivariate model estimate with exponential causality. Coefficient are in the order: VARFI μ_R, μ_V, d_V, $\phi_{1,1}$, $\phi_{1,2}$, $\phi_{2,1}$, $\phi_{2,2}$; GARCH part ω, d, β, ϕ; causality γ_R, γ_V.

Table 12 is divided in two panels, in both I report in the first two columns the benchmark model, while the remaining are devoted to the causality estimates: in panel a) with the introduction of an exogenous variable in the variance equation, while in panel b) with the introduction of a multiplicative exponential factor (EC-GARCH model). Recalling equation (6) the variance component has been modified into

$$\begin{aligned}
\sigma^2_{R,t} &= \omega_1 + \beta_1 \sigma^2_{R,t-1} + \gamma_1 f\left(V_{t-1}\right) + \\
&\quad \left[1 - \beta_1 L - \left(1 - \phi_1 L\right)\left(1 - L\right)^{d_{2,1}}\right] \varepsilon^2_{R,t} \\
\sigma^2_{V,t} &= \omega_2 + \beta_2 \sigma^2_{V,t-1} + \gamma_2 f\left(R_{t-1}\right) + \\
&\quad \left[1 - \beta_2 L - \left(1 - L\right)^{d_{2,2}}\right] \varepsilon^2_{V,t}
\end{aligned} \tag{7}$$

for panel a, where $f\left(V_{t-1}\right) = z^2_{V,t-1}$ and $f\left(R_{t-1}\right) = z^2_{R,t-1}$. For panel b the estimated variance equations are

$$\begin{aligned}
\sigma^2_{R,t} &= \exp\left(\gamma_1 f\left(V_{t-1}\right)\right)\left(\omega_1 + \alpha_1 \varepsilon^2_{R,t-1} + \beta_1 \sigma^2_{R,t-1}\right) \\
\sigma^2_{V,t} &= \exp\left(\gamma_2 f\left(R_{t-1}\right)\right)\left(\omega_2 + \alpha_2 \varepsilon^2_{V,t-1} + \beta_2 \sigma^2_{V,t-1}\right)
\end{aligned} \tag{8}$$

where $f\left(V_{t-1}\right)$ and $f\left(R_{t-1}\right)$ are substituted as before. In the table parameters are reported in the following order: for the VARFI(1,d) μ_R μ_V $d_{1,2}$ $\phi_{1,1}$ $\phi_{1,2}$ $\phi_{2,1}$ $\phi_{2,2}$; for the return ω_1 α_1 β_1; for the volume ω_2 α_2 β_2 both models have a GARCH(1,1) specification, as suggested in the theoretical presentation of the model. The remaining three parameters are, in the order, γ_1 γ_2 ρ. Finally note that the log-likelihoods and the information criteria (to be minimized) can be directly compared across the different specifications. First of all we can note that the causality parameters are non-significant in panel a, while in panel b the causality effect of return variance on volume variance became significant, as we were expecting. Moreover the likelihood clearly increase and all the information criteria prefer the second fitted model. Interestingly the tests on residuals are very close between the two models, this could be expected in the mean since both maintain the same structure, however some difference was expected in the variance because the first model include a long memory behavior while the second is a modified GARCH model, a short memory one. The two models return very close Box-Pierce statistics for elevate lags (50-100), while for the lags up to 20 it seems that the causality GARCH model cannot explain all the correlation. In principle the reverse should appear, the GARCH do not explain high lag correlation since it is a short memory one, however this result has been obtained. We conjecture that, at least in this case, the long memory behavior of the volume and return volatility is generated by the causality structure and an increase of the lags of the GARCH structure should be considered in order to explain the residual correlation for the lags 2-5. Moreover an interesting comparison could be the one of these results with a causality FIGARCH model. These estimations will be the object of future researches.

6 Conclusions

Within this work I focused on the analysis of the seasonal patterns and on the memory and causality structure of the FIB30 market. After the deletion of outliers, particular attention was givn to the identification of the cyclical deterministic components and its relation with the inclusion of the contracts concluded in the first 15 minutes of open market. For these analysis I considered the approach followed by Andersen and Bollerslev (1997) which suggest the use of Fourier regressions. The filtration process evidenced the presence of the wll-known U shape in the variance of the returns, high volatility at the opening and at the close of the market. Moreover, two effects due tothe lunch breack and to the opening of the American markets are noted. The same U pattern is found in the volume mean, while volume variance shows a reverse U pattern which is explained by the increase in the volaility of volumes when the trading decreases during the lunch break. The successive analysis consideres the memory and the causality within univariate and bivariate settings. There is evidence of long memory in the returns variance and in the volume mean and variance. Moreover, there exist a causal relation from the mean returns to the mean volume which contradict the theory. However, this behavious can be explained by the existence in the market of two groups of traders differently informed. Finally, there is also evidence of variance causality again from the returns on the volume. For this last analysis we considered a new type of GARCH model, the EC-GARCH which directly parameterise the variance causality. The EC-GARCH parameterisation was clearly preferred to alternative specifications by standard information criteria. Moreover, the short-memory EC-GARCH models adequately explain the evident variance long memory patterns, a fact which will be object of future studies.

References

[1] Andersed T. and Bollerslev T., 1997,

[2] Baillie, R.T and T. Bollerslev, H.O. Mikkelsen, 1996, *Fractionally integrated generalized autoregressive conditional heteroskedasticity*, Journal of Econometrics 74, 3-30

[3] Blume L., D. Easley and M. O'Hara, 1994, Market statistics and technical analysis: the role of the volume, *Journal of Finance*

[4] Bollerslev T., 1990, Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH approach, *Review of Economic Studies*, 72, 498-505

[5] Caporin M., 2003a, Long memory conditional heteroskedasticity and second order causality, Umpublished PhD dissertation, University of Venice

[6] Caporin M., 2003b, Second order causality with an application to the FIB30 returns and volume, GRETA Working Paper 01.03, www.greta.it

[7] Caporin M., 2003c, Variance (Non-)Causality: a bivariate GARCH-type model, GRETA Working Paper 07.03, www.greta.it

[8] Cheung Y. and L.K. Ng, 1996, A causality-in-variance test and its application to financial market prices, *Journal of Econometrics*, 72, 33-48

[9] Ding, Z., C.W.J. Granger and R.F. Engle 1993, A long memory property of stock market returns and a new model, *Journal of Finance*, 1, 83-106

[10] Engle R.F. and K.F. Kroner, 1995, Multivariate simultaneous generalized ARCH, *Econometric Theory*, 11, 122-150

[11] Glosten, L., R. Jagannathan and D. Runkle, 1993, On the relation between expected value and the volatility of the nominal excess returns on stocks, *Journal of Finance*, 48, 1779-1801

[12] Granger C.W.J., 1980, Testing for causality: a personal viewpoint, *Journal of Economics Dynamic and Control*, 2 (4), 329-352

[13] Karpoff J, 1987, The relation between price changes and trading volume: a survey, Journal of Financial and Quantitative Analysis, 22, 109-126

[14] LeBaron B., 1992, Persistence of the Dow Jones index on rising volumes, Working paper, Wisconsin Madison - Social System

7 Appendix

7.1 Filtering the cyclical component

We apply the Fast Fourier Form modeling, as in Andersen and Bollerslev (1997). This techniques is due to Gallant (1981, 1982) and is just a kind of regression on Fourier frequencies, trying to filter the periodic component via a mixture of harmonics. Assuming the following representation

$$x_{t,n} = E\left[x_{t,n}\right] + \frac{\sigma_{t(,n)} s_{t,n} Z_{t,n}}{\sqrt{N}} \tag{9}$$

that is the return at time t (day), intraday period n is equal to its expected value plus an heteroskedastic (daily or intradaily in that case denominator will be 1) error. Acting as in Bollerslev and Mikkelsen we estimate on unfiltered data the conditional variances with an MA(1)-GARCH(1,1), then we apply the following log-transformation

$$y_{t,n} = 2\ln\left[\left|x_{t,n} - E\left[x_{t,n}\right]\right|\right] - \ln\sigma^2_{t(,n)} + \ln N = \ln s^2_{t,n} + \ln Z^2_{t,n} \tag{10}$$

The resulting relation can be also viewed as the sum of a cyclical function plus a noise

$$y_{t,n} = f\left(\sigma_{t(,n)}, n, \theta\right) + \eta_{t,n} \tag{11}$$

Following the cited papers we specified the function as

$$f\left(\sigma_{t(,n)}, n, \theta\right) = \sum_{j=0}^{J} \sigma_{t(,n)}^{j} \left[\begin{array}{c} \alpha_{1,j} + 2\alpha_{2,j} n / (N+1) + 6\alpha_{3,j} n^2 / ((N+1)(N+2)) \\ + \sum_{i=i}^{I} \gamma_{i,j} I_{n=d_i} + \sum_{l=1}^{L} \left(\delta_{l,j} \cos \frac{nl2\pi}{N} + \phi_{l,j} \sin \frac{nl2\pi}{N}\right) \end{array} \right] \tag{12}$$

It is the sum of three different components: a quadratic term, a set of dummies and a group of harmonics. All are multiplied by a scaling factor that involve volatility. In this paper, given the elevate number of harmonic considered J is set in all cases identically equal to 0, this imply no scaling factor. The seasonal function is evaluated and parameters are estimated with OLS, given a preliminar estimate of daily conditional volatility. Given the estimated seasonal function it is necessary to reconstruct the cyclical component of equation (9), a point solved using

$$\hat{s}_{t,n} = \frac{T \exp\left(\hat{f}_{t,n}/2\right)}{\sum_{t=1}^{T/N} \sum_{n=1}^{N} \exp\left(\hat{f}_{t,n}/2\right)} \tag{13}$$

Given this component the variance filtering is performed with

$$\tilde{x}_{t,n} = x_{t,n} / \hat{s}_{t,n} \tag{14}$$

For the purposes of our analysis we are also interested in a filtration of a cyclical component in the mean and in the variance, the whole model is then

$$y_{t,n} = x_{t,n} - \bar{s}_{t,n} = E\left[x_{t,n} - \bar{s}_{t,n}\right] + \frac{\sigma_{t(,n)} s_{t,n} Z_{t,n}}{\sqrt{N}} \tag{15}$$

where $\bar{s}_{t,n}$ follow directly (12) with $J = 0$. Then the log-transformation correspondent to (10) become

$$y_{t,n} = 2 \ln\left[\left|y_{t,n} - E\left[y_{t,n}\right]\right|\right] - \ln \sigma_{t(,n)}^2 + \ln N = \ln s_{t,n}^2 + \ln Z_{t,n}^2 \tag{16}$$

therefore, given the estimated parameters filtration is performed using

$$\tilde{x}_{t,n} = \left(x_{t,n} - \bar{s}_{t,n}\right) / \hat{s}_{t,n} \tag{17}$$

chosen frequencies 1-2-3-4-6-8-12-16, equivalent respectively to 1 day (8 hrs.), 4 hrs, 2 hrs and 40', 2 hrs, 1 hr, 40' and 30'.

7.2 The Exponential Causality GARCH (EC-GARCH)

Traditional GARCH parameterisations share a common problem: strong zero parameter restrictions or assumptions on the relations among volatilities are required to avoid complex numerical evaluation, unrealistinc computing time and elevate number of parameters. Consequently, simple specifications are preferred, specially in the financial practice. One of the most used if the Constant Conditinanl Correlation GARCH of Bolleslev (1990). However, most of the currest specifications imply non-causality among variances, or provide parameters which are not interpretable (such as the BEKK of Engle and Kroner (1995)). Therefore, in Caporin (2003a and 2003c) I suggest an alternative parameterisation that directly extends the CCC-GARCH introducing in the variance equations a causality function dependant on the innovations of the other variables. The general bivariate model can be represented as

$$\begin{aligned} \begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} &= \begin{bmatrix} \mu_{1,t}\left(I^{t-1}\right) \\ \mu_{2,t}\left(I^{t-1}\right) \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \\ & \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \tilde{\ } iid \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{1,t} & \rho\sigma_{1,t}\sigma_{2,t} \\ \rho\sigma_{1,t}\sigma_{2,t} & \sigma^2_{2,t} \end{bmatrix} \right) \end{aligned} \tag{18}$$

where the mean dynamic is not specified and we can allow for time dependence based on the information set up to time $t-1$ $\left(I^{t-1}\right)$ including therefore also GARCH-in-mean effects. The variances are represented as

$$\sigma^2_{1,t} = \exp\left[f_1\left(I^{t-1}\right)\right]\left[\omega_1 + \beta_1\sigma^2_{1,t-1} + \alpha_1\varepsilon^2_{1,t-1}\right] \tag{19}$$

$$\sigma^2_{2,t} = \exp\left[f_2\left(I^{t-1}\right)\right]\left[\omega_2 + \beta_2\sigma^2_{2,t-1} + \alpha_2\varepsilon^2_{2,t-1}\right] \tag{20}$$

where I included a standard GARCH(1,1) model but there are no constraint to consider a general GARCH(p,q) as well as any alternative GARCH specification such as the FIGARCH of Baillie, Bollerslev and Mikkelsen (1994), the leverage GARCH of Glosten, Jagannathan and Runkle (1993) or the asymmetric power ARCH of Ding, Granger and Engle (1993). Moreover, the different variances can have different structures. The causal relation is driven by the functions $f_1\left(I^{t-1}\right)$ and $f_2\left(I^{t-1}\right)$ that depends on the information sets up to time $t-1$, and are represente ad

$$f_1\left(I^{t-1}\right) = \gamma_1 z^2_{2,t-1} \tag{21}$$

In this setup the causality effect, driven by the parameter γ_1, allows for positive and negative causality: i.e. in the sense that an increase in the variance of the second series imply an increase in the variance of the first series only if the function $f_i(\cdot)$ is greater than 1 (the parameter greater than zero), otherwise we will have a decrease. Non causality is associated with a zero parameter. Moreover parameters need not to be constrained given the exponential formulation. Therefore, a significativity test on the parameter γ_1 will indicate the existence or not of a causal relation between the variances of the two series, while its sign can be interpreted as the causality direction. Moreover, the model is stationary, as verified in Caporin (2003c) and it nests the CCC-GARCH which can be obtained with the zero restrictions on the causality parameters.