

A long memory property of stock market returns and a new model*

Zhuanxin Ding, Clive W.J. Granger and Robert F. Engle

University of California, San Diego, CA, USA

Received June 1992. Final version accepted February 1993

Abstract

A 'long memory' property of stock market returns is investigated in this paper. It is found that not only there is substantially more correlation between absolute returns than returns themselves, but the power transformation of the absolute return $|r_t|^d$ also has quite high autocorrelation for long lags. It is possible to characterize $|r_t|^d$ to be 'long memory' and this property is strongest when d is around 1. This result appears to argue against ARCH type specifications based upon squared returns. But our Monte-Carlo study shows that both ARCH type models based on squared returns and those based on absolute return can produce this property. A new general class of models is proposed which allows the power δ of the heteroskedasticity equation to be estimated from the data.

1. Introduction

If r_t is the return from a speculative asset such as a bond or stock, this paper considers the temporal properties of the functions $|r_t|^d$ for positive values of d . It is well known that the returns themselves contain little serial correlation, in agreement with the efficient market theory. However, Taylor (1986) found that $|r_t|$ has significant positive serial correlation over long lags. This property is examined on long daily stock market price series. It is possible to characterize $|r_t|^d$ to be 'long-memory', with quite high autocorrelations for long lags. It is also found, as an empirical fact, that this property is strongest for $d=1$ or near 1 compared to both smaller and larger positive values of d . This result appears to argue against ARCH type specifications based upon squared returns. The paper examines whether various classes of models are consistent with this observation. A new general class of models is then proposed which allows the power δ of the heteroskedasticity equation to be estimated from the data.

Correspondence to: Clive Granger, University of California, San Diego, CA 92093-0508, USA.

*We thank Jurg Barlocher, Xiaohong Chen, Takeo Hoshi, Bruce Lehman, Victor Ng, and Ross Starr for helpful comments and discussions. We are also grateful to the editor (Richard T. Baillie) and two anonymous referees for their constructive comments. The second and third authors would like to thank NSF for financial support.

The remainder of this paper is organized as follows: In section 2, we give a brief description of the data we use. In section 3 we carry out the autocorrelation and cross-correlation analysis. The special pattern of the autocorrelogram and crosscorrelogram of the stock returns is exploited and presented. Section 4 investigates the effect of temporal aggregation on the autocorrelation structure and examines the short sample autocorrelation property of stock returns. Section 5 presents a Monte-Carlo study of various financial models. Based on this, we propose a new general class of models in section 6. Section 7 concludes the analysis.

2. The data

The data set we will analyze in this paper is the Standard & Poor 500 (hereafter S&P 500) stock market daily closing price index.¹ There are altogether 17055 observations from Jan 3, 1928 to Aug 30, 1991. Denote p_t as the price index for S&P 500 at time t ($t=0, \dots, 17055$). Define

$$r_t = \ln p_t - \ln p_{t-1} \quad (1)$$

as the compounded return for S&P 500 price index at time t ($t=1, \dots, 17054$).

Table 2.1 gives the summary statistics for r_t . We can see from table 2.1 that the kurtosis for r_t of 25.42 is higher than that of a normal distribution which is 3. The kurtosis and studentized range statistics (which is the range divided by standard deviation) show the characteristic ‘fat-tailed’ behavior compared with a normal distribution. The Jarque–Bera normality test statistic is far beyond the critical value which suggests that r_t is far from a normal distribution.

Figs. 2.1, 2.2 and 2.3 give the plots of p_t , r_t and $|r_t|$. We can see from the figures the long run movement of daily p_t , r_t , $|r_t|$ over the past 62 years. There is an upward trend for p_t , but r_t is rather stable around mean $\mu=0.00018$. From the series $|r_t|$, we can clearly see the observation of Mandelbrot (1963) and Fama (1965) that large absolute returns are more likely than small absolute returns to

Table 2.1
Summary statistics of r_t .

data	sample size	mean	std	skewness	kurtosis	min	max	studentized range	normality test
r_t	17054	0.00018	0.0115	-0.487	25.42	-0.228	0.154	33	357788

¹We are indebted to William Schwert for providing us the data.

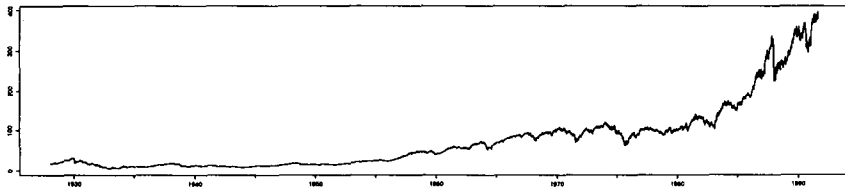


Fig. 2.1. Standard & Poor 500 daily price index 01/03/28–08/30/91.

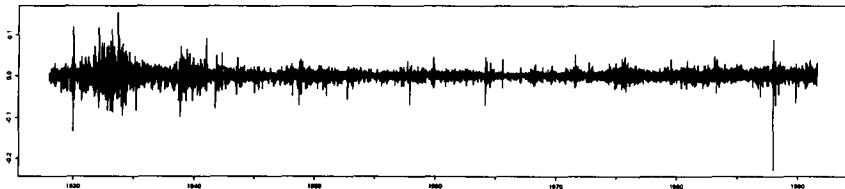


Fig. 2.2. Standard & Poor 500 daily returns 01/04/28–08/30/91.

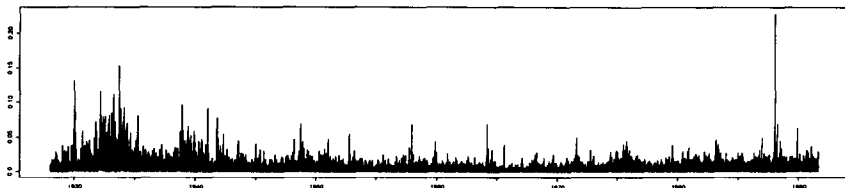


Fig. 2.3. Standard & Poor 500 daily absolute returns 01/04/28–08/30/91.

be followed by a large absolute return. The market volatility is changing over time which suggests a suitable model for the data should have a time varying volatility structure as suggested by the ARCH model. During the Great Depression of 1929 and early 1930s, volatilities are much higher than any other period. There is a sudden drop in prices on Black Monday's stock market crash of 1987, but unlike the Great Depression, the high market volatility did not last very long. Otherwise, the market is relatively stable.

3. Autocorrelation analysis of the return series

It is now well established that the stock market returns themselves contain little serial correlation [Fama (1970), Taylor (1986)] which is in agreement with the efficient market theory. But this empirical fact does not necessarily imply that

returns are independently identically distributed as many theoretical financial models assume. It is possible that the series is serially uncorrelated but is dependent. The stock market data is especially so since if the market is efficient, a stock's price should change with the arrival of information. If information comes in bunches, the distribution of the next return will depend on previous returns although they may not be correlated.

Taylor (1986) studied the correlations of the transformed returns for 40 series and concluded that the returns process is characterized by substantially more correlation between absolute or squared returns than there is between the returns themselves. Kariya et al. (1990) obtained a similar result when studying Japanese stock prices. Extending this line we will examine the autocorrelation of r_t and $|r_t|^d$ for positive d in this section, where r_t is the S&P 500 stock return.

Table 3.1 gives the sample autocorrelations of r_t , $|r_t|$ and r_t^2 for lags 1 to 5 and 10, 20, 40, 70, 100. We plot the autocorrelogram of r_t , $|r_t|$ and r_t^2 from lag 1 to lag 100 in fig. 3.1. The dotted lines show $\pm 1.96/\sqrt{T}$ which is the 95% confidence interval for the estimated sample autocorrelations if the process r_t is independently and identically distributed (hereafter i.i.d.). In our case $T=17054$ so $\pm 1.96/\sqrt{T}=0.015$. It is proved [Bartlett (1946)] that if r_t is a i.i.d process then the sample autocorrelation ρ_τ is approximately $N(0, 1/T)$. In fig. 3.1, about one quarter of the sample autocorrelations within lag 100 are outside the 95% confidence interval for a i.i.d process. The first lag autocorrelation is 0.063 which is significantly positive. Many other researchers [see Fama (1976), Taylor (1986), Hamao et al. (1990)] also found that most stock market

Table 3.1
Autocorrelations of r_t .

data	lag 1	2	3	4	5	10	20	40	70	100
r_t	0.063	-0.039	-0.004	0.031	0.022	0.018	0.017	0.000	0.000	0.004
$ r_t $	0.318	0.323	0.322	0.296	0.303	0.247	0.237	0.200	0.174	0.162
r_t^2	0.218	0.234	0.173	0.140	0.193	0.107	0.083	0.059	0.058	0.045

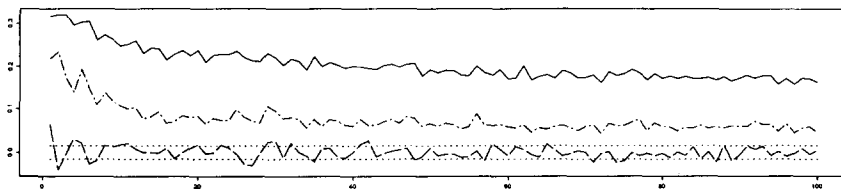


Fig. 3.1. Autocorrelation of $|r|$, r^{**2} , r from high to low.

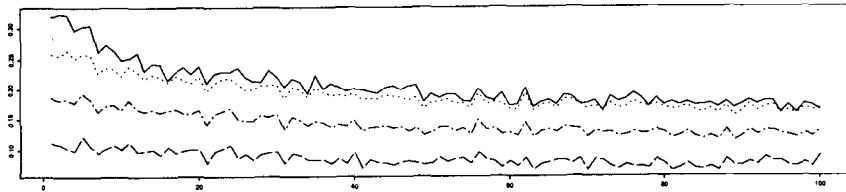
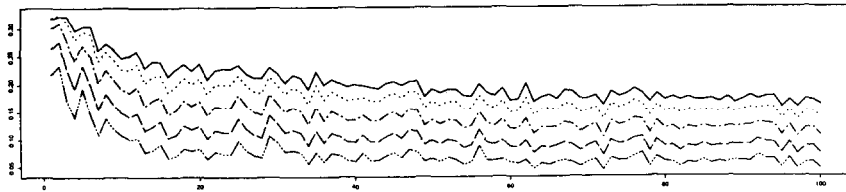
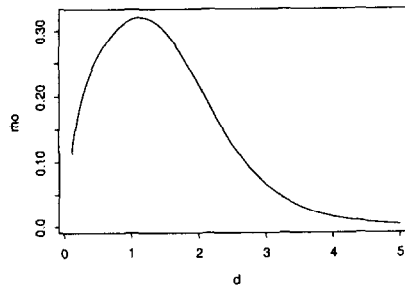
return series have a very small positive first order autocorrelation. The small positive first order autocorrelation suggests that the r_t do have some memory although it is very short and there is a portion of stock market returns that is predictable although it might be a very small one. So the efficient market or random walk hypothesis does not hold strictly. Alternatively, this could be from non-synchronous measurement of prices. The second lag autocorrelation ($= -0.039$) is significantly negative which supports the so called 'mean-reversion' behaviour of stock market returns. This suggests that the S&P 500 stock market return series is not a realization of an i.i.d process.

Furthermore, if r_t is an i.i.d process, then any transformation of r_t is also an i.i.d process, so will be $|r_t|$ and r_t^2 . The standard error of the sample autocorrelation of $|r_t|$ will be $1/\sqrt{T}=0.015$ if r_t has finite variance, the same standard error is applicable for the sample autocorrelation of r_t^2 providing the r_t also have finite kurtosis. But from fig. 3.1, it is seen that not only the sample autocorrelations of $|r_t|$ and r_t^2 are all outside the 95% confidence interval but also they are all positive over long lags. Further, the sample autocorrelations for absolute returns are greater than the sample autocorrelations for squared returns at every lag up to at least 100 lags. It is clear that the S&P 500 stock market return process is not an i.i.d process.

Based on the finding above, we further examined the sample autocorrelations of the transformed absolute S&P 500 returns $|r_t|^d$ for various positive d , Table 3.2 gives $\text{corr}(|r_t|^d, |r_{t+\tau}|^d)$ for $d=0.125, 0.25, 0.50, 0.75, 1, 1.25, 1.5, 1.75, 2, 3$ at lags 1 to 5 and 10, 20, 40, 70, 100. Figs. 3.2, 3.3 show the autocorrelogram of $|r_t|^d$ from lag 1 to 100 for $d=1, 0.50, 0.25, 0.125$ in fig. 3.2 and $d=1, 1.25, 1.5, 1.75, 2$ in fig. 3.3. From table 3.2 and figs. 3.2, 3.3 it is seen that the conclusion obtained above remains valid. All the power transformations of the absolute return have significant positive autocorrelations at least up to lag 100 which supports the claim that stock market returns have long-term memory. The autocorrelations decrease fast in the first month and then decrease very slowly. The most

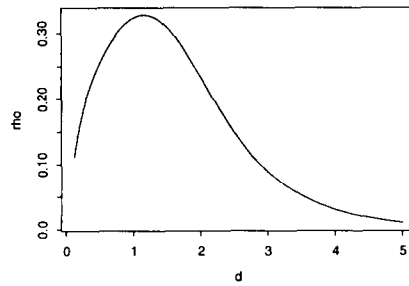
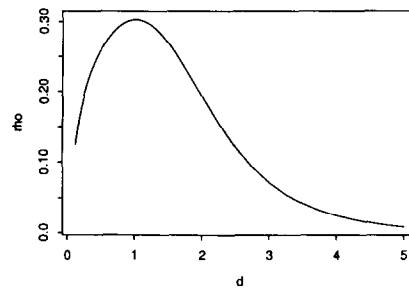
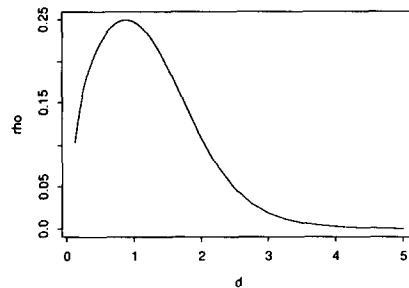
Table 3.2
Autocorrelations of $|r_t|^d$.

d	lag 1	2	3	4	5	10	20	40	70	100
0.125	0.110	0.108	0.102	0.098	0.121	0.100	0.100	0.095	0.065	0.089
0.25	0.186	0.181	0.182	0.176	0.193	0.164	0.164	0.148	0.120	0.131
0.5	0.257	0.255	0.263	0.251	0.259	0.222	0.221	0.192	0.166	0.165
0.75	0.297	0.299	0.305	0.286	0.291	0.246	0.241	0.207	0.180	0.173
1	0.318	0.323	0.322	0.296	0.303	0.247	0.237	0.200	0.174	0.162
1.25	0.319	0.326	0.312	0.280	0.295	0.227	0.211	0.174	0.153	0.138
1.5	0.300	0.309	0.278	0.242	0.270	0.192	0.170	0.136	0.122	0.106
1.75	0.264	0.276	0.228	0.192	0.234	0.149	0.125	0.095	0.088	0.073
2	0.218	0.234	0.173	0.140	0.193	0.107	0.083	0.059	0.058	0.045
3	0.066	0.088	0.036	0.025	0.072	0.019	0.009	0.004	0.006	0.003

Fig. 3.2. $d=1, 0.5, 0.25, 0.125$ from high to low.Fig. 3.3. $d=1, 1.25, 1.50, 1.75, 2$ from high to low.Fig. 3.4. Autocorrelation of $|r|^d$ at lag 1.

interesting finding from the autocorrelogram is that $|r_t|^d$ has the largest autocorrelation at least up to lag 100 when $d=1$ or is near 1. The autocorrelation gets smaller almost monotonically when d goes away from 1.

To illustrate this more clearly, we calculate the sample autocorrelations $\rho_\tau(d)$ as a function of $d, d>0$, for $\tau=1, 2, 5, 10$ and taking $d=0.125, 0.130, \dots, 1.745, 1.750, 2, 2.25, \dots, 4.75, 5$. Figs. 3.4, 3.5, 3.6 and 3.7 give the plots of calculated $\rho_\tau(d)$ at $\tau=1, 2, 5, 10$. It is seen clearly from these figures that the autocorrelation $\rho_\tau(d)$ is a smooth function of d . There is a saddle point \tilde{d} between 2 and 3 such that when $d<\tilde{d}$, $\rho_\tau(d)$ is a concave function and when $d>\tilde{d}$, $\rho_\tau(d)$ is a convex function of d . There is a unique point d^* around 1 such that $\rho_\tau(d)$ reaches its maximum at this point, $\rho_\tau(d^*)>\rho_\tau(d)$ for $d\neq d^*$.

Fig. 3.5. Autocorrelation of $|r|^{**d}$ at lag 2.Fig. 3.6. Autocorrelation of $|r|^{**d}$ at lag 5.Fig. 3.7. Autocorrelation of $|r|^{**d}$ at lag 10.

In fact, $|r_t|^d$ has positive autocorrelations over a much longer lags than 100. Table 3.3 shows the lags (τ^*) at which the first negative autocorrelation of $|r_t|^d$ occurs for various d . It can be seen from the table that in most cases, $|r_t|^d$ has positive autocorrelations over more than 2500 lags. Since there are about 250 working days every year, the empirical finding suggests that $|r_t|^d$ has positive autocorrelations for over 10 years!

Table 3.3
Lags at which first negative autocorrelations of $|r_t|^d$ occurs.

d	0.125	0.25	0.5	0.75	1	1.25	1.5	1.75	2	3
τ^*	2028	2534	2704	2705	2705	2705	2705	2685	2598	520

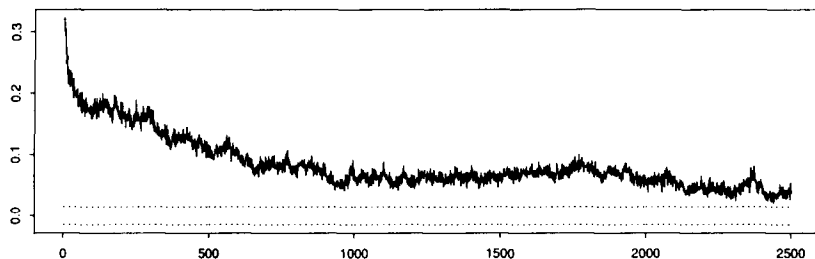


Fig. 3.8. Autocorrelation of $|r|$ up to lag 2500.

We pick $|r_t|$ as a typical transform of the return series here and plot its sample autocorrelations up to lag 2500 in fig. 3.8. The dotted lines are 95% confidence interval for the estimated sample autocorrelation of an i.i.d process as before. It is striking that all the sample autocorrelations are not only positive but also stay outside the confidence interval. Different models have been tried to approximate this sample autocorrelation curve, including: (1) ρ_τ an exponentially decreasing function of τ ($\rho_\tau = \alpha\beta^\tau$) (which is similar to the autocorrelation function of a ARMA model); (2) ρ_τ the same as the autocorrelation function of a fractionally integrated process [see Granger and Joyeux (1980)]

$$\begin{aligned}
 \rho_\tau &= \frac{\Gamma(1-\beta)}{\Gamma(\beta)} \frac{\Gamma(\tau+\beta)}{\Gamma(\tau+1-\beta)} \\
 &= \frac{\Gamma(1-\beta)}{\Gamma(\beta)} \frac{(\tau+\beta-1) \dots \beta \Gamma(\beta)}{(\tau-\beta) \dots (1-\beta) \Gamma(1-\beta)} \\
 &= \frac{(\tau+\beta-1) \dots \beta}{(\tau-\beta) \dots (1-\beta)} \\
 &= \rho_{\tau-1} \frac{(\tau+\beta-1)}{(\tau-\beta)};
 \end{aligned} \tag{2}$$

and (3) ρ_τ a polynomially decreasing function of τ ($\rho_\tau = \alpha/\tau^\beta$) which is approximately the same as (2) when τ is large. It is found, compared to the real data, that the fitted autocorrelation using method (1) decreases too slowly at the beginning

and then too fast at the end while by using methods (2) and (3) the opposite result is found.

The final preferred model is a combination of these methods. A theoretical autocorrelation function is specified as follows:

$$\rho_\tau = \frac{\alpha \rho_{\tau-1}^{\beta_1} \beta_2^\tau}{\tau^{\beta_3}} \quad (3)$$

which can easily be transformed to a linear model

$$\log \rho_\tau = \log \alpha + \beta_1 \log \rho_{\tau-1} + \tau \log \beta_2 - \beta_3 \log \tau. \quad (4)$$

Let $\alpha^* = \log \alpha$, $\beta_1^* = \beta_1$, $\beta_2^* = \log \beta_2$, and $\beta_3^* = -\beta_3$, then

$$\log \rho_\tau = \alpha^* + \beta_1^* \log \rho_{\tau-1} + \beta_2^* \tau + \beta_3^* \log \tau. \quad (5)$$

Ordinary Least Squares gives as estimates:

$$\log \rho_\tau = -0.049 + 0.784 \log \rho_{\tau-1} - 0.195 \times 10^{-4} \tau - 0.057 \log \tau, \quad (6)$$

$(-3.9) \quad (62.9) \quad (-5.9) \quad (-9.1)$

$$R^2 = 0.92, D-W = 2.65.$$

The t -statistics inside parentheses show that all the parameters are significant. After transferring the above equation back to autocorrelations one gets:

$$\rho_\tau = 0.893 \rho_{\tau-1}^{0.784} (0.999955)^\tau / \tau^{0.057}. \quad (7)$$

Fig. 3.9 plots the fitted autocorrelations (dotted line) and the sample autocorrelations themselves. It is seen that the theoretical model fits the actual sample autocorrelations quite well.

Similar studies were also carried out for the New York Stock Exchange daily price index and the German daily stock market price index (DAX) over a shorter sample period (1962–1989 for NYSE, 1980–1991 for DAX); we get similar

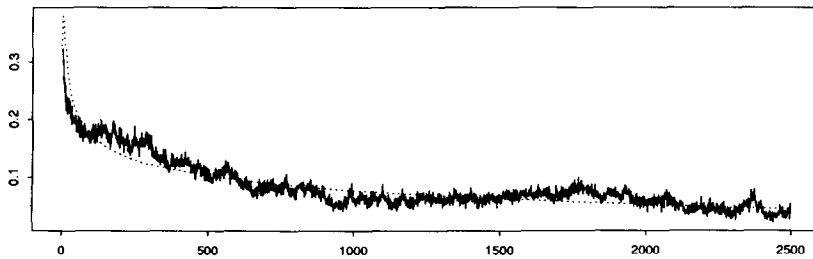


Fig. 3.9. Autocorrelation of $|r|$ (solid line) and its fitted value (dotted line).

autocorrelation structures for transformed returns. Furthermore, we did the cross-correlation analysis of transformed S&P 500 and New York Stock Exchange daily returns series and also found the cross-correlation is the biggest when $d=1$ and that it also has long memory. This suggests there may be volatility co-persistence for these two stock market index prices (see Bollerslev and Engle 1989). Our conjecture is that this property will exist in most financial series.

4. Sensitivity of autocorrelation structure

We now further investigate the effect of temporal aggregation on the autocorrelation structure. Table 3.4 gives the autocorrelations of $|\bar{r}_{t,5}|^d$, where $\bar{r}_{t,5}$ is the 5 day temporal average of r_t . i.e.

$$\bar{r}_{t,5} = \frac{1}{5}(r_{t'+1} + r_{t'+2} + \cdots + r_{t'+5}), \quad (8)$$

where $t=1, 2, \dots, 3410$ and $t'=5(t-1)$. It can be seen that the temporal aggregation does not change the long memory property of the absolute return series. $\rho_\tau(|\bar{r}_{t,5}|^d)$ still reaches a unique maximum when d is around 1 or 1.25 for different lags τ . Compared with the original daily series, the first order autocorrelation for $|\bar{r}_{t,5}|^d$ is much bigger than the second one. Although the temporally aggregated return series here is not exactly the same as a weekly return series, we expect a similar result will hold for the weekly data.

It should also be noted from fig. 2.2 that the volatility structure differs considerably between the pre-war and the post-war period. The pre-war period (1928–1945) is much more volatile than the post-war period (1946–1986). It will be interesting to look at the memory structure for these two periods. Table 3.5

Table 3.4
Autocorrelations of $|\bar{r}_{t,5}|^d$.

d	lag 1	2	3	4	5	10	20	40	70	100
0.125	0.145	0.109	0.148	0.149	0.136	0.105	0.129	0.077	0.072	0.041
0.25	0.187	0.155	0.184	0.184	0.169	0.137	0.158	0.102	0.095	0.052
0.5	0.247	0.213	0.229	0.227	0.204	0.180	0.188	0.136	0.126	0.065
0.75	0.296	0.255	0.261	0.255	0.223	0.212	0.203	0.161	0.149	0.074
1	0.332	0.279	0.279	0.267	0.227	0.233	0.205	0.175	0.163	0.079
1.25	0.352	0.286	0.282	0.263	0.217	0.243	0.197	0.178	0.168	0.080
1.5	0.356	0.277	0.271	0.245	0.196	0.242	0.180	0.173	0.164	0.076
1.75	0.349	0.255	0.250	0.217	0.169	0.231	0.160	0.160	0.153	0.069
2	0.332	0.227	0.223	0.186	0.140	0.214	0.138	0.144	0.138	0.061
3	0.237	0.109	0.115	0.075	0.048	0.124	0.068	0.079	0.073	0.026

Table 3.5
Autocorrelations of $|r_t|^d$ 1928–1945.

d	lag 1	2	3	4	5	10	20	40	70	100
0.125	0.114	0.135	0.126	0.117	0.138	0.131	0.122	0.118	0.067	0.115
0.25	0.201	0.227	0.231	0.204	0.215	0.200	0.197	0.183	0.128	0.158
0.5	0.273	0.298	0.311	0.275	0.276	0.245	0.245	0.216	0.169	0.172
0.75	0.300	0.323	0.332	0.296	0.294	0.251	0.248	0.212	0.172	0.162
1	0.310	0.329	0.329	0.296	0.293	0.241	0.232	0.192	0.159	0.141
1.25	0.310	0.323	0.310	0.280	0.281	0.223	0.205	0.163	0.138	0.116
1.5	0.302	0.310	0.283	0.256	0.260	0.199	0.173	0.130	0.114	0.090
1.75	0.289	0.292	0.251	0.226	0.236	0.175	0.141	0.099	0.091	0.067
2	0.273	0.272	0.218	0.196	0.211	0.151	0.111	0.072	0.070	0.047
3	0.201	0.194	0.114	0.098	0.128	0.076	0.034	0.012	0.020	0.007

Table 3.6
Autocorrelations of $|r_t|^d$ 1946–1986.

d	lag 1	2	3	4	5	10	20	40	70	100
0.125	0.089	0.062	0.054	0.057	0.086	0.047	0.054	0.051	0.041	0.038
0.25	0.129	0.095	0.086	0.102	0.126	0.082	0.082	0.068	0.058	0.053
0.5	0.162	0.128	0.121	0.141	0.158	0.111	0.106	0.082	0.068	0.066
0.75	0.181	0.151	0.143	0.164	0.175	0.126	0.119	0.088	0.067	0.068
1	0.191	0.167	0.157	0.180	0.182	0.133	0.123	0.089	0.062	0.064
1.25	0.194	0.178	0.163	0.191	0.180	0.134	0.120	0.084	0.053	0.056
1.5	0.189	0.182	0.160	0.198	0.170	0.129	0.110	0.074	0.042	0.046
1.75	0.178	0.179	0.150	0.200	0.154	0.119	0.095	0.061	0.031	0.036
2	0.163	0.170	0.135	0.199	0.133	0.105	0.078	0.047	0.021	0.027
3	0.099	0.104	0.066	0.173	0.056	0.047	0.023	0.010	0.002	0.005

shows the autocorrelations of $|r_t|^d$ for the pre-war period (1928–1945). It is seen that the magnitude of the autocorrelation for $|r_t|^d$ is about the same as those in table 3.2. $|r_t|$ has the largest autocorrelation for the first two lags and then this property becomes strongest for $|r_t|^{0.75}$ or $|r_t|^{0.5}$.

Table 3.6 gives the autocorrelations of $|r_t|^d$ for the post-war period (1946–1986). It is clear from the table that during this less volatile period the market has both a smaller and a shorter memory in the sense that the autocorrelations are smaller and decrease faster. The autocorrelations are only about two thirds as big as those of the pre-war period.

Comparing table 3.2, 3.5 and 3.6 we can probably say that the long memory property that was found in the whole sample period can be mainly attributed to the pre-war period. The market has a strong and long memory of big events like the great depression in 1929 and early 1930s when volatility was very high.

5. Monte-Carlo study of various financial time series models

The empirical findings of section 3 and 4 have strong implications for the modeling of financial time series. Taylor (1986) showed neither day-of-the-week effects nor a linear, correlated process can provide satisfactory explanation of the significant correlations among absolute return series, where a linear correlated process can be represented as

$$r_t = r + \sum_{i=0}^{\infty} \alpha_i e_{t-i}. \quad (9)$$

where r and α_i are constants with $\alpha_0 = 1$, e_t is a zero-mean i.i.d process. Taylor concludes that any reasonable model must be a non-linear one. Furthermore, the special autocorrelation pattern of $|r_t|^d$ found in section 3 implies that any theoretical model should be able to capture this before the model can be considered to be 'adequate'.

It should be noted that a process can have zero autocorrelations but have autocorrelations of squares greater than for moduli. For example, consider the following nonlinear model:

$$r_t = |s_t| e_t, \quad (10)$$

$$s_t = \alpha s_{t-\tau} + \eta_t,$$

where $e_t \sim N(0, 1)$, $E(s_t) = E(\eta_t) = E(r_t) = 0$, $|\alpha| < 1$, e_t and η_t are stochastically independent, s_t is independent of $\eta_{t+\tau}$ for $\tau > 0$, $s_t, s_{t-\tau}$ are jointly normally distributed with variance 1, hence we have $\text{var}(\eta_t) = 1 - \alpha^2$ and $\eta_t \sim N(0, 1 - \alpha^2)$. The conditional variance of r_t when s_t is known is s_t^2 , i.e. $\text{var}(r_t | s_t) = s_t^2$. For this model $\text{corr}(r_t, r_{t-\tau}) = 0$ but by using numerical integration it is found that with $|\alpha| < 1$

$$\text{corr}(|r_t|, |r_{t-\tau}|) = \frac{2}{\pi} \left(E |s_t s_{t-\tau} + \alpha s_t^2 - \frac{2}{\pi} \right) < \text{corr}(r_t^2, r_{t-\tau}^2) = \frac{\alpha^2}{4}. \quad (11)$$

It is thus seen that the results of table 3.2 do not necessarily occur.

One possible explanation for the large positive autocorrelation between $|r_t|$ and $|r_{t+\tau}|$ or $|r_t|^d$ and $|r_{t+\tau}|^d$ is the heteroskedasticity of the data, i.e. the variance or conditional variance is changing over time. One family of nonlinear time series models that is able to capture some aspects of the time varying volatility structure is Engle's ARCH (AutoRegressive Conditional Heteroskedasticity) model [Engle (1982)]. In its original setting, the ARCH model is defined as a data generating process for a random variable which has a conditional normal

distribution with conditional variance a linear function of lagged squared residuals. More formally, the ARCH(p) model is defined as follows:

$$\begin{aligned} r_t &= \mu + \varepsilon_t, \\ \varepsilon_t &= s_t e_t, \quad e_t \sim N(0, 1), \\ s_t^2 &= \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2. \end{aligned} \quad (12)$$

It is easily shown that r_t is not autocorrelated with each other but $|r_t|^d$ is. Hence the distribution of r_t is dependent on r_{t-i} , $i > 0$. Since its introduction by Engle (1982), the ARCH model has been widely used to model time-varying volatility and the persistence of shocks to volatility. Much work has also been done both theoretically and empirically. Many modifications and extensions of the original ARCH model have also appeared in the literature.

For example, in order to capture the long memory property of the conditional variance process, Bollerslev (1986) introduced the GARCH(p, q) model, which defines the conditional variance equation as follows:

$$s_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j s_{t-j}^2. \quad (13)$$

Taylor (1986) modeled the conditional standard deviation function instead of conditional variance. Schwert (1989), following the argument of Davidian and Carroll (1987), modeled the conditional standard deviation as a linear function of lagged absolute residuals. The Taylor/Schwert GARCH(p, q) model defines the conditional standard deviation equation as follows:

$$s_t = \alpha_0 + \sum_{i=1}^p \alpha_i |\varepsilon_{t-i}| + \sum_{j=1}^q \beta_j s_{t-j}. \quad (14)$$

One may, at first glance, think that it would be better to use Taylor/Schwert model than Bollerslev's GARCH since the model is expressed in terms of absolute returns rather than squared returns. But this conclusion is not necessarily true when the model is a nonlinear one. In fact, our Monte-Carlo study shows both Bollerslev's GARCH and Taylor/Schwert's model with appropriate parameters can produce the special correlation patterns found in section 3.

Both models were estimated for S&P 500 returns and the following results were obtained:

(1) GARCH

$$\begin{aligned} r_t &= 0.000438 + 0.144\varepsilon_{t-1} + \varepsilon_t, \\ (7.2) \quad (18.4) \end{aligned}$$

$$\begin{aligned} s_t^2 &= 0.0000008 + 0.091\varepsilon_{t-1}^2 + 0.906s_{t-1}^2, \\ (12.5) \quad (50.7) \quad (43.4) \end{aligned} \quad (15)$$

log likelihood: 56822.

(2) Taylor/Schwert

$$r_t = 0.0004 + 0.139\varepsilon_{t-1} + \varepsilon_t, \quad (7.0) \quad (19.6)$$

$$s_t = 0.000096 + 0.104|\varepsilon_{t-1}| + 0.913s_t. \quad (12.6) \quad (67) \quad (517) \quad (16)$$

log likelihood: 56776.

The first order moving average term is in the mean equations of both models to account for the positive first order autocorrelation for the return series. We can see all the parameters are very significant in the above models. The normality test statistic of the standardized residuals for both models are far beyond the critical value of a normal distribution as assumed by both models. This is not surprising since there are definitely other factors affecting the volatility. Nevertheless, the log-likelihood value for Bollerslev's GARCH is significantly larger than that of Taylor/Schwert model.

Based on the estimation results, some simulations have performed using the parameters estimated above assuming $e_t \sim IID N(0, 1)$. Our purpose is to check whether theoretical ARCH models can generate the same type of autocorrelations as stock market return data. Obviously if the theoretical model does not exhibit the same pattern of autocorrelations as stock market return data, then it follows that the theoretical model is misspecified for these data. A total of 18054 observations was generated and the first 1000 were discarded in order to be less affected by the initial value of s_0 which was set to be the unconditional standard deviation of the S&P 500 returns. Figs. 4.1, 4.2, 4.3 and 4.4, 4.5, 4.6 plot the simulated autocorrelogram of the data generated by the two models. It can be seen that the special autocorrelation pattern does exist here. For both models, $|r|^d$ has the largest autocorrelation when $d=1$, and the autocorrelation gets smaller when d goes away from 1. It is interesting that Bollerslev's GARCH model can produce this result even though the conditional variance is a linear function of squared returns. For Bollerslev's GARCH model, the autocorrelation between $|r_t|$ and $|r_{t+\tau}|$ is very close to that between $|r_t|^{1.25}$ and $|r_{t+\tau}|^{1.25}$. But

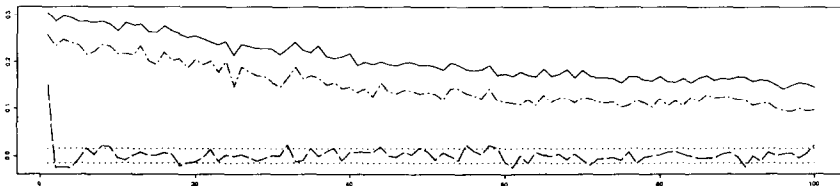
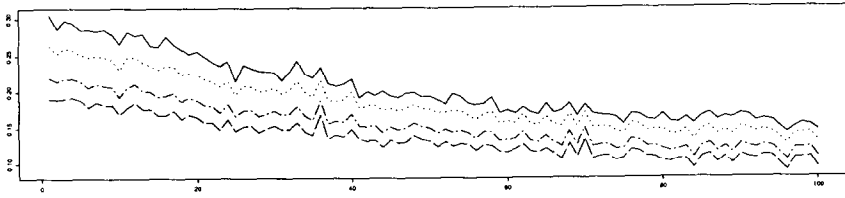
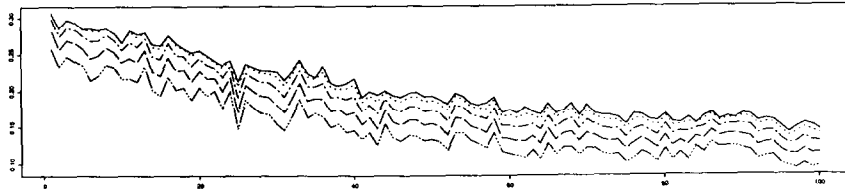
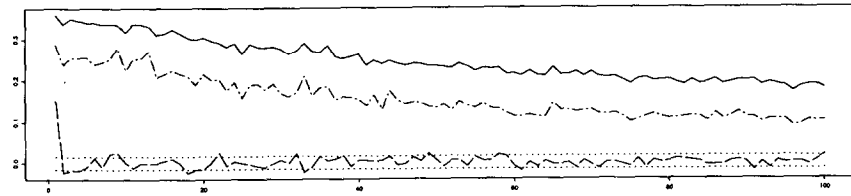
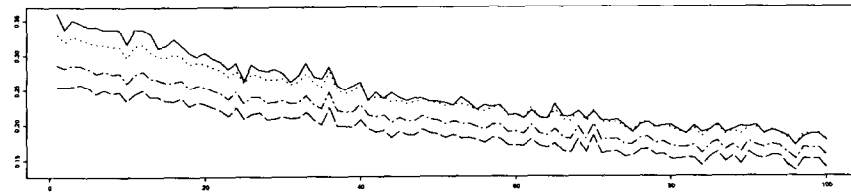


Fig. 4.1. Bollerslev's GARCH model. Autocorrelation of $|r|$, r^{**2} , r from high to low.

Fig. 4.2. Bollerslev's GARCH model. $d = 1, 0.5, 0.25, 0.125$ from high to low.Fig. 4.3. Bollerslev's GARCH model. $d = 1, 1.25, 1.50, 2$ from high to low.Fig. 4.4. Taylor/Schwert model. Autocorrelation of $|r|$, r^{**2} , r from high to low.Fig. 4.5. Taylor/Schwert model. $d = 1, 0.5, 0.25, 0.125$ from high to low.

for the Taylor/Schwert model, the autocorrelation between $|r_t|$ and $|r_{t+\tau}|$ after lag 40 is close to that between $|r_t|^{0.5}$ and $|r_{t+\tau}|^{0.5}$. One major difference between autocorrelograms of the two simulated data series and the real data is that the autocorrelations of the real data decreases rapidly in the first month and then decrease very slowly over a long period, but the autocorrelations of the two simulated data decrease almost constantly over time.

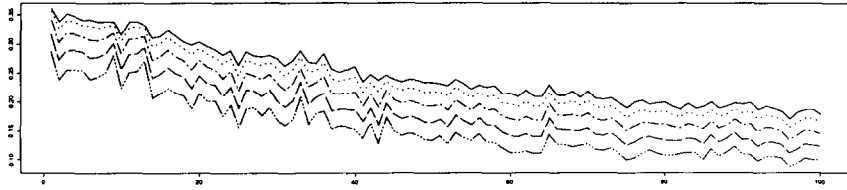


Fig. 4.6. Taylor/Schwert model. $d = 1, 1.25, 1.50, 1.75, 2$ from high to low.

6. A new model – Asymmetric power ARCH

The Monte-Carlo study shows that the ARCH model generally captures the special pattern of autocorrelation existing in many stock market returns data. Both Bollerslev's GARCH and Taylor/Schwert's GARCH in absolute value model can produce this property. It seems there is no obvious reason why one should assume the conditional variance is a linear function of lagged squared returns (residuals) as in Bollerslev's GARCH, or the conditional standard deviation a linear function of lagged absolute returns (residuals) as in Taylor/Schwert model. Fortunately, a more general class of model is available which includes Bollerslev's GARCH, Taylor/Schwert and five other models in the literature as special cases. The general structure is as follows:

$$\begin{aligned} \varepsilon_t &= s_t e_t, \quad e_t \sim N(0, 1), \\ s_t^\delta &= \alpha_0 + \sum_{i=1}^p \alpha_i (|e_{t-i}| - \gamma_i e_{t-i})^\delta + \sum_{j=1}^q \beta_j s_{t-i}^\delta, \quad \text{where} \\ \alpha_0 &> 0, \delta \geq 0, \\ \alpha_i &\geq 0, i = 1, \dots, p, \\ -1 &< \gamma_i < 1, i = 1, \dots, p, \\ \beta_j &\geq 0, j = 1, \dots, q. \end{aligned} \tag{17}$$

The model imposes a Box–Cox power transformation of the conditional standard deviation process and the asymmetric absolute residuals. By using this transformation we can linearize otherwise nonlinear models. The functional form for conditional standard deviation is familiar to economists as the constant elasticity of substitution (CES) production function of Arrow et al. (1961).

The asymmetric response of volatility to positive and negative 'shocks' is well known in the finance literature as the leverage effect of the stock market returns [Black (1976)], which says that stock returns are negatively correlated with changes in return volatility – i.e. volatility tends to rise in response to 'bad news' (excess returns lower than expected) and to fall in response to 'good news'

(excess returns higher than expected) [Nelson (1991)]. Empirical studies by Nelson (1991), Glosten, Jaganathan and Runkle (1989) and Engle and Ng (1992) show it is crucial to include the asymmetric term in financial time series models [for a detailed discussion, see Engle and Ng (1992)].

This generalized version of ARCH model includes seven other models (see appendix A) as special cases. We will call this model Asymmetric Power ARCH model and denote it as A-PARCH.

If we assume the distribution of r_t is conditionally normal, then the condition for existence of Es_t^δ and $E|\varepsilon_t|^\delta$ is (see appendix B):

$$\frac{1}{\sqrt{2\pi}} \sum_{i=1}^p \alpha_i \{ (1+\gamma_i)^\delta + (1-\gamma_i)^\delta \} 2^{\frac{\delta-1}{2}} \Gamma\left(\frac{\delta+1}{2}\right) + \sum_{j=1}^q \beta_j < 1. \quad (18)$$

If this condition is satisfied, then when $\delta \geq 2$ we have ε_t covariance stationary. But $\delta \geq 2$ is a sufficient condition for ε_t to be covariance stationary.

The new model is estimated for S&P 500 return series by the maximum likelihood method using the Berndt–Hall–Hall–Hausman algorithm. The estimated model is as follows:

$$r_t = 0.00021 + 0.145\varepsilon_{t-1} + \varepsilon_t, \quad (3.2) \quad (19.0)$$

$$s_t^{1.43} = 0.000014 + 0.083(|\varepsilon_{t-1}| - 0.373\varepsilon_{t-1})^{1.43} + 0.920s_t^{1.43}, \quad (4.5) \quad (32.4) \quad (-20.7) \quad (474)(33.7) \quad (19)$$

log likelihood: 56974.

The estimated δ is 1.43 which is significantly different from 1 (Taylor/Schwert model) or 2 (Bollerslev GARCH). The t -statistic for the asymmetric term is 32.4 which is very significant implying the leverage effect does exist in S&P 500 returns. By using the log-likelihood values estimated, a nested test can easily be constructed against either Bollerslev's GARCH or Taylor/Schwert model. Let l_0 be the log-likelihood value under the null hypothesis that the true model is Bollerslev's GARCH and l be the log-likelihood value under the alternative that the true model is A-PARCH, then $2(l-l_0)$ should have a χ^2 distribution with 2 degrees of freedom when the null hypothesis is true. But in our example $2(l-l_0) = 2(56974 - 56822) = 304$ which is far beyond the critical value at any reasonable level. Hence we can reject that the data is generated by Bollerslev's GARCH model. The same procedure is applicable to Taylor/Schwert model and we can also reject it.

6. Conclusion

In this paper, a 'long-memory' property of the stock market returns series is investigated. We found not only there is substantially more correlation between absolute returns than returns themselves, but the power transformation of the absolute return $|r_t|^d$ also has quite high autocorrelation for long lags. Furthermore, for fixed lag τ , the function $\rho_\tau(d) = \text{corr}(|r_t|^d, |r_{t+\tau}|^d)$ has a unique maximum point when d is around 1. This result appears to argue against ARCH type specifications based upon squared returns. But our Monte-Carlo study shows both ARCH type of model based upon squared return and those based upon absolute return can produce this property. The ARCH specification based upon the linear relationship among absolute returns is neither necessary nor sufficient to have such a property. Finally, we propose a new general class of ARCH models which we call Asymmetric Power ARCH model and denote A-PARCH. The new model encompasses seven other models in the literature. We estimate S&P 500 returns by the new model and the estimated power δ for the conditional heteroskedasticity function is 1.43 which is significantly different from 1 (Taylor/Schwert model) or 2 (Bollerslev's GARCH).

Appendix A

We now show that the new model includes the following seven ARCH models as a special case.

- (1) Engle's ARCH(p) model [see Engle (1982)], just let $\delta=2$ and $\gamma_i=0$, $i=1, \dots, p$, $\beta_j=0$, $j=1, \dots, q$ in the new model.
- (2) Bollerslev's GARCH(p, q) model (see Bollerslev 1986), let $\delta=2$ and $\gamma_i=0$, $i=1, \dots, p$.
- (3) Taylor/Schwert's GARCH in standard deviation model let $\delta=1$ and $\gamma_i=0$, $i=1, \dots, p$.
- (4) GJR model [see Glosten et al. (1989)], let $\delta=2$.

When $\delta=2$, we have when $0 \leq \gamma_i < 1$

$$\begin{aligned}
 s_t^2 &= \alpha_0 + \sum_{i=1}^p \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^2 + \sum_{j=1}^q \beta_j s_{t-j}^2 \\
 &= \alpha_0 + \sum_{i=1}^p \alpha_i (1 - \gamma_i)^2 \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j s_{t-j}^2 \\
 &\quad + \sum_{i=1}^p \alpha_i \{ (1 + \gamma_i)^2 - (1 - \gamma_i)^2 \} S_i^- \varepsilon_{t-i}^2
 \end{aligned}$$

$$= \alpha_0 + \sum_{i=1}^p \alpha_i (1 - \gamma_i)^2 \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j s_{t-j}^2 + \sum_{i=1}^p 4\alpha_i \gamma_i S_i^- \varepsilon_{t-i}^2, \quad \text{where}$$

$$S_i^- = \begin{cases} 1 & \text{if } \varepsilon_{t-i} < 0 \\ 0 & \text{otherwise.} \end{cases}$$

If we further define

$$\alpha_i^* = \alpha_i (1 - \gamma_i)^2,$$

$$\gamma_i^* = 4\alpha_i \gamma_i,$$

then we have

$$s_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i^* \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j s_{t-j}^2 + \sum_{i=1}^p \gamma_i^* S_i^- \varepsilon_{t-i}^2$$

which is exactly the GJR model.

When $-1 < \gamma_i < 0$ we have

$$s_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i (1 + \gamma_i)^2 \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j s_{t-j}^2 - \sum_{i=1}^p 4\alpha_i \gamma_i S_i^+ \varepsilon_{t-i}^2, \quad \text{where}$$

$$S_i^+ = \begin{cases} 1 & \text{if } \varepsilon_{t-i} > 0 \\ 0 & \text{otherwise,} \end{cases}$$

define

$$\alpha_i^* = \alpha_i (1 + \gamma_i)^2,$$

$$\gamma_i^* = -4\alpha_i \gamma_i,$$

we have

$$s_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i^* \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j s_{t-j}^2 + \sum_{i=1}^p \gamma_i^* S_i^+ \varepsilon_{t-i}^2$$

which allows positive shocks to have a stronger effect on volatility.

(5) Zakoian's TARCh model (see Zakoian 1991), let $\delta = 1$ and $\beta_j = 0$, $j = 1, \dots, q$. We have

$$\begin{aligned} s_t &= \alpha_0 + \sum_{i=1}^p \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i}) \\ &= \alpha_0 + \sum_{i=1}^p \alpha_i (1 - \gamma_i) \varepsilon_{t-i}^+ - \sum_{i=1}^p \alpha_i (1 + \gamma_i) \varepsilon_{t-i}^-, \quad \text{where} \end{aligned}$$

$$\varepsilon_{t-i}^+ = \begin{cases} \varepsilon_{t-i} & \text{if } \varepsilon_{t-i} > 0 \\ 0 & \text{otherwise,} \end{cases} \quad \text{and}$$

$$\varepsilon_{t-i}^- = \varepsilon_{t-i} - \varepsilon_{t-i}^+.$$

So by defining

$$\alpha_i^+ = \alpha_i(1 - \gamma_i),$$

$$\gamma_i^- = \alpha_i(1 + \gamma_i),$$

we have

$$s_t = \alpha_0 + \sum_{i=1}^p \alpha_i^+ \varepsilon_{t-i}^+ - \sum_{i=1}^p \alpha_i^- \varepsilon_{t-i}^-$$

which is the exact TARCH form. If we further let $\beta_j \neq 0, j = 1, \dots, q$ then we get a more general class of TARCH models.

(6) Higgins and Bera's NARCH model [see Higgins and Bera (1990)], let $\gamma_i = 0, i = 1, \dots, p$ and $\beta_j = 0, j = 1, \dots, q$.

Our model becomes

$$s_t^\delta = \alpha_0 + \sum_{i=1}^p \alpha_i |\varepsilon_{t-i}|^\delta, \quad \text{i.e.}$$

$$(s_t^\delta)^\delta / 2 = \alpha_0 + \sum_{i=1}^p \alpha_i (\varepsilon_{t-i}^2)^{\delta/2}.$$

Define

$$\delta^* = \delta/2,$$

$$\alpha_0 = \alpha_0^* \omega^{\delta/2} = \left(1 - \sum_{i=1}^p \alpha_i\right) \omega^{\delta^*}.$$

We have exactly Higgins and Bera's NARCH.

(7) Geweke (1986) and Pantula (1986)'s log-ARCH model. The log-ARCH model is the limiting case of our model when $\delta \rightarrow 0$.

Since

$$s_t^\delta = \alpha_0 + \sum_{i=1}^p \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^\delta + \sum_{j=1}^q \beta_j s_{t-j}^\delta,$$

decompose α_0 as:

$$\begin{aligned} \alpha_0 &= \left\{ 1 - \sum_{i=1}^p \alpha_i E(|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^\delta - \sum_{j=1}^q \beta_j \right\} \omega^\delta \\ &= \alpha_0^* \omega^\delta, \end{aligned}$$

hence $Es_t^\delta = \omega^\delta$. Then we have

$$\frac{s_t^\delta - 1}{\delta} = \left\{ 1 - \sum_{i=1}^p \alpha_i E(|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^\delta - \sum_{j=1}^q \beta_j \right\} \frac{(\omega^\delta - 1)}{\delta}$$

$$\begin{aligned}
& + \sum_{i=1}^p \alpha_i \frac{(|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^\delta - 1}{\delta} + \sum_{j=1}^q \beta_j \frac{s_{t-j}^\delta - 1}{\delta} \\
& - \sum_{i=1}^p \alpha_i \frac{E(|e_{t-i}| - \gamma_i e_{t-i})^\delta - 1}{\delta},
\end{aligned}$$

when $\delta \rightarrow 0$ the above equation becomes

$$\begin{aligned}
\log s_t &= \left\{ 1 - \sum_{i=1}^p \alpha_i \lim_{\delta \rightarrow 0} E(|e_{t-i}| - \gamma_i e_{t-i})^\delta - \sum_{j=1}^q \beta_j \right\} \log \omega \\
& + \sum_{i=1}^p \alpha_i \log(|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i}) + \sum_{j=1}^q \beta_j \log s_{t-j} \\
& - \sum_{i=1}^p \alpha_i \log E(|e_{t-i}| - \gamma_i e_{t-i}) \\
& = \alpha_0^* \log \omega - \sum_{i=1}^p \alpha_i \log \sqrt{2/\pi} + \sum_{i=1}^p \alpha_i \log(|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i}) \\
& + \sum_{j=1}^q \beta_j \log s_{t-j},
\end{aligned}$$

where

$$\begin{aligned}
\alpha_0^* &= \left\{ 1 - \sum_{i=1}^p \alpha_i \lim_{\delta \rightarrow 0} E(|e_{t-i}| - \gamma_i e_{t-i})^\delta - \sum_{j=1}^q \beta_j \right\} \\
&= \left\{ 1 - \sum_{i=1}^p \alpha_i - \sum_{j=1}^q \beta_j \right\},
\end{aligned}$$

since $\lim_{\delta \rightarrow 0} E(|e_{t-i}| - \gamma_i e_{t-i})^\delta = 1$. This is a generalized version of Geweke/Pantula model. If we further let $\gamma_i = 0, i = 1, \dots, p$, and $\beta_j = 0, j = 1, \dots, q$, then we get the exact Geweke/Pantula model.

Appendix B. Conditions for the existence of Es_t^δ and $E|\varepsilon_t|^\delta$

If we assume the distribution is conditional normal, then the condition for existence of Es_t^δ of the new model is

$$\begin{aligned}
& \sum_{i=1}^p \alpha_i E(|e_{t-i}| - \gamma_i e_{t-i})^\delta + \sum_{j=1}^q \beta_j < 1, \quad \text{where} \\
& E(|e_{t-i}| - \gamma_i e_{t-i})^\delta = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (|x| - \gamma_i x)^\delta e^{-\frac{x^2}{2}} dx \\
& = \frac{1}{\sqrt{2\pi}} [(1 + \gamma_i)^\delta + (1 - \gamma_i)^\delta] 2^{\frac{\delta-1}{2}} \Gamma\left(\frac{\delta+1}{2}\right).
\end{aligned}$$

So the condition becomes

$$-\frac{1}{\sqrt{2\pi}} \sum_{i=1}^p \alpha_i [(1+\gamma_i)^\delta + (1-\gamma_i)^\delta] 2^{\frac{\delta-1}{2}} \Gamma\left(\frac{\delta+1}{2}\right) + \sum_{j=1}^q \beta_j < 1. \quad (\text{B1})$$

Since

$$\begin{aligned} E|e_t|^\delta &= E|e_t|^\delta Es_t^\delta \\ &= \frac{1}{\sqrt{\pi}} 2^{\frac{\delta}{2}} \Gamma\left(\frac{\delta+1}{2}\right) Es_t^\delta. \end{aligned}$$

So the condition for the existence of $E|e_t|^\delta$ is the same as that of Es_t^δ . The proof of the above result is almost identical to the proof of theorem 1 in Bollerslev (1986).

When condition (B1) is satisfied, we have the unconditional expectation of s_t^δ as follows

$$\begin{aligned} Es_t^\delta &= \alpha_0 \left/ \left(1 - \sum_{i=1}^p \alpha_i E(|e_{t-i}| - \gamma_i e_{t-i})^\delta - \sum_{j=1}^q \beta_j \right) \right. \\ &= \omega^\delta \quad \text{and} \\ E|e_t|^\delta &= \frac{1}{\sqrt{\pi}} 2^{\frac{\delta}{2}} \Gamma\left(\frac{\delta+1}{2}\right) Es_t^\delta \\ &= \frac{1}{\sqrt{\pi}} 2^{\frac{\delta}{2}} \Gamma\left(\frac{\delta+1}{2}\right) \omega^\delta. \end{aligned}$$

In its special case, when $\delta=2$ and $\gamma_i=0$, we have the covariance stationarity condition for ε_t as

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \sum_{i=1}^p \alpha_i 2^{\frac{2-1}{2}} \Gamma\left(\frac{2+1}{2}\right) + \sum_{j=1}^q \beta_j &= \frac{1}{\sqrt{\pi}} \sum_{i=1}^p \alpha_i 2^{\frac{1}{2}} \Gamma\left(\frac{1}{2}\right) + \sum_{j=1}^q \beta_j \\ &= \sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1 \end{aligned}$$

which is the same as that derived by Bollerslev (1986).

When $\delta=2$ and $\gamma_i \neq 0$, we have the covariance stationarity condition for GJR model as

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \sum_{i=1}^p \alpha_i [(1+\gamma_i)^2 + (1-\gamma_i)^2] 2^{\frac{2-1}{2}} \Gamma\left(\frac{2+1}{2}\right) + \sum_{j=1}^q \beta_j \\ = \sum_{i=1}^p \alpha_i [1 + \gamma_i^2] + \sum_{j=1}^q \beta_j < 1. \end{aligned}$$

When $\delta=1$ and $\gamma_i=0$, we have the condition for existence of Es_t and $E|e_t|$ of Taylor/Schwert model

$$\begin{aligned} & \frac{1}{\sqrt{2\pi}} \sum_{i=1}^p \alpha_i 2\Gamma(1) + \sum_{j=1}^q \beta_j \\ &= \sqrt{2/\pi} \sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1, \end{aligned}$$

Since $\sqrt{(2/\pi)} < 1$, so even if $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j > 1$ it can still be true that Es_t or $E|e_t|$ exists and is finite, this condition is weaker than the covariance stationarity condition of the model. It is possible $E|e_t|^2$ does not exist and e_t is not covariance stationary even if this condition is satisfied.

When $\delta=1$ and $\gamma_i \neq 0$, we have the existence condition of Es_t^δ and $E|e_t|$ for the Asymmetric Taylor/Schwert model or the generalized Zakoian model which is the same as that for the Taylor/Schwert model $\sqrt{2/\pi} \sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$.

Under the assumption that

$$\frac{1}{\sqrt{2\pi}} \sum_{i=1}^p \alpha_i [(1+\gamma_i)^\delta + (1-\gamma_i)^\delta] 2^{\frac{\delta-1}{2}} \Gamma\left(\frac{\delta+1}{2}\right) + \sum_{j=1}^q \beta_j < 1,$$

i.e. the δ th moment of s_t and $|e_t|$ exist, then if $\delta \geq 2$ we have that e_t is covariance stationary. If $\delta \geq 1$ then Es_t and $E|e_t|$ exist and are finite. But $\delta \geq 2$ is a sufficient condition for the process e_t to be covariance stationary.

References

- Black, Fisher, 1976, Studies in stock price volatility changes, Proceedings of the 1976 business meeting of the business and economics statistics section, American Statistical Association, 177–181.
- Bollerslev, T., 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 31, 307–327.
- Bollerslev, T. and R.F. Engle, 1992, Common persistence in conditional variance, *Forthcoming in Econometrica*.
- Davidian, M. and R.J. Carroll, 1987, Variance function estimation, *Journal of American Statistical Association* 82, No. 400, 1079–1091.
- Eatwell, J., M. Milgate and P. Newman (eds.), *The new palgrave: Finance* (New York, Norton).
- Engle, R.F. 1982, Autoregressive Conditional heteroskedasticity with estimates of the variance of U.K. Inflation, *Econometrica*, 50, 987–1008.
- Engle, R.F. 1990, Discussion: stock volatility and the crash of '87, *Review of Financial Studies*, Volume 3, Number 1, 103–106.
- Engle, R.F. and T. Bollerslev, 1986, Modelling the persistence of conditional variances, *Econometric Review*, 5, 1–50, 81–87.
- Engle, R.F. David Lilien and Russ Robins, 1987, Estimating time varying risk premia in the term structure: The ARCH-M Model, *Econometrica* 55, 391–407.
- Engle, R.F. and G. Gonzalez Rivera, 1991, Semiparametric ARCH models, *Journal of Business and Economic Statistics* 9, 345–360.

- Engle, R.F. and V. Ng, 1992, Measuring and testing the impact of news on volatility, *Forthcoming in Journal of Finance*.
- Fama, E.F., 1970, Efficient capital markets: A Review of Theory and Empirical Work, *Journal of Finance*, 25, 383–417.
- Fama, E.F., 1976, *Foundations of finance: Portfolio decision and security prices*, New York: Basic Books Inc.
- French, Ken, William Schwert and Robert Stambaugh, 1986, Expected stock returns and volatility, *Journal of Financial Economics*, 19, 3–29.
- Glosten, L., R. Jaganathan and D. Runkle, 1989, Relationship between the expected value and the volatility of the nominal excess return on stocks, unpublished manuscript, J.L. Kellogg Graduate School, Northwestern University.
- Granger, C.W.J., 1991, Forecasting stock market prices: Lessons for forecasters, UCSD Working Paper.
- Granger, C.W.J., 1980, Long memory relationships and the aggregation of dynamic models. *J. of Econometrics*, 14, 227–238.
- Granger, C.W.J. and A.P. Anderson, 1978, An introduction to bilinear time series model, Vandenhoeck and Ruprecht, Göttingen.
- Granger, C.W.J. and R. Joyeux, 1981, An introduction to long-memory time series models and fractional differencing, *J. of Time Series Analysis*, 1, 15–29.
- Granger, C.W.J. and O. Morgenstern, 1970, Predictability of stock market prices. Heath-Lexington Press.
- Granger, C.W.J. and Paul Newbold, 1986, *Forecasting economic time series*, New York, Academic Press.
- Hamao, Y., R.W. Masulis, V. Ng, 1990, Correlation in price changes and volatility across international stock markets, *The Review of Financial Studies*, V3, 2, 281–307.
- Higgins, M. and A. Bera, 1990, A class of nonlinear ARCH models, Working Paper, Department of Economics, University of Wisconsin at Milwaukee.
- Kariya, T. Tsukuda, Y. Maru, J., 1990, Testing the random walk hypothesis for Japanese stock price in S. Taylor's Model, Working paper, University of Chicago.
- Nelson, D.B., 1990, Stationarity and persistence in the GARCH(1,1) model, *Econometric Theory*, 6, 318–334.
- Nelson, D.B., 1991, Conditional heteroskedasticity in asset returns: A New Approach, *Econometrica*, Vol. 59, No. 2, 347–370.
- Schwert, W., 1990, Stock volatility and the crash of '87, *Review of Financial Studies*, Vol. 3, No. 1, 77–102.
- Taylor, S., 1986, *Modelling financial time series*, New York, John Wiley & Sons.
- Zakoian, J., 1991, Threshold heteroskedasticity model, unpublished manuscript, INSEE.