ELSEVIER

Journal of Econometrics 73 (1996) 185-215

JOURNAL OF Econometrics

Modeling volatility persistence of speculative returns: A new approach

Zhuanxin Ding[a], Clive W.J. Granger*,[b]

[a] *Frank Russell Company, Tacoma, WA 98401, USA*

[b] *Department of Economics, University of California, San Diego, La Jolla, CA 92093-0508, USA*

Abstract

This paper extends the work by Ding, Granger, and Engle (1993) and further examines the long memory property for various speculative returns. The long memory property found for S&P 500 returns is also found to exist for four other different speculative returns. One significant difference is that for foreign exchange rate returns, this property is strongest when $d = 1/4$ instead of at $d = 1$ for stock returns. The theoretical autocorrelation functions for various GARCH(1,1) models are also derived and found to be exponentially decreasing, which is rather different from the sample autocorrelation function for the real data. A general class of long memory models that has no memory in returns themselves but long memory in absolute returns and their power transformations is proposed. The issue of estimation and simulation for this class of models is discussed. The Monte Carlo simulation shows that the theoretical model can mimic the stylized empirical facts strikingly well.

Key words: Long memory; Volatility persistence; Autocorrelation; Power transformation; Aggregation; Long memory ARCH model
JEL classification: C1; C4; C11

1. Introduction

With the availability of high-frequency long time series from returns of speculative asset, much research has been devoted to the study of long-run behavior of financial data. A common finding in much of the empirical literature

* Corresponding author.

We would like to thank Xiaohong Chen, Rob Engle, an editor (Richard T. Baillie), two anonymous referees, participants at workshops at UCSD and the University of Chicago for many helpful comments.

SSDI 0304-4076(95)01737-X

is that the returns themselves contain little serial correlation which is in agreement with the efficient market hypotheses. However, it is also found that the absolute returns and their power transformations are highly correlated. A systematic study of this can be found in Taylor (1986). These empirical regularities culminated by the publication of Ding, Granger, and Engle (1993). In that paper, they investigate the autocorrelation structure of $|r_t|^d$, where $|r_t|$ is the daily S&P 500 stock market returns, and d is a positive number. It is found that $|r_t|$ has significant positive autocorrelations at over 2,700 lags with a series of 17,054 observations. Similar results are also found for other values of d in $|r_t|^d$. This property is found to be strongest when $d = 1$ compared to both smaller or larger d values. These properties are examined for several other long speculative asset returns series, including returns for the Japanese stock market index Nikkei, foreign exchange rate returns for the Deutschmark with the US dollar, individual stock returns for Chevron, and minute-by-minute stock returns for a Japanese food company, Ajinomoto. The results show the long memory property found for S&P 500 returns in Ding, Granger, and Engle (1993) also exist here. One significant difference is that for foreign exchange rate returns, this property is strongest when $d = \frac{1}{4}$ instead of at $d = 1$ for stock returns.

It is also quite common view that the volatility persistence is best represented by the fact that the estimated ARCH and GARCH parameters in a GARCH model sum to very close to one. So the Integrated GARCH(1, 1) model comes out very naturally to model the volatility persistence (see Engle and Bollerslev, 1986). However, in Section 3, we prove that the autocorrelation function for an IGARCH(1, 1) process is exponentially decreasing and is very different from the sample autocorrelation function found for various speculative returns in Section 2. In Section 4, a new general class of models that has no memory in returns themselves but long memory in absolute returns and their power transformations is proposed. The relationship between this class of models and other models is also discussed. Estimation and simulation results for S&P 500 stock market returns using this class of models are presented. The Monte Carlo simulation shows that the theoretical model can mimic the stylized empirical facts strikingly well. Section 5 concludes the analysis.

2. Autocorrelation analysis of various financial time series

In Ding, Granger, and Engle (1993), the sample autocorrelation function for S&P 500 daily stock market returns and their transformations are presented. There, a long memory property for $|r_t|^d$ is established and is found to be strongest for $d = 1$ compared to other values of d. For the convenience of later reference and comparison, Fig. 1 plots the sample autocorrelations for r_t, $|r_t|$, and r_t^2 of S&P 500 stock market returns up to lag 2,500. As pointed out by Ding, Granger, and Engle, the sample autocorrelation function for absolute returns

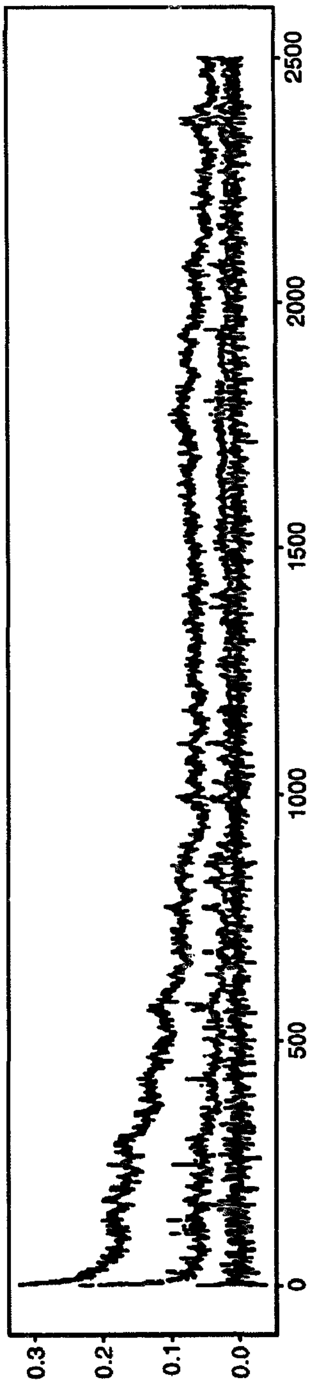

Fig. 1. Autocorrelation of $|r|$, r^2, and r, from high to low: Standard & Poor 500 daily returns, 1/3/28–8/30/91.

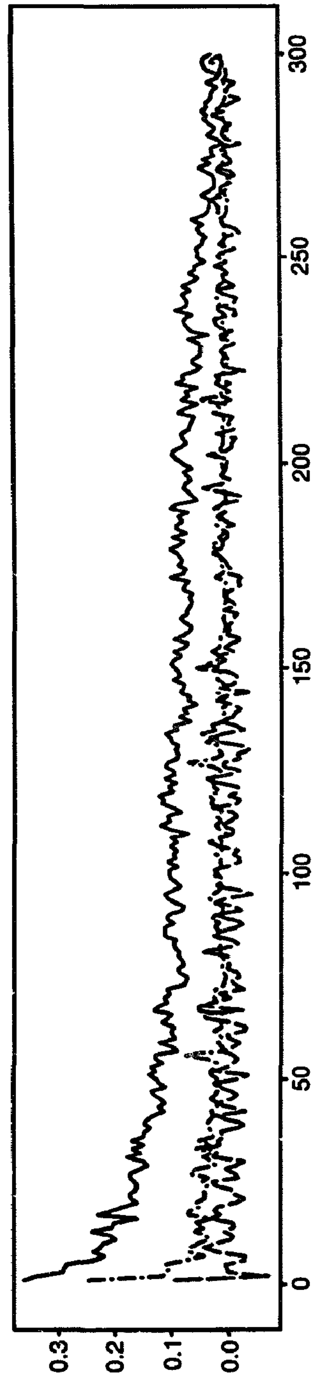

Fig. 2. Autocorrelation of $|r|$, r^2, and r, from high to low: Japanese stock market index Nikkei daily returns, 1970–1992.

and squared returns remains significantly positive for all these lags, while most sample autocorrelations for r_t are not significantly different from zero. The sample autocorrelations for $|r_t|$ are consistently higher than that for r_t^2. It is seen that the sample autocorrelation function decreases very fast at the beginning, and then decreases very slowly and remains significantly positive, which is different from an exponentially decreasing function. Figs. 2 to 5 plot the sample autocorrelations for four other speculative returns: daily returns for Japanese stock market index Nikkei from 1970 to 1992 with 5,594 observations; daily foreign exchange rate returns for the Deutschmark with the US dollar from January 1971 to March 1992 with 5,311 observations; daily individual stock returns for Chevron from July 1962 to December 1991 with 7,420 observations; and minute-by-minute (not necessarily equal time interval) stock returns for a Japanese food company, Ajinomoto, from April 3, 1989 to April 30, 1992 with a total of 25,099 observations. It is clearly seen that the striking regularities found for S&P 500 also exist for all these four different speculative returns. The patterns of the sample autocorrelations are very similar to each other. Fig. 3 also plots the sample autocorrelation functions for the absolute foreign exchange rate returns raising to power $\frac{1}{4}$. It is found that, different from that of stock returns, this property is stronger at $\frac{1}{4}$ than any other values for foreign exchange rate returns.

Table 1 gives the summary statistics for the above five financial series. The mean returns for all five series are very close to zero. The returns for S&P 500, Nikkei, and DM/US are negatively skewed, while the returns for the two individual stocks, Chevron and Ajinomoto, are positively skewed. All five returns are leptokurtic in the sense that the kurtosis for all these returns are bigger than that of a normal distribution, which is 3. The normality test refers to the Jarque–Bera normality test, and the test statistics show that all the five return series are not normal.

Table 2 further gives the numerical values of the sample autocorrelations for r_t, $|r_t|$, and r_t^2 at lags 1, 2, 3, 4, 5, 10, 100, 200, and 500 for these five returns.The last column (Neg. lag) gives the lag at which the first negative sample autocorrelation occurs for the corresponding series. The numbers in parentheses in the first column are two times the standard errors of the sample autocorrelation for the corresponding series if they are not correlated and have finite variances. It is seen that for the return series r_t only one (for exchange rate return), or two (for S&P 500, Nikkei, and Chevron), or at most four (for Ajinomoto) lags of sample autocorrelations of those shown are significantly different from zero. However, $|r_t|$ and r_t^2 have many lags of significantly positive sample autocorrelations. Usually the number of positive sample autocorrelations is increasing with the sample size. For various values of d (not shown here), $|r_t|^d$ has the largest autocorrelation when $d = 1$, except for foreign exchange rate returns, which has the strongest sample autocorrelation when $d = \frac{1}{4}$. This last property, the power transformation has the strongest autocorrelation when $d = 1$ for S&P 500 and

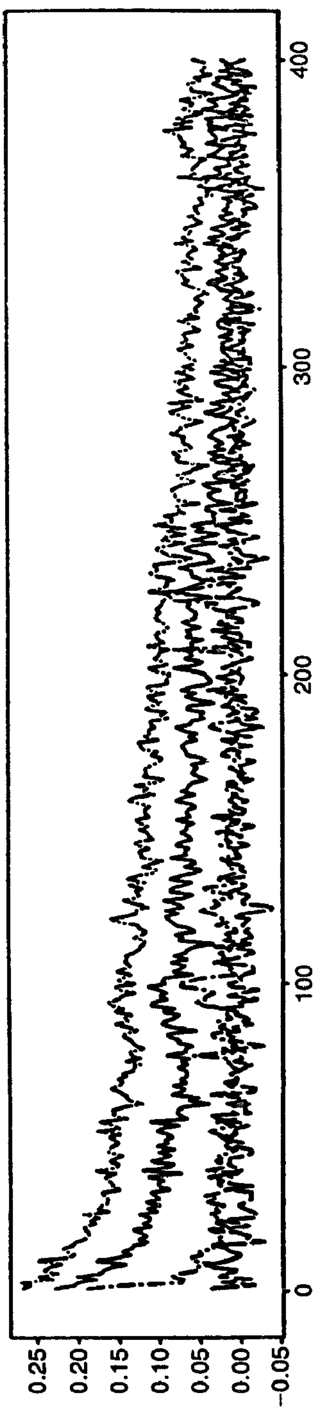

Fig. 3. Autocorrelation of $|r|$, r^2, and r, from high to low: Foreign exchange rate returns, DM/US$, 1/1971–3/1992, highest one for $|r|^{1/4}$.

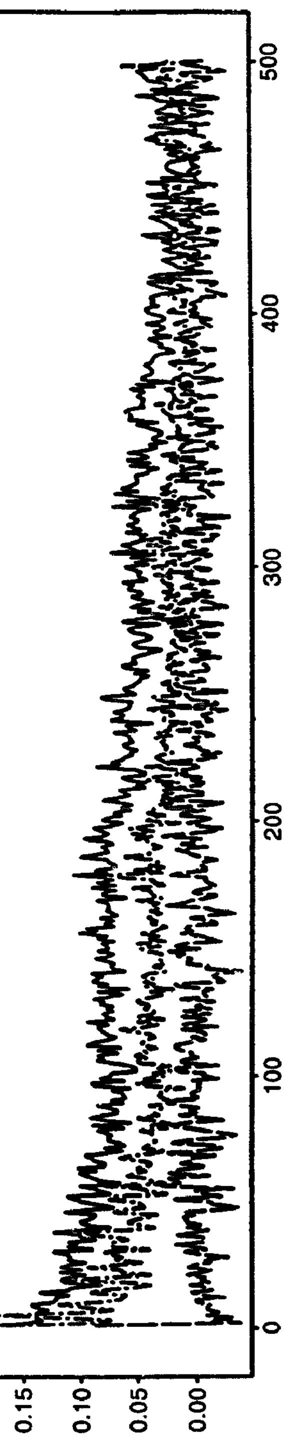

Fig. 4. Autocorrelation of $|r|$, r^2, and r, from high to low: Chevron individual daily stock returns, 7/62–12/91.

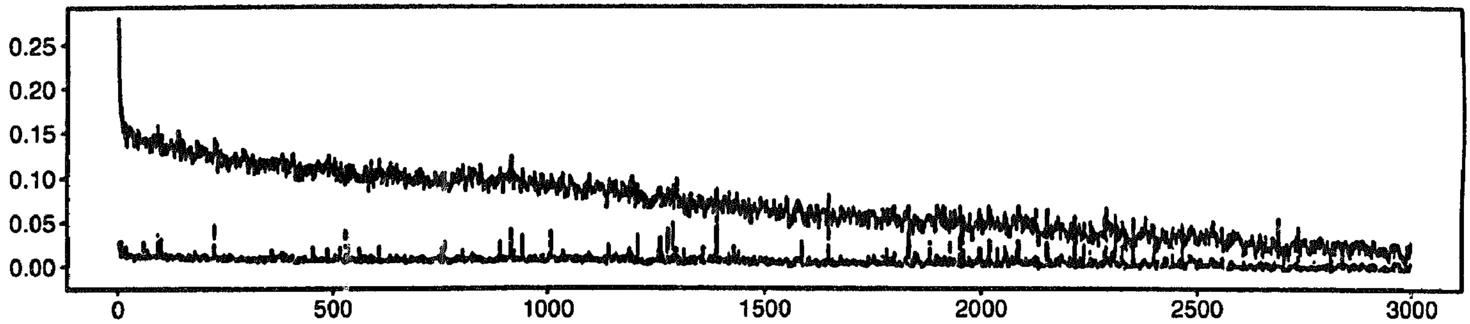

Fig. 5. Autocorrelation of $|r|$, r^2, and r, from high to low: Minutes by minutes stock returns for Ajinomoto, 4/3/89–4/30/92.

Table 1
Summary statistics of r_t

Data	Sample size	Mean	Std.	Skew.	Kurt.	Min.	Max.	Range	Normality test
S&P 500	17,054	0.0002	0.0115	− 0.487	25.42	0.154	− 0.228	0.382	357,788
NIKKEI	5,594	0.0000	0.0091	− 0.799	33.96	0.124	− 0.162	0.286	223,978
DM/US$	5,311	− 0.0001	0.0065	− 0.062	9.21	0.059	− 0.062	0.121	8,534
CHEVRON	7,420	0.0005	0.0152	0.357	5.91	0.105	− 0.079	0.184	2,092
AJINOMOTO	25,099	0.0000	0.0079	0.730	20.18	0.181	− 0.067	0.248	310,733

Table 2
Autocorrelations of r_t

Name	Data	Lag 1	2	3	4	5	10	100	200	500	Neg. lag
S&P 500	r_t	0.063	−0.039	−0.004	0.031	0.022	0.018	0.004	0.012	0.011	2
17,054	$\|r_t\|$	0.3178	0.323	0.322	0.296	0.303	0.247	0.162	0.169	0.108	2705
(0.015)	r_t^2	0.218	0.234	0.173	0.140	0.193	0.107	0.045	0.061	0.027	2598
NIKKEI	r_t	0.094	−0.072	0.002	0.003	0.006	0.023	0.018	0.002	0.010	2
5,594	$\|r_t\|$	0.362	0.338	0.291	0.290	0.285	0.220	0.103	0.077	0.002	380
(0.027)	r_t^2	0.247	0.117	0.112	0.104	0.110	0.053	0.025	0.011	−0.003	231
DM/US$	r_t	0.037	−0.017	0.024	−0.012	0.028	0.015	0.016	0.013	0.017	2
5,311	$\|r_t\|^{1/4}$	0.268	0.259	0.270	0.231	0.249	0.250	0.130	0.083	0.014	489
(0.027)	$\|r_t\|$	0.229	0.206	0.202	0.179	0.203	0.178	0.092	0.042	−0.030	318
	r_t^2	0.191	0.152	0.074	0.071	0.088	0.059	0.077	0.000	−0.025	61
CHEVRON	r_t	0.087	−0.040	−0.014	−0.034	−0.009	−0.000	−0.004	−0.006	−0.007	2
7,420	$\|r_t\|$	0.174	0.147	0.132	0.139	0.139	0.124	0.074	0.081	0.023	572
(0.023)	r_t^2	0.165	0.089	0.074	0.115	0.182	0.085	0.028	0.043	0.000	385
AJINOMOTO	r_t	−0.440	0.127	−0.060	0.018	−0.006	−0.006	−0.004	0.005	−0.000	1
25,099	$\|r_t\|$	0.280	0.224	0.209	0.192	0.181	0.163	0.137	0.130	0.104	3862
(0.013)	r_t^2	0.028	0.022	0.021	0.020	0.017	0.013	0.012	0.013	0.005	1465

other stock returns, has been referred to as the Taylor effect by Granger and Ding (1995). Their research gives strong evidence that the absolute returns have a conditional exponential distribution and pairs of absolute returns have a joint exponential distribution. The result is more significant if the outliers outside the four standard deviations are reduced to four standard deviations. They also showed that the Taylor effect is a consequence of this distribution.

3. The autocorrelation functions of ARCH and GARCH processes

The stylized empirical regularities discussed in Section 2 pose a challenging task for econometricians to develop models to describe the phenomena. Although the Taylor effect has been considered by Granger and Ding (1995), the problem of modeling the long memory behavior of the absolute returns or the squared returns remains largely unsolved. Engle (1982) proposed the ARCH model to forecast the variance conditional upon past information. The conditional variance is a linear function of past squared residuals. The ARCH(p) process is defined as follows:

$$\varepsilon_t = \sigma_t e_t, \qquad e_t \sim \text{i.i.d. } \mathrm{D}(0,1),$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2. \tag{3.1}$$

The ARCH model has the property that the residuals themselves, here ε_t, are not correlated with each other since e_t is i.i.d. over time and is independent of σ_t. But the squared residuals, here ε_t^2, are correlated with each other over time and are forecastable. It follows that $|\varepsilon_t|^d$ or $|r_t|^d$, where d is a positive number, are also correlated with each other and are largely forecastable. So the ARCH model captures some of the empirical facts found in Section 2.

As illustrated by the plots in Section 2, the absolute returns or squared returns are significantly correlated over long lags. This suggests that the data might require a lot of ARCH lags in order to be fully described. Bollerslev (1986) generalizes the ARCH model by adding lagged σ_t in the conditional variance equation. So the GARCH model is as follows:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2. \tag{3.2}$$

The way the GARCH term is introduced, and the reason the GARCH model is so popular are largely due to its convenience in implementation. One can find its counterpart in Box–Jenkins' ARMA technique in modeling the mean.

Taylor (1986) gives the autocorrelation functions for ε_t^2 of an ARCH(p) process, and shows that it follows the same Yule–Walker equation for the associated zero-mean AR(p) process provided the fourth moment of ε_t exists,

$$\rho_k = \sum_{i=1}^{p} \alpha_i \rho_{k-i}. \tag{3.3}$$

Bollerslev (1986) derives the autocorrelation functions for ε_t^2 of a GARCH(p,q) process under the same condition as above and shows

$$\rho_k = \sum_{i=1}^{m} \phi_i \rho_{k-i}, \quad k \geqslant q+1, \tag{3.4}$$

where $m = \max(p,q)$, and $\phi_i = \alpha_i + \beta_i$ for $i = 1,\ldots,m$, $\alpha_i = 0$ for $i > p$, and $\beta_i = 0$ for $i > q$. The above formula is again the same as the Yule–Walker equation for a zero-mean AR(m) process. Unfortunately, no explicit result is available for GARCH(p,q) model when $k \leqslant q$.

In empirical research, the most often used GARCH model is GARCH(1, 1). It will be interesting to see the theoretical autocorrelation functions for GARCH(1, 1) process. For ease of exposition, it will be assumed that the distribution is conditional normal and ε_t is covariance-stationary so that $\alpha + \beta < 1$. These assumptions will be removed later, and it will be shown that similar results still hold.Under the above assumptions the GARCH(1, 1) process can be represented in its asymptotic form as follows:

$$\sigma_t^2 = \sigma^2(1-\alpha-\beta) + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2, \tag{3.5}$$

where σ^2 is the unconditional variance of ε_t. Rearranging the above equation one gets

$$\varepsilon_t^2 - \sigma^2 = (\alpha+\beta)(\varepsilon_{t-1}^2 - \sigma^2) - \beta\sigma_{t-1}^2(e_{t-1}^2 - 1) + \sigma_t^2(e_t^2 - 1). \tag{3.6}$$

Multiplying both sides of the above equation by $(\varepsilon_{t-1}^2 - \sigma^2)$ and taking expectation one has

$$\gamma_1 = (\alpha+\beta)\gamma_0 - 2\beta \mathrm{E}\sigma_{t-1}^4, \tag{3.7}$$

where $\gamma_1 = \mathrm{E}(\varepsilon_t^2 - \sigma^2)(\varepsilon_{t-1}^2 - \sigma^2)$ is the covariance between ε_t^2 and ε_{t-1}^2, and $\gamma_0 = \mathrm{E}(\varepsilon_{t-1}^2 - \sigma^2)^2$ is the variance of ε_{t-1}^2.

If it is further assumed that $3\alpha^2 + 2\alpha\beta + \beta^2 < 1$, so that the fourth moment of ε_t exists, then it is shown in Bollerslev (1988) that

$$\rho_1 = \alpha + \frac{\alpha^2\beta}{1 - 2\alpha\beta - \beta^2}. \tag{3.8}$$

Combining this with (3.4), one has the autocorrelation function for GARCH(1, 1) process, when the fourth moment exists, as follows:

$$\rho_k = \left(\alpha + \frac{\alpha^2\beta}{1 - 2\alpha\beta - \beta^2}\right)(\alpha+\beta)^{k-1}. \tag{3.9}$$

For the convenience of later reference, a detailed derivation of the above formula is given in the Appendix. It should be noted that $\rho_k = 0$ when $\alpha = 0$, which is not surprising. When $\alpha = 0$, the conditional variance process is a deterministically time-varying process independent of ε_t, so $\rho_k = 0$ is what one should have expected. In fact, much of the above discussions is valid even if the fourth moment does not exist or is infinite. Assume now $\alpha + \beta < 1$, but $3\alpha^2 + 2\alpha\beta + \beta^2 \geqslant 1$, so ε_t is covariance-stationary, but its fourth moment does not exist or is infinite. Under this assumption, Eq. (3.7) still holds, but γ_1, γ_0, and $E\sigma_{t-1}^4$ are time-dependent and will go to infinity. However, the sample autocorrelation functions of any observed time series are always defined. By using Jensen's inequality, it is shown in the Appendix that if the process starts a very long time ago, then the first-order autocorrelation function is approximately as follows:

$$\rho_1 \approx \alpha + \tfrac{1}{3}\beta, \tag{3.10}$$

and the autocorrelation function at lag k is approximately

$$\rho_k \approx (\alpha + \tfrac{1}{3}\beta)(\alpha + \beta)^{k-1}. \tag{3.11}$$

From this it is clear that the autocorrelation function still decreases exponentially. It should be noted that this expression is identical to (3.9) when $3\alpha^2 + 2\alpha\beta + \beta^2 = 1$, i.e., the autocorrelation function changes continuously as long as the GARCH(1, 1) model is covariance-stationary.

It is also commonly found in empirical research that the sum of the estimated ARCH and GARCH parameters in a GARCH(1, 1) model is very close to one. For example, Taylor (1986) estimated GARCH(1, 1) models for 40 different financial time series. The results show that for all but six of the 40 series the estimated value of $\alpha + \beta$ is greater than or equal to 0.97. In Ding, Granger, and Engle (1993), the estimated value of $\alpha + \beta$ for daily S&P 500 returns is 0.997. This regularity is widely considered to be a characteristic of volatility persistence. The Integrated GARCH(1, 1) model, which restricts $\alpha + \beta = 1$, is then introduced by Engle and Bollerslev (1986) to model long-run volatility persistence. The IGARCH(1, 1) model is always related to the random walk process in mean. However, Nelson (1990) shows that IGARCH(1, 1) process without drift is strictly stationary and goes to zero almost surely even though it is not covariance-stationary.

More interestingly, using the similar procedure as before, it is shown in the Appendix that the approximate autocorrelation function for ε_t^2 from an IGARCH(1, 1)-type of model is as follows:

$$\rho_k = \tfrac{1}{3}(1 + 2\alpha)(1 + 2\alpha^2)^{-k/2}, \qquad \alpha \neq 0. \tag{3.12}$$

The result is the same no matter whether there is a positive drift or not. It should be noted that the autocorrelation function no longer changes continuously when the parameters change from $\alpha + \beta < 1$ to $\alpha + \beta = 1$ because of the 'jump' in the statistical properties of the model. Formulas (3.9) and (3.11) should not be used to conclude that the autocorrelation function is a constant when $\alpha + \beta = 1$. It is surprising that the autocorrelation function for the Integrated GARCH(1, 1) process is also exponentially decreasing. This is another major difference between the IGARCH(1, 1) model in variance and the random walk process in mean. The IGARCH(1, 1) process is 'not persistent' in volatility at all in the sense that the autocorrelation function for ε_t^2 dies out exponentially. The reason for this is probably due to the fact that the effect of a lagged squared error to the actual conditional variance is exponentially decreasing. Since

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 = \frac{\omega}{1-\beta} + \alpha \sum_{k=1}^{\infty} \beta^{k-1}\varepsilon_{t-k}^2, \tag{3.13}$$

that is, in σ_t^2 equation, the parameter for ε_{t-k}^2 is exponentially decreasing with k as far as $\beta < 1$.

The result here seems to be quite *counterintuitive*. However, as discussed in Nelson (1990), for the GARCH(1, 1) model, whether shocks to conditional variance persist or not depends crucially on the definition of persistence. Researchers usually define volatility persistent by looking at the effect of a shock to future expectation of the variance process. While the effect of a shock to both the expectation and the true process is the same for a random walk process in mean, it is quite different for the IGARCH(1, 1) process in variance. The following example shows that a shock may permanently affect the 'expectation' of a future conditional variance process, but it does not permanently affect the 'true' conditional variance process itself.

Let ε_t be generated from an IARCH(1) process as follows:

$$\varepsilon_t = \sigma_t e_t, \quad e_t \sim \mathrm{N}(0,1), \quad \sigma_t^2 = \varepsilon_{t-1}^2. \tag{3.14}$$

From this IARCH(1) process, the shock to the system at time t comes only from e_t, and this shock will not affect σ_t^2 because σ_t^2 depends only on the past information. Since $\mathrm{E}_t(\varepsilon_{t+k}^2) = \mathrm{E}_t(\sigma_{t+k}^2) = \varepsilon_t^2$, a shock at time t to ε_t^2, which comes from a shock to e_t, will *permanently* change $\mathrm{E}_t(\varepsilon_{t+k}^2)$ and $\mathrm{E}_t(\sigma_{t+k}^2)$, i.e., the 'expectation' of the future squared process and the future conditional variance process. However, the 'true' ε_{t+k}^2 and σ_{t+k}^2 are determined by the following formulas:

$$\sigma_{t+k}^2 = \varepsilon_{t+k-1}^2 = e_{t+k-1}^2 e_{t+k-2}^2 \cdots e_{t+1}^2 \varepsilon_t^2, \quad \varepsilon_{t+k}^2 = e_{t+k}^2 e_{t+k-1}^2 \cdots e_{t+1}^2 \varepsilon_t^2.$$

From here it is clear that the *real* impact of a shock (a change in ε_t^2) to σ_{t+k}^2 and ε_{t+k}^2 are as follows:

$$\partial\sigma_{t+k}^2/\partial\varepsilon_t^2 = \varepsilon_{t+k-1}^2 = e_{t+k-1}^2 e_{t+k-2}^2 \cdots e_{t+1}^2,$$

$$\partial\varepsilon_{t+k}^2/\partial\varepsilon_t^2 = e_{t+k}^2 e_{t+k-1}^2 \cdots e_{t+1}^2,$$

i.e., they are stochastic numbers rather than constants. As shown in Nelson (1990), the above quantities will go to zero almost surely as $k \to \infty$, i.e., the *real* impact of a shock will converge to zero almost surely. This is different from that of a unit root process in mean. For a unit root process in mean, a shock at time t will *permanently* affect both the 'expectation' of the process and the 'true' process itself.

The above discussion can be generalized to the situation where the distribution is not conditional normal, and the conditional variance equation is not a linear function in lagged squared residuals. For example, if the conditional heteroskedasticity equation is Ding, Granger, and Engle's Power-ARCH as follows:

$$\varepsilon_t = \sigma_t e_t, \quad \sigma_t^\delta = \omega^\delta(1-\alpha-\beta) + \alpha|\varepsilon_{t-1}|^\delta + \beta\sigma_{t-1}^\delta, \tag{3.15}$$

where e_t is i.i.d. and, for ease of exposition, it will be assumed that e_t has mean zero, variance σ_e^2, and $\mathrm{E}|e_t|^\delta = 1$, $\mathrm{E}|e_t|^{2\delta} = \zeta$. Then, by using the same steps as before, one can show:

$$\begin{aligned} \rho_1 &= \operatorname{corr}(|\varepsilon_t|^\delta, |\varepsilon_{t-1}|^\delta) \\ &= \alpha + \frac{\beta}{\zeta} - \beta\frac{\zeta-1}{\zeta}\left[\frac{\zeta(1-\alpha-\beta)(1+\alpha+\beta)}{1-(\zeta\alpha^2+\beta^2+2\alpha\beta)} - 1\right]^{-1}, \end{aligned} \tag{3.16}$$

$$\rho_k = \rho_1(\alpha+\beta)^{k-1}, \tag{3.17}$$

provided $\mathrm{E}|\varepsilon_t|^{2\delta}$ exists. Formula (3.16) becomes (3.9) when $\zeta = 3$ which is the normal case. Similar result still holds even when $\mathrm{E}|\varepsilon_t|^{2\delta}$ goes to infinity. Simple algebra shows, when $\alpha + \beta < 1$ but $\zeta\alpha^2 + \beta^2 + 2\alpha\beta > 1$, that

$$\rho_k \approx \left[\alpha + \left(1 - \frac{2}{\zeta}\right)\beta\right](\alpha+\beta)^{k-1}. \tag{3.18}$$

When $\alpha + \beta = 1$ and $\alpha > 0$, so the model is IGARCH(1, 1) in $|\varepsilon_t|^\delta$, one has

$$\rho_k \approx \frac{1+(\zeta-1)\alpha}{\zeta}[1+(\zeta-1)\alpha^2]^{-k/2}. \tag{3.19}$$

However, if $\alpha + \beta = 1$ but $\alpha = 0$ so that the conditional variance process is either a constant or a deterministic step function, then the autocorrelation function is zero. So it is proved that the conclusion in this section is valid under a wide class of models and distributions.

Figs. 6–9 plot the simulated sample autocorrelation functions for various GARCH(1, 1) processes and their corresponding theoretical autocorrelation functions. c is the constant in variance equation, a is the ARCH parameter, and b is the GARCH parameter. A total of 35,000 observations are generated with a conditional normal distribution. The first 15,000 observations are deleted when calculating the sample autocorrelation function in order to be less affected

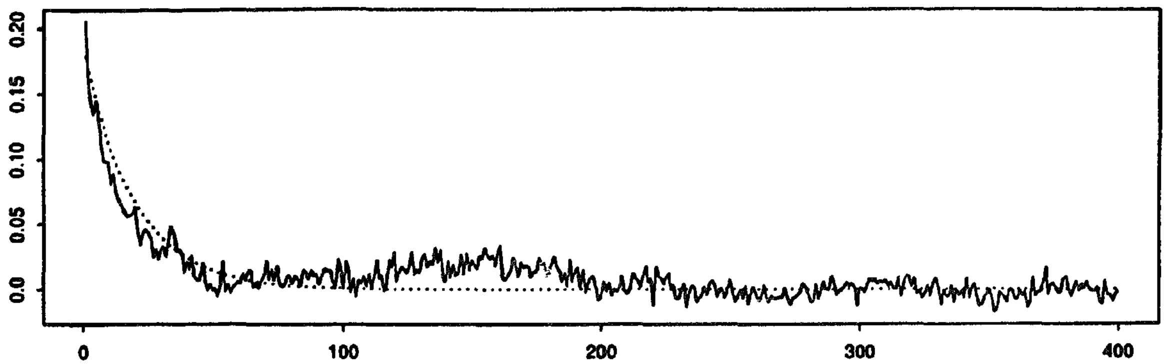

Fig. 6. Theoretical and sample autocorrelation functions for squared GARCH(1, 1) processes: $c = 0.0005$, $a = 0.1$, $b = 0.85$, covariance-stationary, fourth moment exists.

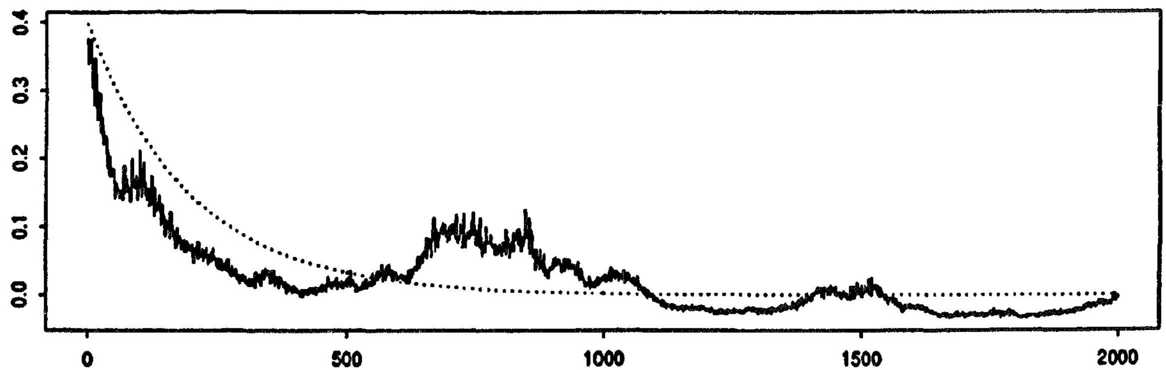

Fig. 7. Theoretical and sample autocorrelation functions for squared GARCH(1,1) processes: $c = 0.00005$, $a = 0.1$, $b = 0.895$, covariance-stationary, fourth moment does not exist.

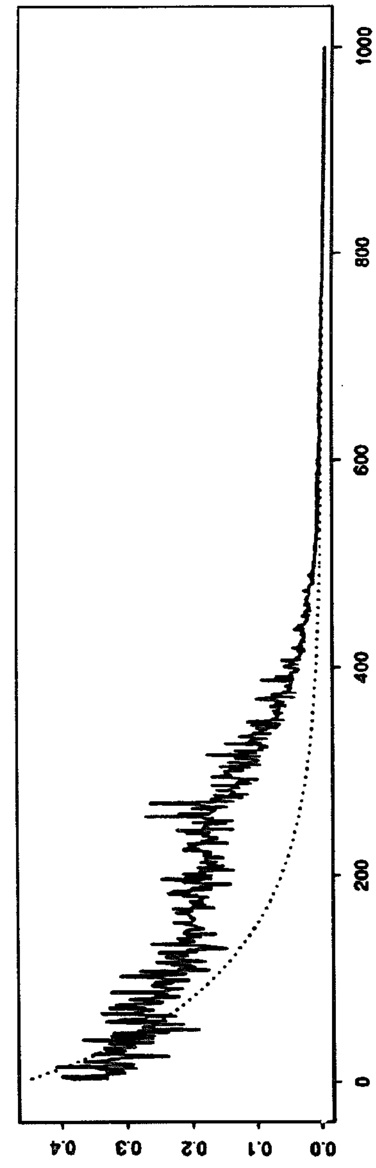

Fig. 8. Theoretical and sample autocorrelation functions for squared GARCH(1,1) processes: $c = 0$, $a = 0.1$, $b = 0.9$, IGARCH(1,1).

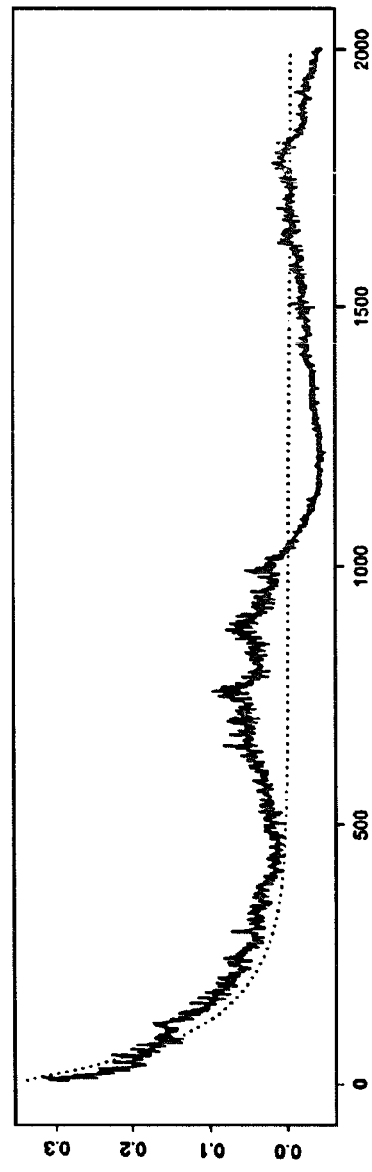

Fig. 9. Theoretical and sample autocorrelation functions for squared GARCH(1,1) processes: $c = 0.000001$, $a = 0.1$, $b = 0.9$, IGARCH(1,1) with drift.

by the initial value, so only 20,000 observations are actually used. The calculated autocorrelations are sample adjusted. The simulated model for Fig. 6 is covariance-stationary and the fourth moment exists, so the theoretical autocorrelation function is precisely defined. It is seen that the sample autocorrelations are very close to the theoretical ones. As predicted by the theory, it decreases exponentially fast. The first negative sample autocorrelation occurs at lag 47 and the sample autocorrelations after this lag are not significantly different from zero. The simulated model for Fig. 7 is also covariance-stationary, but the fourth moment does not exist, so the theoretical autocorrelation function is an approximate one. It is still seen that the theoretical autocorrelation function fits the sample quite well. The sample autocorrelation decreases too fast to account for the long memory property found in the real data. Fig. 8 plots the simulated sample autocorrelations and their theoretical approximations for IGARCH(1, 1) process. One can see that the pattern of the sample autocorrelation function is very different from the previous ones. It decreases like a straight line for about the first 400 lags and then collapses to zero. It is thus seen that the IGARCH(1, 1) process without drift is not persistent in volatility at all. Fig. 9 shows the sample and theoretical autocorrelations for IGARCH(1, 1) process with drift. The shape of the simulated sample autocorrelations is quite similar to those covariance-stationary GARCH(1, 1) processes as shown in Figs. 6 and 7. The theoretical exponentially decreasing autocorrelation function provides a very good approximation. It should be emphasized that the simulations are performed for many different parameter choices, and similar results are found for almost all of them.

4. Modeling volatility persistence: A new approach

From the discussion above, it is seen that the pattern of the sample autocorrelation for various speculative returns discussed in Section 2 is quite different from that of the theoretical autocorrelation functions given by a GARCH(1, 1) or IGARCH(1, 1) process. Usually the real data has a longer memory in volatility than the GARCH(1, 1) model suggested. The autocorrelation for ε_t^2 from a GARCH(1, 1) process decreases exponentially, while the sample autocorrelation usually decreases much faster than exponentially at the beginning and then decreases much slower and remains significantly positive over long lags. Usually a GARCH(1, 1) process can describe the short-run effect better than the long-run effect. It is quite clear from the sample autocorrelation that there are different volatility components that will dominate different time period. Some volatility components may have a very big short-run effect, but die out very quickly. Some of them may have a relatively smaller short-run effect, but they last for a long time period.

For example, consider the following two-component model:

$$\varepsilon_t = \sigma_t e_t, \qquad e_t \sim \text{i.i.d. } D(0,1), \tag{4.1}$$

$$\sigma_t^2 = w\sigma_{1t}^2 + (1-w)\sigma_{2t}^2, \tag{4.2}$$

$$\sigma_{1t}^2 = \alpha_1 \varepsilon_{t-1}^2 + (1-\alpha_1)\sigma_{1t-1}^2, \tag{4.3}$$

$$\sigma_{2t}^2 = \sigma^2(1-\alpha_2-\beta_2) + \alpha_2\varepsilon_{t-1}^2 + \beta_2\sigma_{2t-1}^2. \tag{4.4}$$

So σ_t^2 is a weighted sum of two components, σ_{1t}^2 and σ_{2t}^2, with w and $1-w$ as weights respectively. σ_{1t}^2 is an IGARCH(1,1)-type specification and σ_{2t}^2 is a GARCH(1,1)-type specification. It should be noted that σ_t^2 is not an Integrated GARCH process as will be shown below. Expanding the two variance components σ_{1t}^2 and σ_{2t}^2, one has

$$\sigma_{1t}^2 = \alpha_1 \sum_{k=1}^{\infty} (1-\alpha_1)^{k-1}\varepsilon_{t-k}^2, \tag{4.5}$$

$$\sigma_{2t}^2 = \sigma^2 \frac{1-\alpha_2-\beta_2}{1-\beta_2} + \alpha_2 \sum_{k=1}^{\infty} \beta_2^{k-1}\varepsilon_{t-k}^2. \tag{4.6}$$

So

$$\begin{aligned}\sigma_t^2 &= \sigma^2(1-w)\frac{1-\alpha_2-\beta_2}{1-\beta_2} \\ &\quad + \sum_{k=1}^{\infty} [w\alpha_1(1-\alpha_1)^{k-1} + (1-w)\alpha_2\beta_2^{k-1}]\varepsilon_{t-k}^2. \end{aligned} \tag{4.7}$$

It is then clear that, when $w\alpha_1(1-\alpha_1)^{k-1} > (1-w)\alpha_2\beta_2^{k-1}$, the first variance component will have a bigger effect than the second one, and reversing the inequality reverses this result. If one substitutes σ_{1t}^2, σ_{2t}^2 into σ_t^2, it is readily seen that ε_t follows a GARCH(2,2) process and the conditional variance equation becomes

$$\begin{aligned}\sigma_t^2 &= \sigma^2(1-w)\alpha_1(1-\alpha_2-\beta_2) + [w\alpha_1 + (1-w)\alpha_2]\varepsilon_{t-1}^2 \\ &\quad - [w\alpha_1\beta_2 + (1-w)(1-\alpha_1)\alpha_2]\varepsilon_{t-2}^2 \\ &\quad + (1-\alpha_1+\beta_2)\sigma_{t-1}^2 - (1-\alpha_1)\beta_2\sigma_{t-2}^2. \end{aligned} \tag{4.8}$$

Interestingly, although the ARCH(2) and GARCH(2) parameters are negative in the above equation, the conditional variance process is still guaranteed to be positive provided the parameters in the components representation are positive. The sum of the ARCH and GARCH parameters in the above equation is

$$\begin{aligned}\Sigma &= w\alpha_1 + (1-w)\alpha_2 - w\alpha_1\beta_2 - (1-w)(1-\alpha_1)\alpha_2 \\ &\quad + 1 - \alpha_1 + \beta_2 - (1-\alpha_1)\beta_2 \\ &= 1 - (1-w)\alpha_1(1-\alpha_2-\beta_2), \end{aligned}$$

which is bigger than zero and less than one when $0 < w < 1$, $0 < \alpha_1 < 1$, and $0 < \alpha_2 + \beta_2 < 1$. Under this assumption, the process is covariance-stationary with

$$\mathrm{E}\sigma_t^2 = \mathrm{E}\sigma_{1t}^2 = \mathrm{E}\sigma_{2t}^2 = \mathrm{E}\varepsilon_t^2 = \sigma^2. \tag{4.9}$$

The theoretical autocorrelation function for this two-component model is not available. However, by Eq. (3.4), an exponentially decreasing type autocorrelation function should be expected for this model.

If one fits this two-component model for S&P 500 daily returns with a conditional normal distribution, then the estimated result is as follows:

$$\underset{}{r_t} = \underset{(7.4)}{0.000463} + \underset{(18.2)}{0.148}\varepsilon_{t-1} + \varepsilon_t, \quad \varepsilon_t = \sigma_t e_t, \quad e_t \sim \mathrm{N}(0,1), \tag{4.10}$$

$$\sigma_t^2 = \underset{(39.3)}{0.704}\sigma_{1t}^2 + 0.296\sigma_{2t}^2, \tag{4.11}$$

$$\sigma_{1t}^2 = \underset{(29.0)}{0.153}\varepsilon_{t-1}^2 + 0.847\sigma_{1t-1}^2, \tag{4.12}$$

$$\sigma_{2t}^2 = \underset{(7.6)}{0.162 \times 10^{-6}} + \underset{(12.3)}{0.008}\varepsilon_{t-1}^2 + \underset{(16.84)}{0.991}\sigma_{2t-1}^2, \tag{4.13}$$

Log-likelihood: 56,894.

This model gives a significantly higher likelihood function than a GARCH(1, 1) model (see Ding, Granger, and Engle, 1993). For the estimation result here, the volatility-persistent part comes from the GARCH(1, 1)-type component instead of the IGARCH(1, 1)-type component. In the conditional variance equation, σ_{1t}^2 starts very big $(0.704 \times 0.153 = 0.1077)$ in magnitude, but decays very fast with a decay rate of 0.847, while σ_{2t}^2 starts very small $(0.296 \times 0.008 = 0.0024)$ in magnitude, but decays very slowly with a decay rate of 0.991. So σ_{1t}^2 captures the short-term fluctuation, while σ_{2t}^2 models the long-term, gradual movements in volatility. For the specific parameters estimated here, $w\alpha_1(1-\alpha_1)^{k-1} > (1-w)\alpha_2\beta_2^{k-1}$ when $k \leqslant 25$. So the short-run volatility fluctuation will die out within 25 days or about one month, and the long-run mean-reversion volatility component will dominate thereafter.

Fig. 10 plots the simulated sample autocorrelations for the GARCH(1, 1) model estimated in Ding, Granger, and Engle (1993). Fig. 11 plots the simulated sample autocorrelations for the two-component model estimated above. Clearly, the two-component model gives a much closer sample autocorrelations than GARCH(1, 1), compared to the real one. The two-component model is also more persistent in the sense that it has more lags of positive sample autocorrelations.

The intuition behind this two-component model is that one can use two different variance components, each of them has an exponentially decreasing

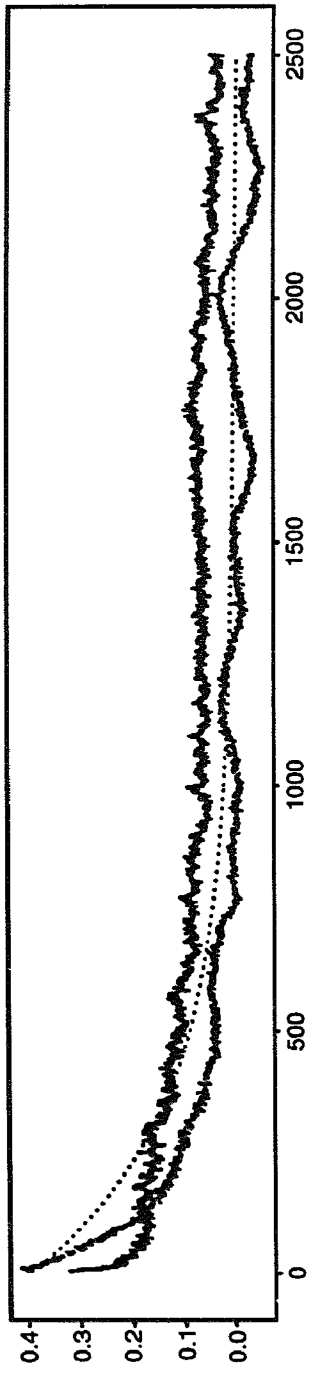

Fig. 10. Sample autocorrelations (solid line) of $|r|$ for S&P 500 and their simulated values: GARCH(1, 1) model, $c = 0.0000008$, $\alpha = 0.091$, $\beta = 0.906$.

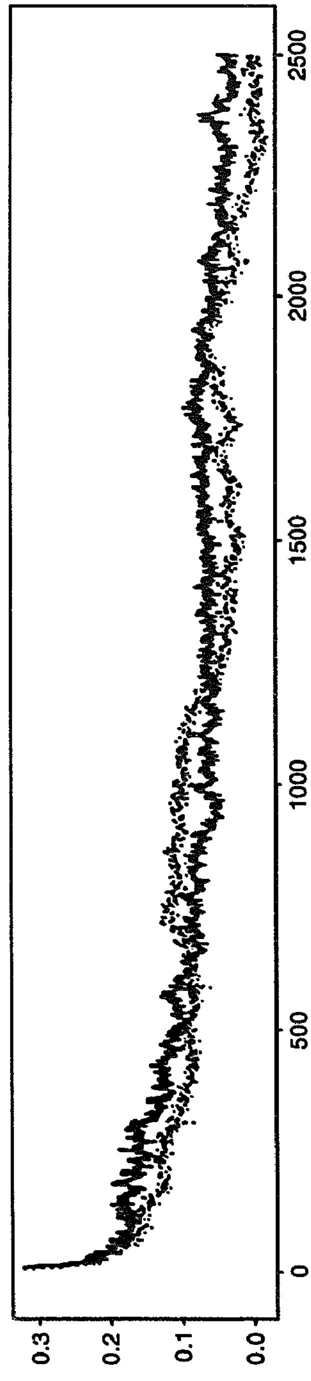

Fig. 11. Sample autocorrelations (solid line) of $|r|$ for S&P 500 and their simulated values: Two-component model.

autocorrelation pattern, to model the long-term and short-term movements in volatility. This can also be seen from the two different slopes in sample autocorrelations. However, it is also seen that the two-component model is still not enough to account for the whole autocorrelation pattern, as it might be reasonable to think that there are many more components in volatility. To generalize this idea, an N-component model is defined as follows:

$$\varepsilon_t = \sigma_t e_t, \quad e_t \sim \text{i.i.d. } \mathrm{D}(0,1), \tag{4.15}$$

$$\sigma_t^2 = \sum_{i=1}^{N} w_i \sigma_{it}^2, \quad \sum_{i=1}^{N} w_i = 1, \tag{4.16}$$

$$\sigma_{it}^2 = \sigma^2(1 - \alpha_i - \beta_i) + \alpha_i \varepsilon_{t-1}^2 + \beta_i \sigma_{it-1}^2, \qquad i = 1, 2, \ldots, N, \tag{4.17}$$

where w_i is the weight for volatility component i. Since this model is a very general one and encompasses the models discussed above as special cases, one would expect it to give a better approximation to the real data. The sample autocorrelation, which consists of N different sloped exponentially decreasing components, will essentially envelope the real one. Similar to the two-component model, the N-component model corresponds to a GARCH(N,N) model. However, when the number of components N is too large, it will be difficult to estimate α_i and β_i for such a model. On the other hand, one does not have any prior knowledge how many variance components there really are. Our interest is at the limiting case when N goes to infinity, and α_i and β_i can take any value in some region, so the component model will correspond to no GARCH process having finite number of parameters.

Under this situation it is necessary to assume some distributional forms for α and β in order to go further. One very general distribution defined on the range $[0,1]$, which can be transformed to any other range $[a,b]$, is the *Beta* distribution. The density function for a *Beta* distribution with parameters p,q is as follows:

$$f(x) = \frac{x^{p-1}(1-x)^{q-1}}{B(p,q)}, \qquad 0 \leqslant x \leqslant 1, \tag{4.18}$$

where $B(p,q)$ is the *Beta* function defined as follows:

$$B(p,q) = \int_0^1 x^{p-1}(1-x)^{q-1}\,\mathrm{d}x. \tag{4.19}$$

Beta distribution can have various shapes depending on the parameters p and q. It includes many other distributions, such as uniform, as special cases. Johnson and Kotz (1970) give a detailed discussion for this family of distributions. In the following analysis it will be assumed that β has a form of *Beta* distribution on the range $[0,1]$ with parameters p and q. It will be argued later that the exact form selected for this distribution function is not of critical importance except near $\beta = 1$. It will also be assumed that $\alpha = (1-\beta)\alpha^*$, where α^* is from any distribution defined on the range $[0,1]$ with a mean $\mu > 0$. It then follows that $0 \leqslant \alpha + \beta \leqslant 1$ as required. For mathematical convenience it will be further assumed that α^* is independent of β even though it can be seen later that this is not necessary. Since Eq. (4.17) can be rewritten in the following form:

$$\sigma_{it}^2 = \sigma^2 \frac{1-\alpha_i-\beta_i}{1-\beta_i} + \frac{\alpha_i}{1-\beta_i L}\varepsilon_{t-1}^2, \qquad i = 1,2,\ldots,N,$$

substituting this into (4.16) one gets

$$\sigma_t^2 = \sum_{i=1}^{N} w_i \left[\sigma^2 \frac{1-\alpha_i-\beta_i}{1-\beta_i} + \frac{\alpha_i}{1-\beta_i L}\varepsilon_{t-1}^2\right], \tag{4.20}$$

where L is the lag operator. When $N \to \infty$ and under the distributional assumptions for α and β given above, one has

$$\begin{aligned}
\sigma_t^2 &= \sigma^2 \int_0^1 \frac{1-\alpha-\beta}{1-\beta}\,\mathrm{d}F(\alpha,\beta) + \int_0^1 \frac{\alpha}{1-\beta L}\varepsilon_{t-1}^2\,\mathrm{d}F(\alpha,\beta) \\
&= \sigma^2 \int_0^1 (1-\alpha^*)\,\mathrm{d}F(\alpha^*) + \sum_{k=1}^{\infty} \varepsilon_{t-k}^2 \int_0^1 \alpha^*(1-\beta)\beta^{k-1}\,\mathrm{d}F(\alpha^*,\beta) \\
&= \sigma^2(1-\mu) + \sum_{k=1}^{\infty} \varepsilon_{t-k}^2 \int_0^1 \alpha^*\,\mathrm{d}F(\alpha^*) \int_0^1 (1-\beta)\beta^{k-1}\,\mathrm{d}F(\beta) \\
&= \sigma^2(1-\mu) + \mu \sum_{k=1}^{\infty} \frac{B(p+k-1,q+1)}{B(p,q)}\varepsilon_{t-k}^2.
\end{aligned} \tag{4.21}$$

The conditional variance process decided by the above formula has the long memory property since, when k large,

$$\begin{aligned}
a_k = \frac{B(p+k-1,q+1)}{B(p,q)} &= \frac{q\Gamma(p+q)\Gamma(p+k-1)}{\Gamma(p)\Gamma(p+q+k)} \\
&\approx \frac{q\Gamma(p+q)}{\Gamma(q)} k^{-1-q},
\end{aligned} \tag{4.22}$$

which is a characteristic of a long memory process (see Granger, 1980). We will refer to this general class of models as Long Memory ARCH model of order q and denote as LM(q)-ARCH model. It should be noted that independent research by Baillie, Bollerslev, and Mikkelsen (1996) leads to a closely related model, which they call Fractionally Integrated GARCH, or FIGARCH model. The result here appears to provide an interesting argument for how such volatility processes might arise.

It is seen from Eq. (4.22) that except for at the extreme point of $p = 0$, the decay pattern of a_k is solely decided by q and so the discussion in Granger (1980) applies here. According to Granger (1980), the shape of the *Beta* distribution is of little relevance in getting the long memory property of the model except near $\beta = 1$, where q determines the slope of the *Beta* density function in the form chosen. If the upper limit of the range is changed from one to b, where b is some quantity strictly less than one, then a_k will be dominated by exponentially decreasing component again and the model will lose the long memory property. So, as far as a proportion of β is taken from regions arbitrarily close to one, the model will have the long memory property desired. It is thus seen that the most important parameter for getting volatility persistence is β instead of $\alpha + \beta$. Important difference also occurs when p takes the value of zero. When $p = 0$, the distribution for β collapses to the lower limit which is zero, so each component follows an ARCH(1) process instead of a GARCH(1, 1) process. The aggregated conditional variance process (4.21) becomes

$$\sigma_t^2 = \sigma^2(1 - \mu) + \mu\varepsilon_{t-1}^2,$$

which, as discussed in Section 3, does not have long memory property even if $\mu = 1$.

The coefficient for ε_{t-k}^2 from the Long Memory ARCH model decays hyperbolically which is much slower than that of any GARCH models having finite lags of ARCH and GARCH terms. Table 3 gives some numerical values for coefficients at different lags of these two classes of models. The first five columns are coefficients at lag k for long memory models with different parameters p and

Table 3
Coefficients at lag k (a_k) for long memory and GARCH(1, 1) models

Lag k	$p = 1$, $q = 0.1$	$p = 2$, $q = 0.2$	$p = 3$, $q = 0.3$	$p = 4$, $q = 0.4$	$p = 5$, $q = 0.5$	$p = 0.75$, $q = 0.25$	$a = 0.1$, $b = 0.9$
1	0.09091	0.09091	0.09091	0.09091	0.09091	0.25000	0.10000
2	0.04329	0.05682	0.06342	0.06734	0.06993	0.09375	0.09000
3	0.02793	0.04058	0.04787	0.05261	0.05594	0.05469	0.08100
4	0.02044	0.03122	0.03799	0.04266	0.04607	0.03760	0.07290
5	0.01603	0.02518	0.03122	003555	0.03880	0.02820	0.06561
10	0.00752	0.01227	0.01567	0.01824	0.02028	0.01166	0.03874
15	0.00484	0.00855	0.01191	0.01525	0.01877	0.00691	0.02288
20	0.00353	0.00605	0.00819	0.01019	0.01219	0.00482	0.01351
50	0.00129	0.00202	0.00249	0.00283	0.00308	0.00153	0.00057
100	0.00060	0.00088	0.00101	0.00107	0.00109	0.00065	0.00000
150	0.00038	0.00054	0.00060	0.00061	0.00059	0.00039	0.00000
200	0.00028	0.00038	0.00041	0.00041	0.00039	0.00027	0.00000
500	0.00010	0.00013	0.00012	0.00011	0.00010	0.00009	0.00000
1,000	0.00005	0.00006	0.00005	0.00004	0.00003	0.00004	0.00000
1,500	0.00003	0.00003	0.00003	0.00002	0.00002	0.00002	0.00000
2,000	0.00002	0.00002	0.00002	0.00002	0.00001	0.00002	0.00000

q. They all start at the same coefficient at lag 1, but after that they differ from each other. Column 6 is also a long memory model with $p = 0.75$ and $q = 0.25$, so that $p + q = 1$. The model is of form (4.27) which will be discussed below. The coefficient at lag 1 is very big compared to the first five columns. However, starting from lag 2, the coefficient decreases much faster than that of the first five models. The last column is the coefficients at different lags for an IGARCH (1, 1) model with ARCH parameter 0.1 and GARCH parameter 0.9. Comparing this model with the long memory models, it can be seen that the coefficients for this model decrease to zero much faster. The coefficients after lag 100 are all zero to the fifth decimal point, while all the long memory models still have a coefficient around 0.00002 at lag 2,000.

Since $\sum_{k=1}^{\infty} a_k = 1$, which is the same as an IGARCH process, the Long Memory ARCH model provides a mechanism to distribute more weight on the far past shocks which enables the model to have the long memory property desired. Several special cases arise from (4.21). When $\mu = 1$, the distribution for α will put all its probability mass on its upper limit, i.e., $\alpha_i = 1 - \beta_i$, so each component follows an IGARCH(1, 1) type process and one has

$$\sigma_t^2 = \sum_{k=1}^{\infty} \frac{q\Gamma(p+q)\Gamma(p+k-1)}{\Gamma(p)\Gamma(p+q+k)} \varepsilon_{t-k}^2, \tag{4.23}$$

which also has the property that the coefficients for ε_{t-k}^2 sum to 1. When $p + q = 1$, one has

$$\begin{aligned} \sigma_t^2 &= \sigma^2(1-\mu) + \mu \sum_{k=1}^{\infty} \frac{q\Gamma(k-q)}{\Gamma(1-q)\Gamma(k+1)} \varepsilon_{t-k}^2 \\ &= \sigma^2(1-\mu) + \mu(1-(1-L)^q)\varepsilon_t^2, \end{aligned} \tag{4.24}$$

which is the same as

$$(1-L)^q \varepsilon_t^2 = \left(\frac{1}{\mu} - 1\right)(\sigma^2 - \sigma_t^2) + \sigma_t^2(e_t^2 - 1). \tag{4.25}$$

When $0 < q < \frac{1}{2}$, ε_t^2 is stationary, the mean exists and is equal to σ^2, $\mathrm{E}\varepsilon_t^2 = \mathrm{E}\sigma_t^2 = \sigma^2$. So the right-hand side is a white noise and ε_t^2 is an $\mathrm{I}(q)$ process. If further conditions are imposed so that the fourth moment of ε_t exists, then the theoretical autocorrelation function for ε_t^2 from this model is

$$\rho_k = \mathrm{corr}(\varepsilon_t^2, \varepsilon_{t-k}^2) = \frac{\Gamma(1-q)}{\Gamma(q)} \frac{\Gamma(k+q)}{\Gamma(k+1-q)} \approx \frac{\Gamma(1-q)}{\Gamma(q)} k^{2q-1}, \tag{4.26}$$

which is the same as an $\mathrm{I}(q)$ process in mean. Finally, when $\mu = 1$ in (4.25), one gets

$$(1-L)^q \varepsilon_t^2 = \sigma_t^2(e_t^2 - 1), \tag{4.27}$$

which is the long memory model proposed by Granger and Ding (1995) and is also a special case of the FIGARCH model of Baillie, Bollerslev, and Mikkelsen (1996).

The above discussion in ε_t^2 is solely for convenience. It is not difficult to generalize to other cases. For example, if each volatility component follows Ding, Granger, and Engle's Power–ARCH:

$$\varepsilon_t = \sigma_t e_t, \quad e_t \sim \text{i.i.d. } \mathrm{D}(0,1), \tag{4.28}$$

$$\sigma_t^\delta = \sum_{i=1}^{N} w_i \sigma_{it}^\delta, \quad \sum_{i=1}^{N} w_i = 1, \tag{4.29}$$

$$\sigma_{it}^\delta = \sigma^\delta(1 - \alpha_i - \beta_i) + \alpha_i |\varepsilon_{t-1}|^\delta/\lambda + \beta_i \sigma_{it-1}^\delta, \qquad i = 1,2,\dots,N, \tag{4.30}$$

where $\lambda = \mathrm{E}|e_t|^\delta$ and δ is a positive number to be estimated from the data set. Then, under the same assumption as before, one has

$$\sigma_t^\delta = \sigma^\delta(1 - \mu) + \mu \sum_{k=1}^{\infty} \frac{B(p+k-1, q+1)}{B(p,q)} |\varepsilon_{t-k}|^\delta/\lambda. \tag{4.31}$$

The discussions for σ_t^2 or ε_t^2 above are all applicable to σ_t^δ or $|\varepsilon_t|^\delta$ here. $|\varepsilon_t|^\delta$ will have the long memory property discussed above. When $p + q = 1$, $|\varepsilon_t|^\delta$ will have a theoretical autocorrelation function as in (4.26). If $\delta = 1$ is the true data-generating process, then the absolute returns will also have the properties.

In order to examine the practical relevance of this new class of models, various long memory models with alternative distributions are estimated for S&P 500 daily returns using the estimation by simulation method discussed in Ding (1994). [Interested readers are referred to Ding (1994) for a detailed discussion of simulation and estimation for this general class of Long Memory ARCH models.] The finally preferred model and its estimation result is as follows:

$$\underset{(13.3)}{r_t = 0.000608} + \underset{(16)}{0.103\varepsilon_{t-1}} + \varepsilon_t, \quad \varepsilon_t = \sigma_t e_t, \tag{4.32}$$

$$\sigma_t = \sum_{k=1}^{\infty} \frac{B(p+k-1, q+1)}{B(p,q)} |\varepsilon_{t-k}|/\lambda, \tag{4.33}$$

where e_t has a double exponential distribution and $\lambda = \mathrm{E}|e_t| = \sqrt{\frac{1}{2}}$. The estimated p equals 5.41, with a t-statistic 4.93, and the estimated q equals 0.597, with a t-statistic 8.72. The likelihood function is 57,226, which is much higher than that of the GARCH(1,1) model and the two-component model. It should also be noted that the parameters used in this long memory model is less than the GARCH(1,1) model and the two-component model. Fig. 12 plots the sample autocorrelations for S&P 500 daily returns and simulated absolute returns using the parameters estimated above. A total of 200,000 variance components are generated using β_i from the *Beta* (5.41, 0.597) distribution and $\alpha_i = 1 - \beta_i$. It can

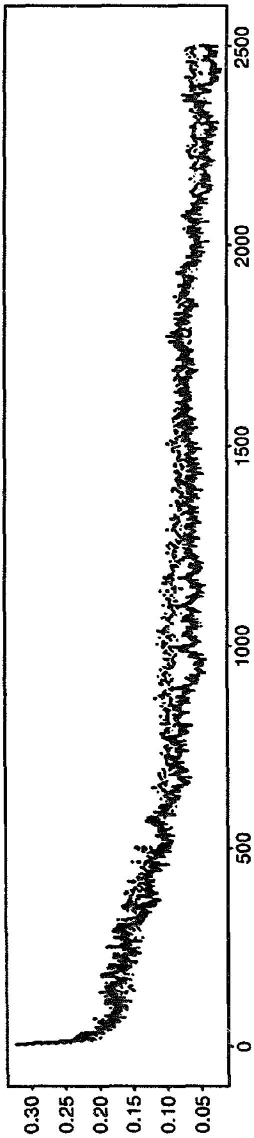

Fig. 12. Sample autocorrelations (solid line) of $|r|$ for S&P 500 and their simulated values: Long memory model, $p = 5.4$, $q = 0.6$.

be seen that for the first 500 lags the two sample autocorrelations move very closely. From lag 500 to lag 2,500, the simulated data has relatively higher sample autocorrelations than the real data. However, compared with any other models, this model still gives the best approximation to the reality.

5. Conclusion

This paper gives new evidence of long-term dependence that exists in speculative returns series. Five speculative returns series from different places and different markets are examined. It is found that the absolute returns and their power transformations all have long, positive autocorrelations. Usually this property is strongest for the absolute returns. One exception is the exchange rate return which has the strongest property when taking to power $\frac{1}{4}$ (this property will be examined in more detail in a later paper). The theoretical autocorrelation functions for various GARCH(1, 1) models are derived and found to be exponentially decreasing, which is rather different from the sample autocorrelation function for the real data. A general class of long memory models that has no memory in returns themselves, but long memory in absolute returns and their power transformations is proposed. The issue of estimation and simulation for this class of models is discussed. The estimated results show that this model gives a much better description for the real data. The Monte Carlo simulation also shows that the theoretical model can mimic the stylized empirical facts strikingly well.

Appendix

We first derive the autocorrelation functions for covariance-stationary GARCH(1, 1) model under conditional normal distribution.

When $\alpha + \beta < 1$, the GARCH(1, 1) process can be represented in its asymptotic form as follows:

$$\sigma_t^2 = \sigma^2(1 - \alpha - \beta) + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2, \tag{A.1}$$

where σ^2 is the unconditional variance of ε_t. Rearranging the above equation one gets

$$\varepsilon_t^2 - \sigma^2 = (\alpha + \beta)(\varepsilon_{t-1}^2 - \sigma^2) - \beta\sigma_{t-1}^2(e_{t-1}^2 - 1) + \sigma_t^2(e_t^2 - 1). \tag{A.2}$$

Multiplying both sides of the above equation by $(\varepsilon_{t-1}^2 - \sigma^2)$ and taking expectation one has

$$\gamma_1 = (\alpha + \beta)\gamma_0 - 2\beta \mathrm{E}\sigma_{t-1}^4, \tag{A.3}$$

where $\gamma_1 = \mathrm{E}(\varepsilon_t^2 - \sigma^2)(\varepsilon_{t-1}^2 - \sigma^2)$ is the covariance between ε_t^2 and ε_{t-1}^2 and $\gamma_0 = \mathrm{E}(\varepsilon_{t-1}^2 - \sigma^2)^2$ is the variance of ε_{t-1}^2.

If it is further assumed $3\alpha^2 + 2\alpha\beta + \beta^2 < 1$, so that the fourth moment of ε_t exists, then dividing both sides of (A.3) by γ_0, which is finite, gives

$$\rho_1 = \alpha + \beta - 2\beta \mathrm{E}\sigma_{t-1}^4/\gamma_0. \tag{A.4}$$

Also by definition

$$\gamma_0 = \mathrm{E}(\varepsilon_t^2 - \sigma^2)^2 = 3\mathrm{E}\sigma_t^4 - \sigma^4, \tag{A.5}$$

so one has

$$\mathrm{E}\sigma_t^4 = (\gamma_0 + \sigma^4)/3. \tag{A.6}$$

Substituting (A.6) into (A.4) gives

$$\rho_1 = \alpha + \beta - 2\beta \frac{\gamma_0 + \sigma^4}{3\gamma_0} = \alpha + \tfrac{1}{3}\beta - \tfrac{2}{3}\beta\sigma^4/\gamma_0. \tag{A.7}$$

From the conditional variance equation one can get

$$\mathrm{E}\sigma_{t-1}^4 = \frac{\sigma^4(1 - \alpha - \beta)(1 + \alpha + \beta)}{1 - (3\alpha^2 + 2\alpha\beta + \beta^2)}. \tag{A.8}$$

Substituting this into (A.5), one has

$$\gamma_0 = 3\mathrm{E}\sigma_t^4 - \sigma^4 = \sigma^4 \frac{2(1 - 2\alpha\beta - \beta^2)}{1 - (3\alpha^2 + 2\alpha\beta + \beta^2)}, \tag{A.9}$$

so that

$$\frac{\sigma^4}{\gamma_0} = \frac{1 - (3\alpha^2 + 2\alpha\beta + \beta^2)}{2(1 - 2\alpha\beta - \beta^2)}. \tag{A.10}$$

Substituting this into (A.7) and some simple algebra shows

$$\rho_1 = \alpha + \frac{\alpha^2\beta}{1 - 2\alpha\beta - \beta^2}. \tag{A.11}$$

Combining this with (3.4), one has the autocorrelation function for GARCH(1,1) process, when the fourth moment exists, as follows:

$$\rho_k = \left(\alpha + \frac{\alpha^2\beta}{1 - 2\alpha\beta - \beta^2}\right)(\alpha + \beta)^{k-1}. \tag{A.12}$$

Much of the above discussions are valid even if the fourth moment does not exist. Assume now $\alpha + \beta < 1$, but $3\alpha^2 + 2\alpha\beta + \beta^2 \geqslant 1$, so ε_t is covariance-stationary, but its fourth moment does not exist. Under this assumption, Eq. (A.3) still holds but γ_1, γ_0, and $\mathrm{E}\sigma_{t-1}^4$ will go to infinity. However, the sample autocorrelation functions are always defined. By Jensen's inequality, one has $\gamma_{0,t-1} > \gamma_{1,t}$ and $\gamma_{0,t-1} > \mathrm{E}\sigma_{t-1}^4$, where $\gamma_{0,t} = \mathrm{E}(\varepsilon_t^2 - \sigma^2)^2$ is the variance of ε_t^2 and $\gamma_{1,t} = \mathrm{E}(\varepsilon_t^2 - \sigma^2)(\varepsilon_{t-1}^2 - \sigma^2)$ is the first-order autocovariance function of ε_t^2. There is a time subscript for the variance and autocovariance function because they are usually changing with time when the fourth moment of ε_t does not exist. If we define the time-varying first-order autocorrelation as usual,

$$\rho_{1,t} = \gamma_{1,t}/\gamma_{0,t-1}, \tag{A.13}$$

then

$$\rho_{1,t} = \alpha + \beta - 2\beta \mathrm{E}\sigma_{t-1}^4/\gamma_{0,t-1}. \tag{A.14}$$

From the conditional variance equation one has

$$\begin{aligned} \mathrm{E}\sigma_t^4 &= \sigma^4(1-\alpha-\beta)(1+\alpha+\beta) + (3\alpha^2+\beta^2+2\alpha\beta)\mathrm{E}\sigma_{t-1}^4 \\ &= \sigma^4(1-\alpha-\beta)(1+\alpha+\beta)[1+(3\alpha^2+\beta^2+2\alpha\beta) \\ &\quad + \cdots + (3\alpha^2+\beta^2+2\alpha\beta)^t \mathrm{E}\sigma_0^4]. \end{aligned} \tag{A.15}$$

Assume, without loss of generality, that $\mathrm{E}\sigma_0^4 = 1$, then

$$\mathrm{E}\sigma_t^4 = \sigma^4(1-\alpha-\beta)(1+\alpha+\beta)\sum_{i=0}^{t}(3\alpha^2+\beta^2+2\alpha\beta)^i. \tag{A.16}$$

So substituting (A.5) and (A.16) into (A.14) gives

$$\begin{aligned} \rho_{1,t} &= \alpha+\beta-2\beta\frac{\sigma^4(1-\alpha-\beta)(1+\alpha+\beta)\sum_{i=0}^{t-1}(3\alpha^2+\beta^2+2\alpha\beta)^i}{3\sigma^4(1-\alpha-\beta)(1+\alpha+\beta)\sum_{i=0}^{t-1}(3\alpha^2+\beta^2+2\alpha\beta)^i-\sigma^4} \\ &= \alpha+\tfrac{1}{3}\beta-\tfrac{2}{3}\beta\left[3(1-\alpha-\beta)(1+\alpha+\beta)\sum_{i=0}^{t-1}(3\alpha^2+\beta^2+2\alpha\beta)^i-1\right]^{-1}, \end{aligned} \tag{A.17}$$

which is changing with time. If it is assumed that the process starts at a very long time ago, so that the last term can be ignored, then

$$\rho_1 \approx \alpha + \tfrac{1}{3}\beta, \tag{A.18}$$

and the autocorrelation function at lag k is approximately

$$\rho_k \approx (\alpha + \tfrac{1}{3}\beta)(\alpha+\beta)^{k-1}. \tag{A.19}$$

It should be noted that this expression is identical to (A.12) when $3\alpha^2 + 2\alpha\beta + \beta^2 = 1$, i.e., the autocorrelation function changes continuously as far as the GARCH(1, 1) model is covariance-stationary.

The situation is quite different when the covariance-stationary assumption is removed. Assume

$$\varepsilon_t = \sigma_t e_t, \quad e_t \sim \text{i.i.d. N}(0,1), \quad \sigma_t^2 = \alpha\varepsilon_{t-1}^2 + (1-\alpha)\sigma_{t-1}^2, \tag{A.20}$$

and $\sigma_0 = 1$, a constant. Then

$$\varepsilon_t^2 = (\alpha\varepsilon_{t-1}^2 + (1-\alpha)\sigma_{t-1}^2)e_t^2, \tag{A.21}$$

and in general it is not difficult to get

$$\begin{aligned}
&\mathrm{E}\sigma_t^2 = 1,\\
&\mathrm{E}\sigma_t^4 = (1+2\alpha^2)^t \mathrm{E}\sigma_0^4 = (1+2\alpha^2)^t,\\
&\mathrm{E}\varepsilon_t^2\varepsilon_{t-k}^2 = (1+2\alpha)\mathrm{E}\sigma_{t-k}^4 = (1+2\alpha)(1+2\alpha^2)^{t-k},\\
&\mathrm{E}\varepsilon_t^2\mathrm{E}\varepsilon_{t-k}^2 = (\mathrm{E}\sigma_{t-k}^2)^2 = 1,\\
&\mathrm{V}\varepsilon_t^2 = 3(1+2\alpha^2)^k\mathrm{E}\sigma_{t-k}^4 - (\mathrm{E}\sigma_{t-k}^2)^2 = 3(1+2\alpha^2)^t - 1,\\
&\mathrm{V}\varepsilon_{t-k}^2 = 3\mathrm{E}\sigma_{t-k}^4 - (\mathrm{E}\sigma_{t-k}^4)^2 = 3(1+2\alpha^2)^{t-k} - 1.
\end{aligned}$$

So by definition the time-varying autocorrelation function becomes

$$\rho_{k,t} = \frac{\mathrm{E}\varepsilon_t^2\varepsilon_{t-k}^2 - \mathrm{E}\varepsilon_t^2\mathrm{E}\varepsilon_{t-k}^2}{\sqrt{\mathrm{V}\varepsilon_t^2}\sqrt{\mathrm{V}\varepsilon_{t-k}^2}} = \frac{(1+2\alpha)(1+2\alpha^2)^{t-k} - 1}{\sqrt{3(1+2\alpha^2)^t - 1}\sqrt{3(1+2\alpha^2)^{t-k} - 1}}. \tag{A.22}$$

When $t \gg k > 0$ and $\alpha \neq 0$, one has approximately

$$\rho_k \approx \frac{1+2\alpha}{3}(1+2\alpha^2)^{-k/2}. \tag{A.23}$$

It is seen that, like a GARCH(1, 1) process, the autocorrelation function decreases exponentially. In the extreme case of $\alpha = 0$, so that $\sigma_t^2 = \sigma_{t-1}^2 = \cdots = \sigma_1^2 = \sigma^2$, i.e., the variance is constant over time and there is no heteroskedasticity in the series concerned, then (A.22) gives $\rho_k = 0$ as might be expected. On the other extreme, if $\alpha = 1$, so that $\sigma_t^2 = \varepsilon_{t-1}^2$, then (A.23) gives

$$\rho_{k,t} = \sqrt{(3^{t-k+1} - 1)/(3^{t+1} - 1)}. \tag{A.24}$$

When $t \gg k > 0$, (A.24) becomes

$$\rho_k \approx 3^{-k/2}, \tag{A.25}$$

and again it is exponentially decreasing.

Similar results can be derived for the IGARCH(1, 1) process with a drift. Assume now

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + (1-\alpha)\sigma_{t-1}^2 \tag{A.26}$$

and $\sigma_0^2 = \omega$, a constant. Then

$$\mathrm{E}\sigma_t^2 = \mathrm{E}(\omega + \alpha\varepsilon_{t-1}^2 + (1-\alpha)\sigma_{t-1}^2) = \omega + \mathrm{E}\sigma_{t-1}^2 = (t+1)\omega \tag{A.27}$$

and

$$\begin{aligned}\mathrm{E}\sigma_t^4 &= \mathrm{E}[\omega^2 + \alpha^2\varepsilon_{t-1}^4 + (1-\alpha)^2\sigma_{t-1}^4 + 2\omega\alpha\varepsilon_{t-1}^2 \\ &\quad + 2\omega(1-\alpha)\sigma_{t-1}^2 + 2\alpha(1-\alpha)\varepsilon_{t-1}^2\sigma_{t-1}^2] \\ &= \omega^2 + (1+2\alpha^2)\mathrm{E}\sigma_{t-1}^4 + 2\omega\mathrm{E}\sigma_{t-1}^2 \\ &= \omega^2 + (1+2\alpha^2)[\omega^2 + (1+2\alpha^2)\mathrm{E}\sigma_{t-2}^4 + 2\omega\mathrm{E}\sigma_{t-2}^2] + 2\omega\mathrm{E}\sigma_{t-1}^2 \\ &= \omega^2 + \omega^2(1+2\alpha^2) + \omega^2(1+2\alpha^2)^2 + \cdots + \omega^2(1+2\alpha^2)^t\mathrm{E}\sigma_0^4 \\ &\quad + 2\omega\mathrm{E}\sigma_{t-1}^2 + 2\omega(1+2\alpha^2)\mathrm{E}\sigma_{t-2}^2 + \cdots + 2\omega(1+2\alpha^2)^{t-1}\mathrm{E}\sigma_0^2 \\ &= \omega^2\sum_{i=0}^{t}(1+2\alpha^2)^i + 2\omega^2\sum_{i=1}^{t} i(1+2\alpha^2)^{t-i}.\end{aligned} \tag{A.28}$$

When $\alpha \neq 0$, (A.28) becomes

$$\mathrm{E}\sigma_t^4 = \omega^2\frac{(1+2\alpha^2)^{t+1}-1}{2\alpha^2} + \omega^2\frac{(1+2\alpha^2)^{t+1}-(1+2\alpha^2)-2\alpha^2 t}{2\alpha^4}.$$

When t is large, one has approximately

$$\mathrm{E}\sigma_t^4 \approx \frac{\omega^2}{2\alpha^2}\left(1+\frac{1}{\alpha^2}\right)(1+2\alpha^2)^{t+1}. \tag{A.29}$$

By the definition of an autocovariance function, one has

$$\begin{aligned}\gamma_{k,t} &= \mathrm{E}\varepsilon_t^2\varepsilon_{t-k}^2 - \mathrm{E}\varepsilon_t^2\mathrm{E}\varepsilon_{t-k}^2 \\ &= k\omega\mathrm{E}\sigma_{t-k}^2 + (1+2\alpha)\mathrm{E}\sigma_{t-k}^4 - (\omega k + \mathrm{E}\sigma_{t-k}^2)\mathrm{E}\sigma_{t-k}^2 \\ &= (1+2\alpha)\mathrm{E}\sigma_{t-k}^4 - (\mathrm{E}\sigma_{t-k}^2)^2.\end{aligned} \tag{A.30}$$

So

$$\begin{aligned}\rho_k &= \frac{(1+2\alpha)\mathrm{E}\sigma_{t-k}^4 - (\mathrm{E}\sigma_{t-k}^2)^2}{\sqrt{[3\mathrm{E}\sigma_t^4 - (\mathrm{E}\sigma_t^2)^2][3\mathrm{E}\sigma_{t-k}^4 - (\mathrm{E}\sigma_{t-k}^2)^2]}} \\ &= \left[\frac{(1+2\alpha)}{2\alpha^2}(1+2\alpha^2)^{t-k+1}\left(1+\frac{1}{\alpha^2}\right) - (t-k+1)^2\right] \\ &\quad \times\left[\frac{3}{2\alpha^2}(1+2\alpha^2)^{t+1}\left(1+\frac{1}{\alpha^2}\right) - (t+1)^2\right]^{-1/2} \\ &\quad \times\left[\frac{3}{2\alpha^2}(1+2\alpha^2)^{t-k+1}\left(1+\frac{1}{\alpha^2}\right) - (t-k+1)^2\right]^{-1/2}.\end{aligned} \tag{A.31}$$

When $t \gg k > 0$, one has approximately

$$\rho_k \approx \frac{1 + 2\alpha}{3}(1 + 2\alpha^2)^{-k/2}, \tag{A.32}$$

which satisfies $0 \leqslant \rho_k \leqslant 1$ when $0 \leqslant \alpha \leqslant 1$. Comparing (A.23) with (A.32), it is seen that the autocorrelation functions for IGARCH(1,1) models with or without a drift are the same.

When $\alpha = 0$, the conditional variance process becomes a deterministically increasing step function independent of ε_t, which further gives us that ε_t is an independent process with deterministically increasing variance. No autocorrelation should be expected for ε_t^2 under this situation. The following algebra confirms this. Substitute $\alpha = 0$ into (A.28), one gets

$$\mathrm{E}\sigma_t^4 = \omega^2(t + 1) + \omega^2(t + 1)t = \omega^2(t + 1)^2, \tag{A.33}$$

substituting (A.33) into (A.30), one has

$$\gamma_{k,t} = (1 + 2\alpha)\mathrm{E}\sigma_{t-k}^4 - (\mathrm{E}\sigma_{t-k}^2)^2 = \omega^2(t + 1)^2 - \omega^2(t + 1)^2 = 0.$$

So, as for the IGARCH(1,1) process without a drift, the autocorrelation function is zero when the lagged error term does not enter the conditional variance process.

References

Baillie, R.T., T. Bollerslev, and H. Mikkelsen, 1996, Fractionally integrated generalized autoregressive conditional heteroskedasticity, Journal of Econometrics, forthcoming.

Bollerslev, T., 1986, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31, 307–327.

Bollerslev, T., 1988, On the correlation structure for the generalized autoregressive conditional heteroskedasticity process, Journal of Time Series Analysis 9, 121–131.

Ding, Zhuanxin, 1994, Time series analysis of speculative returns, Ph.D. dissertation (University of California, San Diego, CA).

Ding, Zhuanxin, C.W.J. Granger, and R.F. Engle, 1993, A long memory property of stock market returns and a new model, Journal of Empirical Finance 1, 83–106.

Engle, R.F., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, Econometrica 50, 987–1008.

Engle, R.F. and T. Bollerslev, 1986, Modeling the persistence of conditional variances, Econometric Review 5, 1–50, 81–87.

Granger, C.W.J., 1980, Long memory relationships and the aggregation of dynamic models, Journal of Econometrics 14, 227–238.

Granger, C.W.J. and Z. Ding, 1994, Stylized facts on the temporal and distributional properties of daily data from speculative markets, Economics Department working paper 94-19 (University of California, San Diego, CA).

Granger, C.W.J. and Z. Ding, 1995, Some properties of absolute returns, an alternative measure of risk, Annales d'Economie et de Statistique 40, 67–91.

Granger, C.W.J. and Z. Ding, 1996, Varieties of long memory models, Journal of Econometrics, this issue.

Granger, C.W.J. and R. Joyeux, 1981, An introduction to long-memory time series models and fractional differencing, Journal of Time Series Analysis 1, 15–29.

Hosking, J.R.M., 1981, Fractional differencing, Biometrika 68, 165–176.

Johnson, N.L. and S. Kotz, 1970, Continuous univariate distributions – 2 (Wiley, New York, NY).

Nelson, D.B., 1980, Stationarity and persistence in the GARCH(1, 1) model, Econometric Theory 6, 318–334.

Taylor, S., 1986, Modeling financial time series (Wiley, New York, NY).