The Contrast of Parametric and Nonparametric Volatility Measurement Based on Chinese Stock Market

Xinwu Zhang, Yan Wang, and Handong Li*

School of Management, Beijing Normal University,
Beijing, 100875, People's Republic of China
li_handong@sina.com

Abstract. Most procedures for modeling and forecasting financial asset return volatilities rely on restrictive and complicated parametric GARCH or stochastic volatility models. The method of realized volatility constructed from high-frequency intraday returns is an alternative choice for volatility measurement. In this paper we make an empirical analysis on Chinese stock index data by using the method of nonparametric realized volatility. We find that the realized volatility can describe the Chinese stock index volatility very well. The original Chinese stock index return series show obvious leptokurtic, fat-tailed relative to the Gaussian distribution.We show that the return series standardized instead by the realized volatility are very nearly Gaussian distribution, and we find that the four minutes is a better choice as the best time interval to describe the volatility of Chinese stock market. We also make a contrast with the popular method of GARCH model, but the return series standardized instead by GARCH model don't accord with Gaussian distribution. The result shows that the realized volatility can describe the dynamic behaviors of Chinese stock market well. In a way, it indicates that the Chinese stock market is effective.

Keywords: realized volatility, GARCH, volatility measurement, conditional distribution.

1 Introduction

In econophysics research, much effort has been devoted on both the empirical and the theoretical level to such phenomena like fat-tailed distributions of financial fluctuations, persistent correlations in volatility of financial asset returns.The distributional characteristics of asset returns are the key ingredients for the pricing of financial instruments, portfolio allocation, performance evaluation, and managerial decision making. The most critical feature of the conditional return distribution is arguably its second moment structure, which is empirically the dominant time-varying characteristic of the distribution. Because volatility persistence renders

* Correspondence Author. This research is supported by National Science Foundation of China under grant No. 70601002.

J. Zhou (Ed.): Complex 2009, Part I, LNICST 4, pp. 618–627, 2009.
© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2009

high-frequency returns temporally dependent (Bollerslev, Chou and Kroner [1]), it is the conditional return distribution, and not the unconditional distribution, that is of relevance for risk management. This is especially true in high-frequency situations, such as monitoring and managing the risk associated with day-to-day operations of a trading desk, where volatility clustering is omnipresent.

It is well known that most of high-frequency financial asset return time series, such as the Shanghai stock market index, are leptokurtic, fat-tailed relative to the Gaussian distribution, and that the fat tails are typically reduced but not eliminated when return series are standardized by volatilities estimated from popular models such as GARCH and SV. A sizable literature explicitly attempts to model the fat-tailed conditional distribution, including, for example, Engle and Gonzalez-Rivera [2], and K. Chen, C. Jayprakash and B. Yuan [3].

Andersen, Bollerslev, Diebold and Labys[4] consider two major dollar exchange rates, and they show that returns standardized instead by the realized volatilities are nearly Gaussian. It indicates that it may be very important to find a suitable volatility measure to depict the conditional distribution of return series.

Assuming that return series dynamics operate only through the conditional variance, a standard decomposition of the time-t return series is:

$$r_t = \sigma_t * \varepsilon_t \tag{1}$$

Where σ_t refers to the time-t conditional standard deviation, and $\varepsilon \sim N(0, 1)$. Thus, given σ_t it would be straightforward to back out ε_t and assess its distributional properties. Of course σ_t is not directly observable. When using an estimate of σ_t from GARCH or SV model the distributions of the resulting standardized returns are typically found to be fat-tailed, or leptokurtic.

The research about the volatility of Chinese stock market returns is concentrated in low frequency data by the method of GARCH model in recent years. We analyze high frequency data using the method of realized volatility, and find that the realized volatility method can describe the dynamic volatility behaviors of Chinese stock market return well.

The stock index return series show obvious leptokurtic, fat-tailed distribution, but after we standardize the return series which is processed with realized volatility, it doesn't refuse Gaussian distribution hypothesis. We also make a compare between the realized volatility and GARCH model, the result indicates that return distribution standardized by GARCH model refuse Gaussian distribution hypothesis.

This shows that the realized volatility can describe the dynamic behaviors of Chinese stock market well. Meanwhile the result is according with the hypothesis that the series of stock index obey Semimartingale Stochastic Differential Equations, which suggest that the Chinese stock market is available.

2 Realized Volatility and *GARCH* Model

2.1 Realized Volatility

In 1980, Merton noticed that the variance in the fixed period of independent and identically distributed random variables could be estimated by the square sum of the return realization value ,and as long as the frequency is enough high, the estimation is very exact. French and Schwert etc [5] use the daily income in months to estimate the variance of every month. Andersen and Bollerslev[6], Hesieh [7], and Taylor S, Xu X [8] respectively use the square sum of within-day return to estimate the variance of daily return. In the recent years, Andersen and Bollerslev [9] put forward using the high-frequency to calculate the volatility rate. They put forward the method of measuring the realized volatility, that is to use the square sum of the return for some time as the estimation of the volatility. This estimation method is different from ARCH models and SV models ,and it doesn't depend on the model and doesn't need complex parameter estimation.

We first give some usual symbols.

Logarithm yield: As observed the financial assets price data in the time interval in time length $[0, T]$, we define return:

$$r^*(i, \delta) = lnS_i - lnS_{i-\delta} \tag{2}$$

It denotes the Continuously Compounded Return of financial assets in a certain time interval. Notice it is logarithmic.

In the financial environment of risk-free arbitrage, logarithm yield r_i^* of financial assets obeys Special Half Martingale process. If (Ω, I, P) supposing is a complete probability space, Information Filtration $(I_t)_{t\in[0,T]} \subseteq I$ is a increasing subalgebra series, $I_t$is p-complete and right continuity $S_t, t \in [0, t]$. This definition represent the price of financial primary assets in this spatial , then S_t is included in the information set I_t in t time.

Logarithm yield in the Δ period: The logarithm yield in the Δ period in the t-day invested in some financial primary assets is

$$r^*(t, \triangle) = X_t - X_{t-\Delta} \tag{3}$$

Where t represent the t-day $\Delta > 0$. As existing market microstructure noise, there exists the deviation in some degree between high-frequency financial data observed and potential real data, so we need proceed data according to following method

$$\tau_i = t - 1 + \frac{i}{n}, (i = 0, 1, \cdots, n) \tag{4}$$

Where τ_i represent observation value at the closing time point of i-observation period in t-day.

Note the observation value of log-price to be

$$Y_{\tau_i} = X_{\tau_i} + \varepsilon_{\tau_i} \tag{5}$$

where ε_{τ_i} is microstructure noise.

For simplification, let

$$E(\varepsilon_{\tau_i}) = E(\varepsilon), E(\varepsilon_{\tau_i}^2) = E(\varepsilon^2), \tag{6}$$

Where ε and log-price process are mutually independent.

$$Let \qquad r(\tau_i, \frac{1}{n}) = Y_{\tau_i} - Y_{\tau_i - 1}, (i = 1, \cdots, n) \tag{7}$$

It is logarithm yield in $[\tau_{i-1}, \tau_i]$ of financial assets base on observation data. when Δ=1, realized volatility in t-day is

$$RV_t = \sum_{i=1}^{n} [r(\tau_i, \frac{i}{n})]^2 \tag{8}$$

It is the realized volatility of financial primary assets in t-day in which the observation frequency of financial assets price data in t-day is $n + 1$.

2.2 GARCH Model

GARCH models are very popular for representing the dynamic evolution of the volatility of financial returns and have been extensively analyzed in the literature [see, e.g., Bollerslev, Engle, and Nelson [10], Bera and Higgins [11], Diebold and Lopez [12], and McAleer and Oxley [13], among many others].

The condition variance of GARCH model is represented as follows:

$$y_t = x_t \beta + \varepsilon_t \tag{9}$$

$$\varepsilon_t \mid \psi_{t-1} \sim N(0, h_t) \tag{10}$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{11}$$

$$= \alpha_0 + \alpha(L)\varepsilon_t^2 + \beta(L) h_t \tag{12}$$

To guarantee the condition variance , demand :

$$\alpha_0 > 0 \tag{13}$$

$$\alpha_i \geq 0, i = 1, \cdots, q \tag{14}$$

$$\beta_j \geq 0, j = 1, \cdots, p \tag{15}$$

$GARCH(p, q)$ is represented as GARCH process in which the order is p and q. Relative to ARCH, the advantages of GARCH model are that the lower orders GARCH model can represent higher order ARCH model, which can reduce the number of lag order of the model and then the recognition and estimation of the model become easier than ARCH. The stationary condition of GARCH model is as following:

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1 \tag{16}$$

The model of $GARCH(1,1)$,which is often used in finance analysis, is the popular and sample model in the GARCH families. It is usually found to be the better choice to describe the volatility of financial assets. It's formulation is

$$h_t = \alpha_0 + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1} \tag{17}$$

Where $\alpha_0 > 0, \alpha \geq 0, \beta \geq 0$.The necessary and sufficient condition of the stationary process of $GARCH(1,1)$ model is $\alpha + \beta < 1$.

3 Return Standardization

The asset return series are naturally decomposed as $r_t = \sigma_t * \varepsilon_t$, where $\varepsilon \sim N(0,1)$, and $\sigma_t$is the time conditional standard deviation. On rearranging this decomposition, we get the standardized return series,

$$\frac{r_t}{\sigma_t} = \varepsilon_t \tag{18}$$

Obviously, the result of the distribution is mainly dependent on the σ_t.

In practice, people have made a lot of research about σ_t, and many volatility models have been proposed. However, as formally shown by Andersen, Bollerslev, Diebold and Labys, the ex-post volatility over a day may be estimated to any desired degree of accuracy by summing sufficiently high-frequency returns within the day.

In this paper we use the realized volatility and GARCH model to get the conditional σ_t, and then compare the return distribution standardized instead by realized volatility with the one by GARCH model.

4 Empirical Analysis

4.1 Data

The sample data we used in this paper are the Shanghai securities integrated index from September 11 2006 to March 31 2008, which includes 82560 minutes datum of 344 trading day. There is a lunch break time in the stock market, where has a small fluctuate on the data. Because the fluctuate is no different from the other minute intervals, we connect them directly. For the two methods, we extract two datum groups. One is minute datum for the realized volatility method, the other is daily datum for the GARCH model.

4.2 Original Return Distribution

The Shanghai Securities Integrated Index(SSII) return is gained by following method

$$r_n = lnS_n - lnS_{n-1} \tag{19}$$

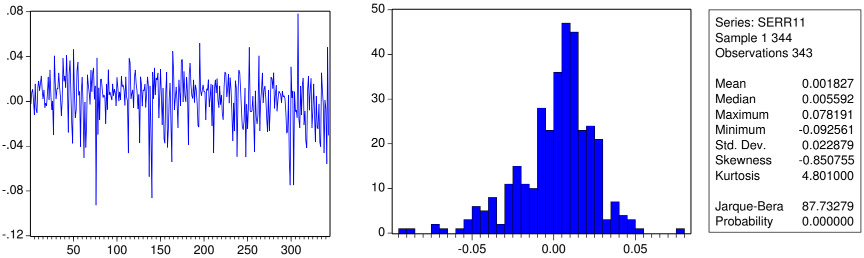

Fig. 1. The Fluctuation Diagram and the Histogram of daily returns

Where S_n represents the closing price of the n-th day in the observable sample series. We can get the original return distribution. The result is described in figure 1.

The value of skewness of original return distribution is -0.850755. The value of kurtosis of one is 4.801. The return distribution is leptokurtic, right deviation, and volatility clustering. The P-value of JB statistic is nearly to zero, so the result refuse the null hypothesis of Gaussian distribution. By testing, we also find that the return series show Heteroscedasticity.

4.3 Realized Volatility Analysis

We can use different time intervals to calculation daily realized volatility. On one hand, smaller the time interval is, less the information lose is, but bigger the error of microstructure is. On the other hand, longer the time interval is, more the information loses is, but the measurement error will become the main error. So the time interval should balance between the two kind errors. Of course, there may be the best time interval about realized volatility. Now we calculate the realized volatility at the time interval from 1 minute to 30 minutes by using the formula.

$$RV_t = \sum_{i=1}^{n} [r(\tau_i, \frac{i}{n})]^2 \tag{20}$$

Next we make the returns distribution standardized by the realized volatilities

$$\frac{r_t}{\sigma_t} = \varepsilon_t \tag{21}$$

The standardized results of day return by different time intervals realized volatility is indicated in table 1.

The JB statistic obeys the $\chi^2(2)$ distribution. The P- value represents the reception level of Gaussian distribution. The null hypothesis of the test is that the standardized return obeys Gaussian distribution. At the test significance level of 0.05, all of the P-values of statistics are not significant, it indicates that

Table 1. The standardized results of day return

	Mean	median	Std.dev	skewness	kurtosis	Jarque-bera	probability
1	0.42766	0.58696	1.9484	-0.15212	3.1275	1.5553	0.45949
2	0.36039	0.43083	1.5168	-0.06704	3.0276	0.26785	0.87465
3	0.31757	0.37075	1.3231	-0.03238	3.0351	0.077498	0.96199
4	0.29917	0.34856	1.2235	-0.01076	3.0022	0.006693	0.99666
5	0.28341	0.32807	1.1775	0.000132	3.0269	0.010369	0.99483
6	0.28425	0.32286	1.1599	0.018464	3.0681	0.085759	0.95803
7	0.27592	0.31516	1.1625	0.044382	3.118	0.31155	0.85575
8	0.28733	0.30697	1.1845	-0.00123	2.9191	0.093557	0.9543
9	0.29608	0.34409	1.2161	0.12277	3.2531	1.7775	0.41117
10	0.29499	0.32648	1.238	0.035407	3.1288	0.30863	0.857
11	0.31367	0.30706	1.2409	0.068909	2.9929	0.27217	0.87277
12	0.29474	0.35038	1.2749	-0.05773	2.9378	0.24584	0.88433
13	0.30299	0.33594	1.2421	-0.02154	2.8266	0.45607	0.7961
14	0.29761	0.33909	1.2292	0.014576	2.9816	0.01698	0.99155
15	0.3166	0.35388	1.2783	0.03567	3.0846	0.17491	0.91626
16	0.29395	0.35401	1.2842	-0.1579	3.2056	2.0294	0.3625
17	0.3009	0.34399	1.2394	0.090443	3.0826	0.56503	0.75388
18	0.3266	0.33246	1.2575	0.11249	3.0649	0.78344	0.67589
19	0.30935	0.36246	1.2972	0.074532	3.1412	0.60254	0.73988
20	0.32698	0.35293	1.3334	0.10732	3.3089	2.0218	0.36389
21	0.30482	0.33763	1.2705	0.083091	3.1353	0.6564	0.72022
22	0.31655	0.32452	1.3127	-0.01417	3.0678	0.077192	0.96214
23	0.30349	0.31914	1.2738	-0.04182	3.0113	0.10179	0.95038
24	0.29482	0.36018	1.3276	-0.21238	3.1978	3.1377	0.20829
25	0.30208	0.3545	1.2686	-0.14879	2.9742	1.275	0.5286
26	0.30632	0.32425	1.2604	-0.0118	2.8626	0.27773	0.87035
27	0.31121	0.32553	1.346	0.081218	3.1796	0.83824	0.65763
28	0.30952	0.3517	1.2952	0.096925	3.1082	0.70443	0.70313
29	0.33009	0.3457	1.2737	0.256	3.3665	5.6659	0.058839
30	0.33415	0.35503	1.4182	0.095115	3.4102	2.9217	0.23204

the test result accepts the null hypothesis of Gaussian distribution. So the return standardized by the realized volatilities is very nearly Gaussian.

There is a big span at different time intervals, we plot the P- values at the figure 2.

From figure 2 we can see that the P-values from 3,4,5,6,8,14,15,22,23 minutes interval realized volatility are larger than 0.9, most of P-values are bigger than 0.6, which show that normality is obvious. Furthermore we can see that the P-values of the time interval from 3 to 6 minutes are bigger than the other sections, so we think that the time section of 3 to 6 minutes is the best time interval of realized volatility about Chinese stock market, and 4 minutes is a proper choice.

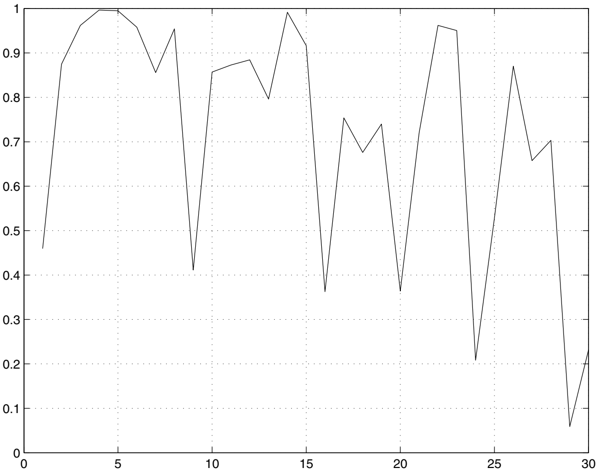

Fig. 2. The P-values of different time intervals

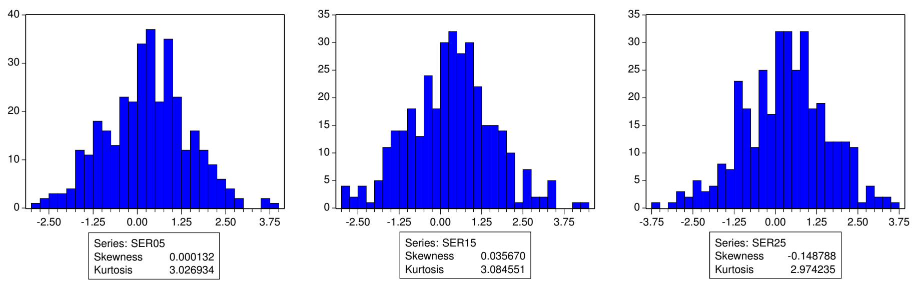

Fig. 3. The Histogram at the time intervals of 5,15,25 min

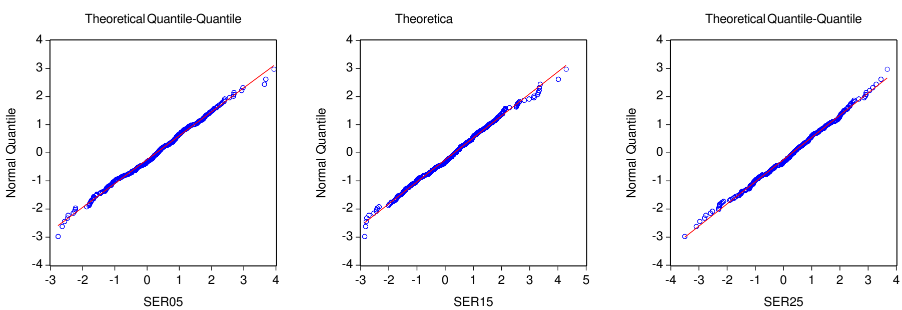

Fig. 4. The Quantile-Quantile Plots at the time intervals of 5,15,25 min

After standardized, the skewness value of the return series is nearly to zero, and the kurtosis value is nearly to 3. For example, the standardized result of 5,15,25 minutes are in figures 3, and 4.

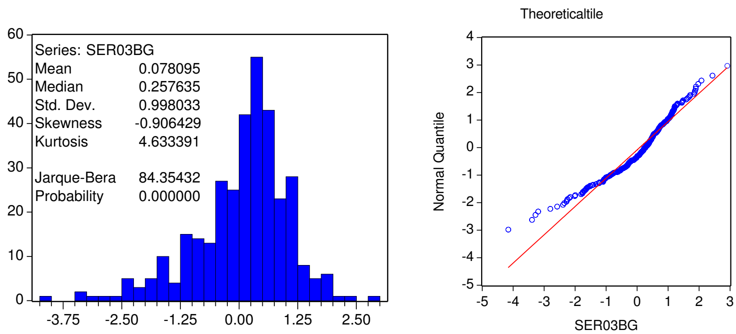

Fig. 5. The Histogram and Quantile-Quantile Plots of standardized return

4.4 GARCH Model Analysis

To understand the differences between the realized volatility and GARCH model, we also make an empirical analysis by utilizing the GARCH model.

A large number of literature show that $GARCH(1,1)$ model can describe the time variation of volatility well. So we choice $GARCH(1,1)$ model to describe the heteroscedesticity of return series. We choose the $AR(3) - GARCH(1,1)$ model.

$$R_n = 0.1138R_{n-3} + \varepsilon_t \tag{22}$$

$$h_t = 0.0000434 + 0.0601\varepsilon_{t-1}^2 + 0.8620h_{t-1} \tag{23}$$

After obtaining the conditional variance series, we can gain day return series standardized by time-varying conditional standard deviation from $GARCH(1,1)$. The distribution plot is in figure 5.

From the figures 5, we can see that $GARCH(1,1)$ model may describe the volatility of SSII, but it isn't accurate. The fitting Residual Error series are leptokurticfat-tailedright deviation. The P-value of JB statistic is nearly to zero, so it refuses the Hypothesis of Gaussian distribution.

5 Conclusion and Directions for Future Research

In this paper we use the nonparametric model, realized volatility and the Parametric model, $GARCH(1,1)$ respectively to describe the volatility of logarithm yield by utilizing the high frequency data of SSII in Chinese stock market. The result shows that realized volatility can not only gain the most of the high frequency information, but also return series standardized by realized volatility is basically accord with Gaussian distribution after taking logarithm and standardizing, which supports the Efficient Market Hypothesis in Chinese stock market. The standardized return series of $GARCH(1,1)$ are leptokurtic, fat-tailed, right deviation, which aren't accord with Gaussian distribution. It indicates that GARCH model can not correctly describe the dynamic of return series in Chinese stock market. The empirical results show that the volatility measurement of realized volatility is more accurate than GARCH model's.

Furthermore, we also study the best time interval of sampling about realized volatility. The result show that 4 minutes is a best time interval, which can make the daily realized volatility more nearly to Gaussian distribution.

Based on our research result, we plan to go on our research about Chinese stock market volatility at the following aspects: The first is to investigate the dynamic distribution of stock index return, and to determine the effects of different sampling frequency realized volatility. The second is to analyze the economics meanings of realized volatility to Chinese stock market, and to investigate the dynamic mechanism of practical volatility. The third is to make a contrast of accuracy and reliability of forecasting by different methods, so as to provide better methods for empirical research in future.

References

1. Bollerslev, T., Chou, R.J., Kroner, K.F.: ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence. Journal of Econometrics 52, 5–59 (1992)
2. Engle, R.F., Gonzalez-Rivera, G.: Semiparametric ARCH Models. Journal of Business and Economic Statistics 9(4), 345–359 (1991)
3. Chen, K., Jayprakash, B.Y.: Conditional Probability as a Measure of Volatility Clustering in Financial Time Series. Europhysics Letters 18, 1–6 (2005)
4. Anderson, T.G., Bollerslev, T.: Exchange rate returns standardized by realized volatility are nearly Gaussian. Multinational Finance Journal 4, 159–179 (2000)
5. Kenneth, F., Schwert, G.W., Stambaugh, R.: Excepted Stock Returns and Volatility. Journal of Financial Economics 19, 3–30 (1987)
6. Anderson, T.G., Bollerslev, T.: Answering the critics: Yes, ARCH models do provide good vohtility forecasts. National Bureau of Economic Research (NBER) Working paper, No. 6023 (1997)
7. Hsieh, D.A.: Chaos and nonlinear dynamics: application to financial markets. The Journal of Finance 46, 1839–1877 (1991)
8. Taylor, S.J., Xu, X.: The incremental volatility information in one million foreign exchange quotations. Journal of Empirical Finance 4, 317–340 (1997)
9. Andersen, T.G., Bollerslev, T., Diebold, F.X.: Parametric and nonparametric volatility measurement. In: Hansen, L.P., AytSahalia, Y. (eds.) Handbook of Financial Econometrics. North Holland, Amsterdam (2002) (forthcoming)
10. Bollerslev, T., Engle, R., Nelson, D.: ARCH Models. In: Handbook of Econometrics, vol. IV, pp. 2959–3038. North-Holland, Amsterdam (1994)
11. Bera, A.K., Higgings, M.L.: A Survey of ARCH Models: Properties, Estimation and Testing. Journal of Economic Surveys 7, 305–366 (1993)
12. Diebold, F.X., Lopez, J.A.: Macroeconomics: Developments, Tensions and Prospects. Blackwell, Oxford (1995)
13. McAleer, M., Oxley, L.: Contributions to Financial Econometrics. Blackwell, Oxford (2003)