Long Memory in emerging markets: evidence from Chinese Stock Market

Chaoqun Ma\textsuperscript{a} Hongquan Li\textsuperscript{b,*} Lin Zou\textsuperscript{a}

(a College of Business Administration, Hunan University, Changsha, P.R.China, 410082)
(b School of Business, Hunan Normal University, Changsha, P.R.China, 410081)

Abstract: The notion of long memory, or long-term dependence, has received considerable attention in empirical finance. This paper makes two main contributions. First, the paper aims to provide evidence of nonlinear (long memory) dynamics in the equity market of china. Analysis of market patterns in china market (a typical emerging market) instead of U.S. market (a developed market) will be meaningful because little previous research on the behaviors of emerging markets has been carried out. Secondly, we aim at the comprehensive search of long memory feature in China stock market returns as well as volatility. While many empirical works were done on the detection of long memory in return series, very few investigations focused on the market volatility, though the long-term dependence in volatility may lead to some types of volatility persistence as observed in financial markets and affect volatility forecasts and derivative pricing formulas. So, using modified rescaled range analysis and ARFIMA model testing, this study examined long-term dependence in Chinese stock market returns and volatility. The results show that although the returns themselves contain little serial correlation, the variability of returns has significantly long-term dependence. It would be beneficial to encompass long memory structure to assess the behavior of stock prices and research on financial market theory.

Key words: ARFIMA model, Long memory, Modified rescaled range analysis, Stock market

1 Introduction

Over the past decade, a number of empirical studies have documented the existence of nonlinear structure in financial markets. The analyses in nonlinear framework are important for at least two principal reasons. First, the effort devoted to study time series reflects the fact that nonlinearities convey information about the structure of the series under study. Second, these nonlinearities provide insight into the nature of the process governing the structure of these time series.

Long memory, a key symptom of nonlinear dynamics, implies the presence of significant autocorrelation, which invalidates the weak form of market efficiency. Recently, many empirical studies have been performed to detect the presence of long memory pattern in various stocks and indices returns (see [1], [2], [3]). Most of them concluded weak or no evidence of long memory in US, UK stock markets. However, some evidence of long-term dependence was found in Singaporean, Korean, Athens and Taiwanese indices (see [4], [5], [6]).

While many empirical works were done on the detection of long memory in return series, very few investigations focused on the market volatility, though the long-term dependence in volatility may lead to some types of volatility persistence as observed in financial markets. An innovative study is given in Ding et al. (see [7], [8]), which discovered that the autocorrelations of

\* Corresponding author. E-mail Address: Lhquan2000@126.com (Hongquan Li).
the absolute return for the S&P 500 index decay very slowly. The same phenomenon was observed in some fractional power transformations of the absolute return. The presence of this long memory persistence property provides us an important guidance in modeling volatility. Strong evidence of long memory volatility was detected in an equally weighted and a value-weighted index. There are two implication of the existence of long memory volatility. First, the long-term dependence in volatility means that it is problematic to use short memory models, such as ARMA models, to do volatility forecasting. Second, as the volatility dynamic plays a very important role in derivative pricing, it may be beneficial to incorporate the long-term volatility structure in deriving pricing formulas.

This paper makes two main contributions. First, the paper aims to provide evidence of nonlinear (long memory) dynamics in the equity market of China. Analysis of market patterns in China market (a typical emerging market) instead of U.S. market (a developed market) will be meaningful because little previous research on the behaviors of emerging markets has been carried out. Secondly, we aim at the comprehensive search of long memory feature in Chinese stock market returns as well as volatility. Besides using returns and squared returns, we use absolute returns as well. Comparing with the squared returns, the absolute returns are more robust to outliers and so long memory testing results based on the absolute transformation are less sensitive to larger stock price movements.

This paper is divided into four sections. Section 2 describes the methods used to detect and model long memory in time series. The third section discusses the data and presents the empirical results. The final section provides a brief summary and conclusion.

2 Methodology

A long memory process is a stationary process, which exhibits significant correlation at large lags. In other words, a stationary process has long memory if its autocorrelation function, say $\rho(k)$, has a hyperbolic decay:

$$\rho(k) \sim c k^{-2d-1}, \text{ as } k \to \infty, \text{ where } c \neq 0, \text{ and } d < 0.5 \quad (2.1)$$

Thus the autocorrelation function of a long memory process follows a power law, slower than the exponential decay. Power-law decay is slower than the exponential decay and, since $d<0.5$, the sum of the autocorrelation coefficients of such series approaches infinity. The speed of decay of the series autocorrelation function is related to the Hurst exponent by $H=0.5+d$.

A popular method of capturing the type of behavior defined in Equation (2.1) is the Autoregressive Fractally Integrated Moving Average (ARFIMA) model:

A time series $\{X_t\}$ follows an ARFIMA $(p, d, q)$ process if:

$$\Phi_p(B)(1-B)^d X_t = \Theta_q(B)\epsilon_t \quad (2.2)$$

Where

$$\Phi_p(B) = 1 - \Phi_1 B - \cdots - \Phi_p B^p,$$

$$\Theta_q(B) = 1 + \Theta_1 B + \cdots + \Theta_q B^q,$$
\[(1 - B)^d = 1 - dB - \frac{d(1-d)}{2!} B^2 - \ldots\]

and \( \varepsilon_t \) are i.i.d. distributional with mean zero and variance \( \sigma^2 < \infty \).

The properties of the ARFIMA model presented by Granger and Joyeux (see [9]) and in Hosking (see [10]): (1): if the roots of \( \Phi_p(B) \) and \( \Theta_q(B) \) are outside the unit circle and \(|d| < 0.5\), then \( X_t \) is both stationary and invertible; (2): if \( 0 < d < 0.5 \), the ARFIMA model is capable of generating stationary series which are persistent. In this case the process displays long memory characteristics, such as a hyperbolic autocorrelation decay to zero; (3): if \( d \geq 0.5 \) the process is non-stationary; (4): when \( d = 0 \) there is an ARMA process and it exhibits short memory; (5): When \(-0.5 < d < 0\) the ARFIMA process is said to exhibit intermediate memory or antipersistence.

There are many statistical tests to detect the existence of long memory in a time series. The first procedure was based on the behavior of the average range (R) rescaled by the average standard deviation (S), as a function of sample size. Hurst (1951) defined the R/S statistic as follows: for a time series with total observations T, there is an integer n, \( n \leq T \) then there exists the R/S statistic defined as

\[ Q(n) = \frac{R(n)}{S(n)} \quad (2.3) \]

where \( R(n) \) is the range given by

\[ R_n = \max_{1 \leq k \leq n} \sum_{j=1}^{k} (x_j - \bar{x}) - \min_{1 \leq k \leq n} \sum_{j=1}^{k} (x_j - \bar{x}) \]

and \( S(n) \) is the sample standard deviation of \( X_t \) over the period of n. As n increase, the following holds:

\[ \log Q(n) = \text{cons} \tan t + H \log(n) \quad (2.4) \]

where H is the Hurst exponent.

Thus, the Hurst exponent can be obtained by regressing \( \log(Q(n)) \) on \( \log(n) \) for different values of n.

The main advantage of the above R/S analysis as that the procedure of H estimation is independent of the distribution assumption for a given time series. However, the R/S statistic was reported to have bias when (i) the series contains the short-term memory, (ii) the series is characterized by heterogeneities and (iii) the series is nonstationary.

Lo (see [1]) proposed a modified version of the R/S statistic, which is robust even if the
presence of a short memory and heterogeneity. Lo’s modified R/S statistic can be defined as follows:

$$Z_n = \frac{R(n)}{\sigma_n(q)} = \frac{1}{\sigma_n(q)} \left[ \max_{1 \leq k \leq n} \sum_{j=1}^{k} (x_j - \bar{X}) - \min_{1 \leq k \leq n} \sum_{j=1}^{k} (x_j - \bar{X}) \right]$$

(2.5)

$$\sigma^2_n(q) = S_n^2(1 + 2 \sum_{j=1}^{q} W_j(q) \rho_j)$$

$$\rho(j)$$ is the jth order autocorrelation of \( \{x_t\} \) and \( W_j(q) \) is the Bartlett window weight

$$W_j(q) = 1 - \frac{j}{q+1}, \quad q < n$$

and q is the optimal lag of autocorrelation function. It cannot be too small; otherwise the effect of the short-term dependence cannot be entirely removed. However, using a large lag value may lower the power of the test and hence a compromise has to be selected in practice.

An alternative way of testing the long memory is based on the fractional differencing model (ARFIMA model). The value of d in ARFIMA model may be estimated using several techniques, such as semiparametric estimation (Geweke and Porter-Hudak, 1983. See [12]), approximate maximum likelihood estimation and Bayesian techniques. Here, the d value is estimated using GPH technique. It presents the following spectral regression method to estimate d and perform hypothesis testing:

$$\ln \{ I(\hat{\lambda}_j, T) \} = c - d \ln \{ 4 \sin^2 (\hat{\lambda}_j, T/2) \} + \varepsilon, \quad j = 1, 2, \ldots, n$$

(2.6)

Where \( I(\hat{\lambda}_j, T) \) is the periodogram value of frequency \( \hat{\lambda}_j \), which will depend on the sample size T, and \( n = g(T) \ll T \) is the number of Fourier frequencies included in the spectral regression.

In practice, \( g(T) = T^{0.5} \).

3 Results

In the section, the empirical results for China stock markets data are discussed. The data used in this study consist of daily stock index observations over the period 1 January 1991 to 9 March 2004. The data under consideration are: ShangHai Composite Index, denoted as SHCI; ShenZhen Component Index, denoted as SZCI. All return series are calculated using the continuously compounded approach. In other words, returns are defined as the first difference of the logarithmic prices. We denote the return of a series at time \( t \) by \( r_t \). Two transformed series are consideration, namely the square mean deviation \((r_t - \bar{r})^2\) and the absolute deviation \(|r_t - \bar{r}|\), where \( \bar{r} \) is the sample mean of \( r_t \). In the traditional GARCH models, the variable \((r_t - \bar{r})^2\), playing the role of...
‘variance’, follows an ARMA process where long memory should not be detected. Similarly, the variable $|r_t - \bar{r}|$ is used as a proxy of the volatility at time $t$. All the two variables serve the same purpose of providing some measures for the returns fluctuation. Similar consideration can be found in Ding et al. (see [7], [8]) where different power transformations of the absolute returns were used to investigate a long memory property of stock market returns.

The classical R/S test, modified R/S test and the GPH test (estimate the value of $d$ in ARFIMA model) were adopted in this study. When applying the modified R/S test, the effect of the short-term dependence on the long memory test should be removed by choosing a suitable lag $q$ in equation (2.5). In order to check the sensitivity of the results towards the choice of the lag value, we perform the modified R/S test for $q = 0, 3, 6, 9$ and $q^*$, where $q^*$ was selected by the data-dependent rule (see [1], [11]).

In the application of GPH test, it is crucial to select the number of terms $g(T)$ included in the spectral regression of Equation (2.6). In order to yield the asymptotic result, Geweke and Porter-Hudak (1983) suggested $g(T) = T^\alpha$, where $0 < \alpha < 1$. In this paper, we use $\alpha = 0.5$, 0.55 and 0.6 and denote the semiparametric estimates of $d$ associated with $\alpha$ by $d(\alpha)$.

Tab.1 and Tab.2 present the empirical results. All test statistics come to similar conclusion. Although the returns themselves contain little serial correlation, the variability of returns has significantly long dependence. The important findings from the tables are summarized below.

### Table 1

<table>
<thead>
<tr>
<th>Index</th>
<th>Series</th>
<th>$H$</th>
<th>$\rho_1$</th>
<th>$q^*$</th>
<th>$V_n(0)$</th>
<th>$V_n(3)$</th>
<th>$V_n(6)$</th>
<th>$V_n(9)$</th>
<th>$V_n(q^*)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SZCI</td>
<td>$r_t$</td>
<td>0.62</td>
<td>0.059</td>
<td>4</td>
<td>1.36</td>
<td>1.20</td>
<td>1.08</td>
<td>1.02</td>
<td>1.15</td>
</tr>
<tr>
<td></td>
<td>$</td>
<td>r_t - \bar{r}</td>
<td>$</td>
<td>0.82</td>
<td>0.320</td>
<td>13</td>
<td>7.03*</td>
<td>3.67*</td>
<td>2.65*</td>
</tr>
<tr>
<td></td>
<td>$(r_t - \bar{r})^2$</td>
<td>0.71</td>
<td>0.216</td>
<td>10</td>
<td>4.18*</td>
<td>2.69*</td>
<td>2.12*</td>
<td>1.87*</td>
<td>1.82*</td>
</tr>
<tr>
<td>SHCI</td>
<td>$r_t$</td>
<td>0.64</td>
<td>0.053</td>
<td>4</td>
<td>1.41</td>
<td>1.22</td>
<td>1.13</td>
<td>1.10</td>
<td>1.18</td>
</tr>
<tr>
<td></td>
<td>$</td>
<td>r_t - \bar{r}</td>
<td>$</td>
<td>0.93</td>
<td>0.277</td>
<td>12</td>
<td>7.98*</td>
<td>4.26*</td>
<td>3.04*</td>
</tr>
<tr>
<td></td>
<td>$(r_t - \bar{r})^2$</td>
<td>0.59</td>
<td>0.020</td>
<td>2</td>
<td>2.40*</td>
<td>2.21*</td>
<td>2.05*</td>
<td>1.95*</td>
<td>2.28*</td>
</tr>
</tbody>
</table>

Note: (i) $\rho_1$ is the estimated first-order autocorrelation coefficient of the data. (ii) * significant at the 5% level,

### Table 2

<table>
<thead>
<tr>
<th>Index</th>
<th>Series</th>
<th>$H$</th>
<th>$\rho_1$</th>
<th>$q^*$</th>
<th>$V_n(0)$</th>
<th>$V_n(3)$</th>
<th>$V_n(6)$</th>
<th>$V_n(9)$</th>
<th>$V_n(q^*)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SZCI</td>
<td>$r_t$</td>
<td>0.62</td>
<td>0.059</td>
<td>4</td>
<td>1.36</td>
<td>1.20</td>
<td>1.08</td>
<td>1.02</td>
<td>1.15</td>
</tr>
<tr>
<td></td>
<td>$</td>
<td>r_t - \bar{r}</td>
<td>$</td>
<td>0.82</td>
<td>0.320</td>
<td>13</td>
<td>7.03*</td>
<td>3.67*</td>
<td>2.65*</td>
</tr>
<tr>
<td></td>
<td>$(r_t - \bar{r})^2$</td>
<td>0.71</td>
<td>0.216</td>
<td>10</td>
<td>4.18*</td>
<td>2.69*</td>
<td>2.12*</td>
<td>1.87*</td>
<td>1.82*</td>
</tr>
<tr>
<td>SHCI</td>
<td>$r_t$</td>
<td>0.64</td>
<td>0.053</td>
<td>4</td>
<td>1.41</td>
<td>1.22</td>
<td>1.13</td>
<td>1.10</td>
<td>1.18</td>
</tr>
<tr>
<td></td>
<td>$</td>
<td>r_t - \bar{r}</td>
<td>$</td>
<td>0.93</td>
<td>0.277</td>
<td>12</td>
<td>7.98*</td>
<td>4.26*</td>
<td>3.04*</td>
</tr>
<tr>
<td></td>
<td>$(r_t - \bar{r})^2$</td>
<td>0.59</td>
<td>0.020</td>
<td>2</td>
<td>2.40*</td>
<td>2.21*</td>
<td>2.05*</td>
<td>1.95*</td>
<td>2.28*</td>
</tr>
</tbody>
</table>

Note: (i) $\rho_1$ is the estimated first-order autocorrelation coefficient of the data. (ii) * significant at the 5% level,
Table 2: Semiparametric estimates of $d$

<table>
<thead>
<tr>
<th>Index</th>
<th>Series</th>
<th>$d(0.5)$</th>
<th>$d(0.55)$</th>
<th>$d(0.6)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SZCI</td>
<td>$r_t$</td>
<td>-0.043</td>
<td>0.086</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.48)</td>
<td>(1.19)</td>
<td>(0.72)</td>
</tr>
<tr>
<td></td>
<td>$</td>
<td>r_t - \bar{r}</td>
<td>$</td>
<td>0.329*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.36)</td>
<td>(4.75)</td>
<td>(4.96)</td>
</tr>
<tr>
<td></td>
<td>$(r_t - \bar{r})^2$</td>
<td>0.279*</td>
<td>0.310*</td>
<td>0.119**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.66)</td>
<td>(3.87)</td>
<td>(1.94)</td>
</tr>
<tr>
<td>SHCI</td>
<td>$r_t$</td>
<td>0.047</td>
<td>0.009</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.50)</td>
<td>(0.11)</td>
<td>(0.07)</td>
</tr>
<tr>
<td></td>
<td>$</td>
<td>r_t - \bar{r}</td>
<td>$</td>
<td>0.486*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.07)</td>
<td>(3.26)</td>
<td>(3.88)</td>
</tr>
<tr>
<td></td>
<td>$(r_t - \bar{r})^2$</td>
<td>0.069</td>
<td>0.053</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.14)</td>
<td>(1.17)</td>
<td>(1.37)</td>
</tr>
</tbody>
</table>

Note: * significant at the 5% level, ( ) denotes t-value.

Using the classical R/S method, we found the Hurst exponent for all the series is higher than the expected 0.5 (random walk series). This result is an indication of the existence of long-range dependence in stock markets (see Tab.1). However, after correcting for short-range dependence using Lo’s modified R/S analysis technique, the evidence of long-term memory in returns disappeared. As far as the two volatility proxies, the absolute deviation $|r_t - \bar{r}|$ and square mean deviation $(r_t - \bar{r})^2$, the modified R/S test shows positive results in all cases. It indicates strong and robust evidence of long memory in stock market volatility.

Tab.2 presents the estimates of $d$. The two returns series, which are determined as having no long memory by the modified R/S test, accord with the results by the GPH test as not rejecting $H_0: d = 0$. Next, we turn to the evidence of the absolute deviation $|r_t - \bar{r}|$. It is easy to observe that most of the estimates of $d$ for $|r_t - \bar{r}|$ are greater than those for $(r_t - \bar{r})^2$. Therefore, it is not surprising to see that applying the GPH test on the absolute mean deviations reveals strong evidence of long memory in the two indices.

4 Conclusion

In this paper, three procedures are applied, the classical R/S test, the modified R/S test and GPH test, to detect the existence of long-term dependence in volatility as well as returns for Chinese stock market composite indices. Two proxies of the variability of returns: the absolute mean deviation and the squared mean deviation were adopted in the study. The results show that though the returns themselves contain little serial correlation, the variability of returns has significantly long-term
dependence. It would be beneficial to encompass long memory structure to assess the behavior of stock prices and research on financial market theory.

5 Acknowledgement

This paper is supported by National Natural Science Foundation of China (70471030).

6 References


Email: mcq999@sina.com
Lhquan2000@126.com
Zoulin2800@126.com