An Examination of Long-term Memory Using the Intraday Stock Returns

Dr. Bwo-Nung Huang  
Institute of International Economics  
National Chung Cheng University  
Chia-Yi, 621 Taiwan  
Fax: 011886-5-2720816  
e-mail: ecdbh@ccunix.ccu.edu.tw

Dr. Chin W. Yang  
Department of Economics  
Clarion University of Pennsylvania  
Clarion, Pennsylvania 16214  
USA  
Fax: 814-2261910  
e-mail: "YANG@VAXA.clarion.edu

March 31, 1999
An Examination of Long-term Memory Using the Intraday Stock Returns

Abstract

In this paper, we show that by applying the modified R/S technique to the intraday data, there exists the phenomenon of long-term memory in both NYSE and NASDAQ indices. The policy implication is that an abnormal profit is entirely possible for certain time intervals during a trading day. The efficient market hypothesis may actually coexist with a global structure to comprise a fractal market hypothesis. In addition, via a univariate ARMA model, we can make more accurate forecasts on the stock returns in the time period when a long-term memory is found.

Keywords: modified R/S, fractal market hypothesis, intraday data

(JEL Classification No: G10)
Introduction

Modern portfolio theory has had a profound influence on investment behaviors. Without doubt, the foundation of capital market equilibrium lies on the efficient market hypothesis. In its extreme form, the random walk hypothesis, i.e., independent and identically distributed (iid) return reign its continued supremacy in the realm of empirical studies. It is crucially important that the random walk hypothesis holds or most of the statistical techniques in analyzing capital market equilibrium such as CAPM are open to question. A violation of the random walk hypothesis is fraught with statistical implications in such a way that it needs to be reevaluated with great caution. Are stock returns run by the vagaries of their history or history repeats itself? In the spectrum of noises, a random walk is characterized by the white noise in which the next outcome is completely unpredictable. A persistent pattern is represented by the black noise where a long-term dependence is present. (e.g., a streak of increases in stock returns). In contrast, a mean-reversion manifests itself in the form of the pink noise in which an increase in stock returns will have more than 50% chance to encounter a decrease. While the black and pink noise abound in nature (e.g., sun spots, tree rings and turbulence), such occurrences are extremely rare in the financial literature. In other words, the random walk hypothesis reigns supreme thus far. There is no reason why the field of financial economics is overwhelmingly dominated by the efficient market hypothesis which is in fact only a part of more general fractal market hypothesis. Thus far, anamolies to the efficient market hypothesis are limited to the weekend effect (Cross 1973; French 1980; Gibbons & Hess 1981; Lakonishok & Levi 1982; Keim & Stambaugh 1984; Jaff & Westerfield 1985) or January effect (Branch & Chang, 1990; Chang & Pinegar, 1990; Jones & Wilson, 1989; Thaler, 1987) in a given period of time of a week or a year. In the case of emerging markets, Huang (1995), Huang and Yang (1995) find evidence of inefficient market components using daily returns.

However, the findings of the efficient markets using relatively low frequency data does not necessarily imply that the result can be carried over to the scenario in which higher frequency data (intraday returns) are employed. There is a good possibility that the empirical results using daily returns may well misrepresent the true return generating process because daily returns are usually obtained from using returns at market open or
market close. Needless to say such a practice grossly ignores the return generating process during a given trading day. Little wonder that the great majority of empirical studies employing aggregate data (daily, weekly, monthly and annual stock returns) have found no trace of a long memory (Lo, 1991). On the surface, such a conclusion is hard to find fault with since the methodology (Lo, 1991) is very robust with respect to the distributional assumption of stock returns. However, when we look close up at the intraday returns, the picture may well be different. The explanation can be rather straightforward: in deriving aggregate stock returns, the anomalies may be aggregated out. In this paper, we examine the potential existence of the long memory using one-minute price data. The next section presents statistical properties of the data. In section 3, we present the test results based on the modified rescaled range (R/S) technique developed by Lo (1991). Section 4 illustrates a univariate ARMA model to better forecast the stock returns. A conclusion is given in section 5.

\[ \text{Statistical Properties of the Intraday Data} \]

Previous analyses using intraday data are, on the whole, limited to investigations on the shapes of the mean and the standard deviation of stock returns (e.g., Wood et al. 1985; Harris, 1986; Jain and Joh, 1988; McInish and Wood, 1990; Lockwood and Linn, 1990; Stoll and Whaley, 1990). More recently McInish and Wood (1991) observe a U-shape curve for the first-order autocorrelation coefficients. Most studies have found much greater volatilities in stock returns at market open and/or market close. This being the case, employing daily returns can grossly misrepresent the return generating process for a given trading day. The minute by minute stock returns of 1992 New York Stock Exchange (NYSE) and 1992 National Association of Securities Dealers over the Counter (NASDAQ) are obtain from the 1992 CRSP tape and Institute for the Study of Security Markets (ISSM) data file. We believe that 1992 is perhaps a typical year and NASDAQ data are not as frequently used.\(^1\) Besides, it is important to know if any difference in the stock returns exists between these two markedly different exchanges.

In total, there were 254 trading days in 1992 for the NYSE and the NASDAQ; and each trading days has 6½ hours (390 minutes). The markets open at 9:30 AM and close at 4:00 PM. In a regular trading day. There are 391 one-minute stock returns for each market. Parallel to that by McInish and Wood (1991), the mean and standard deviation of

---

\(^1\) Chan, Christie and Shultz (1995) analyze the NASDAQ trade patterns.
the stock returns are U-shaped as shown in Figure 1-4. This is to say that these statistics are markedly different at market open and market close especially in the NASDAQ. The shape of the standard deviation indicates a relatively wider bottom in the NASDAQ since changes in standard deviations are milder during a trading day as shown in Figure 2. In addition, an incomplete U (or J) curve of the standard deviation of the NYSE is found (Figure 2). The J curve phenomenon is in agreement with the result of the average bid-ask spread found by McInish and Wood (1993). On close inspection, the mean stock returns of the NYSE at market open is mildly different from that for the rest of a trading day (Figure 1). In contrast, the stock returns of the NASDAQ exhibit a more pronounced shift from market open. Viewed in this perspective, we speculate that the J Curve phenomenon is more conspicuous in the NASDAQ than that of the NYSE found by McInish and Wood (1992).

Clearly, the means and standard deviations of both exchanges at market close do not represent that of other time segments in a giving trading day. A moment’s inspection will reveal that they are noticeably greater, and as such should not be used as daily returns. The result derived from employing such daily returns is at best questionable. It needs to be reexamined carefully in the lens of modified R/S technique with the intraday stock returns.

\*\*\* Modified Rescaled Range Analysis and the Empirical Results

Does history repeat itself or it just wanders aimlessly in the domain of space-time configuration. Vagaries of natural occurrences have been well-documented throughout the history and perhaps the most popular one is the Brownian motion in classic physics. Given a sequence of stock returns (logarithmic price difference) \( r_1, \ldots, r_n \), the range (maximum distance particles travel) of the sums of cirrulative deviations from means of various lengths after adjusting for its standard deviation can be represented by

\[
\bar{Q}_n \equiv \frac{1}{S_n} \max_{1 \leq k \leq n} \left( \sum_{j=1}^{k} d_j - \bar{r}_n i \right) - \min_{1 \leq k \leq n} \left( \sum_{j=1}^{k} d_j - \bar{r}_n i \right)
\]

(1)

\[\] \[\]

\[\] \[\]
where $\bar{r}_n$ is the sample mean return; and $S_n$ is the sample standard deviation. For the time length $n>3$, we can calculate a $\tilde{Q}_n$ with different starting point. This is known as traditional recaled range (R/S) analysis developed by Hurst (1951) in which he found persistency patterns in water flow of the Nile. Rather surprisingly, the R/S technique is quite robust in detecting long-term dependence. In their classic paper, Mandelbrot and Wallis (1969) via Monte Carlo simulations showed the clear superiority of the R/S technique especially in the presence of non-Gaussian stock returns with pronounced skewness and significant kurtosis. Departure from normality assumption can lead to serious error in statistical inference (Huang and Yang, 1996). The almost sure convergence property even in the case of infinite variance adds yet another descrable property to the technique. For a long enough length of time (ten or more cyclical lengths of time), existence of nonperiodical cycles may be calculated. These desirable properties notwithstanding, the test statistic is very sensitive to the short-term dependence (e.g., autocorrelation). In order to take the short-term dependence into consideration, Lo (1991) propose the standard deviation estimator that properly accounts for weighted autocovariances up to lag $q$. With this modified standard deviation estimator, the modified R/S statistic is

$$\tilde{Q}_n = \frac{1}{\tilde{\sigma}_n(q)} \left[ \sum_{i=1}^{k} d_i \bar{r}_n - \text{Min}_{i \leq k \leq n} \sum_{j=1}^{k} d_j \bar{r}_n \right]$$

in which

$$\tilde{\sigma}_n = \frac{1}{n} \sum_{j=1}^{n} d_j \bar{r}_n + \frac{2}{n} \sum_{j=1}^{q} w_j(q) \sum_{j+1}^{n} d_j \bar{r}_n$$

Note that the weight is set as $w_j = 1 - j/(q+1)$ for $q < r$ so that the modified standard deviation of (3) is assured to be positive (see Newet and West, 1987). Unfortunately the choice of an optimum lag period $q$ is data dependent as shown by Andrew (1991). It can be rather arbitrary in the case of a finite sample.

The asymptotic property of the modified R/S, according to Kennedy (1976) and Siddiqui (1976) can be expressed in following relation:

$$V_n(q) \equiv \frac{1}{\sqrt{n}} Q_a \xrightarrow{a} V$$

(4)
where \( v \) follows the probability distribution as shown below:

\[
F_r(v) = 1 + 2 \sum_{k=1}^{\infty} (1 - 4k^2v^2)e^{-2(kv)^2}
\]  

(5)

Within this framework, the critical values, as are reported by Lo (1991) may be computed from (5). In Table 4, we present a list of the R/S statistics for the holding period of \( q=10 \) (two weeks) and \( q=20 \) (four weeks) and graph them in one-minute intervals as shown in Figure 5 and 6 respectively.

It becomes immediately clear that we fail to reject the null hypothesis for both indices at market open and market close. The result that there is no long-term memory (e.g., no evidence of fractionally-differenced ARIMA \((p, d, q)\) models with \(-\frac{1}{2} < d < \frac{1}{2}\) is consistent with Lo’s study (1991) in which daily returns were used. The hypothesis testing based on aggregate data that leads inexorably to the conclusion of no long-term memory may well be misleading.

Surprisingly, the existence of long-term memory in the stock returns of both NYSE and NASDAQ is detected in other time intervals during a trading day. As shown in Table 1, there are 44 (NYSE) and 49 (NASDAQ) time intervals in which we can reject the null hypothesis at 10% significance level. In the case of NYSE, 33% of the time intervals in which we reject the null hypothesis occurred within one hour after the market open while 25% of such time intervals were clustered within thirty minutes around the noon break. In the case of NASDAQ, 39% of 49 rejections took place within the time period of 10:00 ~ 12:00. The rest of the rejections were well dispersed in the remaining time intervals during a trading day.

In both cases, 11.28% (44/390) for NYSE and 12.56% (49/390) for NASDAQ represent percentages of time in which a persistence pattern can be recognized. In particular, one hour after the market open in the case of NYSE, there were 15 rejections or 25% (15/60) of the time in which abnormal profit can be realized. In the case of NASDAQ, the probability for abnormal profit is 15.83% (19/120) from 10:00 to noon. In either case, the probability for potential profit exceeds that allowed by the sampling error represented type one error of 10%.

Evidently, some portion of the stock returns could have been predicted in both markets. Such a finding is significant: a discernible and observant investor can beat the
odds by entering the market at the appropriate time intervals. Our results differ drastically from that using the daily returns (Lo, 1991). From another perspective, a study on cumulative returns using 15-minute intervals data (Harris, 1986) shows that a skillful investor could procure abnormal profit at appropriate time intervals.

Well-known in the literature, intraday stock returns exhibit to various degrees, spurious autocorrelations due to infrequent and/or nonsynchronous trading. As pointed out by Cohen (1983), infrequent and nonsynchronous tradings are considered the culprit of spurious correlation found in the stock returns. In general, such artificial correlations can be attributed to the infrequent trading of small-capitalization stocks. As new information is impounded into the stocks of large capitalization, it trickles down to stocks of small capitalization with lags. Thus a positive serial correlation is generated through the diffusion process. Such a induced positive serial correlation, according to Lo and MacKinlay (1988), causes more impacts on the equal-weighted index. Nonetheless it is difficult to ascertain that the positive serial correlation is a direct consequence of the infrequent tradings. By employing the variance ratio approach, Lo and MacKinlay (1988) are able to reject the random walk hypothesis. However the rejection of the null hypothesis is not attributed to infrequent tradings. In the similar vein, their second study (1990) indicates that there is sufficient evidence to support that nonsynchronous tradings is the source of the spurious correlation.

The studies by Lo and Mackinlay (1988, 1990) clearly suggest that the R/S model possesses several desirable statistic properties. For instance it explicitly takes short-run autocorrelation into consideration as compared to the classical R/S model (Mandelbrot, 1975). As reported in the previous section, about 12% of the time intervals in both NYSE and NASDAQ are not consistent with the efficient market hypothesis. In other words, one can secure considerable amount of abnormal profit via entering the market at appropriate time intervals. Granted that a well-specified structural model might have better forecasts on stock returns at various time intervals such a model is not feasible due to the lack of high-frequency data of economic variables. As such, we employ a time series model to compare the forecasting performance between the in-sample and out-sample time periods. In order to facilitate such comparison, we make use of Theil's U inequality coefficient as shown below:
The U statistic, being a unit-free measurement, is amenable to direct comparisons. Typically, the size of U reflects immediately the forecasting performance of the model. With U>1, it is indicative of the poor forecasting ability, i.e., worse than that of the naive model. In contrast, U<1 suggests that the model in general has reasonable forecasting power. Note that a small U may be the direct result of the presence of a strong trend in the time series.

To facilitate the comparison of forecasts of the stock returns, we employ the univariate Box & Jenkins model as shown below:\textsuperscript{5}

\[ r_t = \mu + \Phi_1 r_{t-10} + \Phi_2 r_{t-20} + a_t - \theta_1 a_{t-10} - \theta_2 a_{t-20} \]  

(7)

where \( r_t \) is the stock return in a given interval during a trading day; and \( a_t \) represents innovations consistent with the white noise process. For ease of comparison, the forecasting process is based on equation (7); and it consists of two cases. First, we estimate the coefficients of (7), using the entire 254 days of 1992, to perform within the sample or in-sample forecast. Second, the out-sample forecast is based upon the estimation made on the first 150 trading days, and then we compare the forecasting results over the sample period from the 151st to the 254th day. The resulting U statistic is reported in Table 2 for both NYSE and NASDAQ.

An examination of the U statistic in Table 2 suggests, in either in-sample or out-sample scenario, the existence of a better forecasting ability for the time interval in which the random walk hypothesis is rejected than that when the random walk hypothesis cannot be rejected for the market of NYSE.\textsuperscript{6} The similar result is found in the in-sample scenario

\textsuperscript{5}The form of equation (7) is specified such that it incorporates long lag terms to account for the potential long-term memory phenomenon.

\textsuperscript{6}In Lo's model, various forms of short-term dependence (strong-mixing) are impeded in the null hypothesis. If the null hypothesis is not rejected, one may test the existence of such a dependence against the random walk hypothesis.
for NASDAQ. However, in the case of out-sample the mean value of the U statistic for
the time intervals in which the random walk hypothesis cannot be rejected is slightly
smaller (better) than that when the null hypothesis is rejected (0.7699 < 0.7707). In terms
of maximum and minimum values of the U statistic, it clearly indicates (Table 2) that the
forecasting ability during the interval when the null hypothesis is rejected is superior to
that in which the null cannot be rejected. In summary, the results based on the univariate
time series model unravels that stock returns can be more accurately forecasted during the
time intervals in which the random walk hypothesis is rejected. Consequently, abnormal
profit could have been realized by a shrewd investor at an appropriate time interval.

© Conclusion

Empirical evidence abounds in supporting the efficient market hypothesis. The great
majority of studies have pointed out to the deep-rooted conviction that in the long run an
investor cannot out perform the market. Such an i.i.d. process precludes the recognition of
a pattern, and as a result, the efficient market hypothesis reigns supreme in most cases.
However, as is pointed out in the paper, stock returns at market close have at least two
weaknesses. In addition to its representation problem, the statistical moment at market
close is found significantly different from that of other time interval during a trading day.
Hence, the result obtained from employing daily returns can grossly mask the true return
generating process. In this paper, we make use of intraday stock returns of both NYSE
and NASDAQ with one-minute intervals. Parallel to the findings by McInish and Wood
(1991), the means and standard deviations of the intraday returns exhibit U-shaped
relation in both cases. Furthermore, we reject the null hypothesis in 44 and 49 of a total of
390 one-minute time intervals in a given trading day for NYSE and NASDAQ
respectively. The startling result signifies the existence of a long-term memory (in terms of
high frequency data) after short-term dependence is properly accounted for. Since the
speculation that infrequent and nonsynchronous trading may be related to the spurious
autocorrelation is largely refuted by Lo and Mackinlay (1990), the conclusion of a long-
term memory using high frequency data presents a real possibility that local randomness
and global determinism can actually coexist (Peters, 1994). Applying a univariate ARMA
model, we are able to show the existence of better forecasting performances for the time
intervals in which the long-term memory is found.
References


<table>
<thead>
<tr>
<th></th>
<th>NYSE</th>
<th></th>
<th>NASDAQ</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>9:32</td>
<td>12:19</td>
<td></td>
<td>9:30</td>
<td>12:08</td>
</tr>
<tr>
<td>9:34&lt;sup&gt;a&lt;/sup&gt;</td>
<td>12:24</td>
<td></td>
<td>9:32</td>
<td>12:10&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>9:36</td>
<td>12:47</td>
<td></td>
<td>9:40</td>
<td>12:18</td>
</tr>
<tr>
<td>9:38&lt;sup&gt;b&lt;/sup&gt;</td>
<td>12:59</td>
<td></td>
<td>9:50</td>
<td>12:52</td>
</tr>
<tr>
<td>9:40</td>
<td>1:16</td>
<td></td>
<td>9:51</td>
<td>12:54</td>
</tr>
<tr>
<td>9:43</td>
<td>1:17</td>
<td></td>
<td>9:57</td>
<td>1:02</td>
</tr>
<tr>
<td>9:44&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1:26&lt;sup&gt;a&lt;/sup&gt;</td>
<td>10:04</td>
<td>1:07</td>
<td></td>
</tr>
<tr>
<td>9:52</td>
<td>1:38&lt;sup&gt;b&lt;/sup&gt;</td>
<td>10:05</td>
<td>1:29&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>10:01&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1:46</td>
<td></td>
<td>10:21</td>
<td>1:31</td>
</tr>
<tr>
<td>10:07&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1:47</td>
<td></td>
<td>10:26&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1:41</td>
</tr>
<tr>
<td>10:17</td>
<td>2:06</td>
<td></td>
<td>10:31</td>
<td>1:44</td>
</tr>
<tr>
<td>10:20</td>
<td>2:52</td>
<td></td>
<td>10:32</td>
<td>1:48</td>
</tr>
<tr>
<td>10:25</td>
<td>2:53</td>
<td></td>
<td>10:38&lt;sup&gt;b&lt;/sup&gt;</td>
<td>2:04</td>
</tr>
<tr>
<td>10:26</td>
<td>3:13&lt;sup&gt;a&lt;/sup&gt;</td>
<td>10:45</td>
<td>2:10</td>
<td></td>
</tr>
<tr>
<td>10:30</td>
<td>3:20</td>
<td></td>
<td>10:49&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2:22</td>
</tr>
<tr>
<td>10:33</td>
<td>3:23</td>
<td></td>
<td>11:00</td>
<td>2:34</td>
</tr>
<tr>
<td>10:39&lt;sup&gt;a&lt;/sup&gt;</td>
<td>3:37&lt;sup&gt;b&lt;/sup&gt;</td>
<td>11:01</td>
<td>2:35</td>
<td></td>
</tr>
<tr>
<td>10:53</td>
<td>3:46</td>
<td></td>
<td>11:13</td>
<td>2:40</td>
</tr>
<tr>
<td>10:55</td>
<td>3:59</td>
<td></td>
<td>11:18</td>
<td>3:15</td>
</tr>
<tr>
<td>11:19</td>
<td></td>
<td></td>
<td>11:24</td>
<td>3:44</td>
</tr>
<tr>
<td>11:22</td>
<td></td>
<td></td>
<td>11:27</td>
<td>3:50</td>
</tr>
<tr>
<td>11:38</td>
<td></td>
<td></td>
<td>11:37&lt;sup&gt;a&lt;/sup&gt;</td>
<td>3:53</td>
</tr>
<tr>
<td>12:00</td>
<td></td>
<td></td>
<td>11:46</td>
<td>3:58</td>
</tr>
<tr>
<td>12:05</td>
<td></td>
<td></td>
<td>11:51</td>
<td>3:59</td>
</tr>
<tr>
<td>12:10</td>
<td></td>
<td></td>
<td>11:55</td>
<td></td>
</tr>
</tbody>
</table>

Note: a=significant at 10% level for q=5 and q=20; b=significant at 10% for q=20; and the rest denote significant at 10% level for q=5.
Table 2. In-sample and Out-sample U statistic Comparison: NYSE and NASDAQ

<table>
<thead>
<tr>
<th></th>
<th>NYSE</th>
<th>NASDAQ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>reject RW</td>
<td>can’t reject H₀</td>
</tr>
<tr>
<td>in-sample</td>
<td>T</td>
<td>44</td>
</tr>
<tr>
<td>Mean</td>
<td>0.6992</td>
<td>0.7105</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.8795</td>
<td>0.9657</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.5651</td>
<td>0.3857</td>
</tr>
<tr>
<td>Std dev.</td>
<td>0.0571</td>
<td>0.0685</td>
</tr>
<tr>
<td>out-sample</td>
<td>T</td>
<td>44</td>
</tr>
<tr>
<td>Mean</td>
<td>0.7376</td>
<td>0.7543</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.9633</td>
<td>0.9880</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.6019</td>
<td>0.6306</td>
</tr>
<tr>
<td>Std dev.</td>
<td>0.6760</td>
<td>0.6306</td>
</tr>
</tbody>
</table>

Note: T=sample size, std dev.=standard deviation. "Reject RW" refers to time intervals in which the random walk hypothesis is rejected in the context of the modified R/S statistic. Conversely "can't reject H₀" represents that when the null hypothesis cannot be rejected. In-sample denotes forecasting within the sample while out-sample is the forecast outside the sample.
Figure 1 Variation of Mean Stock Returns (1-minute Interval)
1992 NYSE

Figure 2 Variation of the Standard Deviation of Stock Returns (1-minute Interval)
1992 NYSE
Figure 3 Variation of Mean Stock Returns (1-minute Interval)  
1992 NASDAQ

Figure 4 Variation of the Standard Deviation of Stock Returns (1-minute Interval)  
1992 NASDAQ
Figure 5 Variation of the Modified R/S Statistics with 1-minute Interval (q=5)
1992 NYSE

Figure 6 Variation of the Modified R/S Statistics with 1-minute Interval (q=20)
1992 NYSE
Figure 7 Variation of the Modified R/S Statistics with 1-minute Interval (q=5)
1992 NASDAQ

Figure 8 Variation of the Modified R/S Statistics with 1-minute Interval (q=20)
1992 NASDAQ