Nilay Shah and David DeHorn
Stats For Finance
4/11/01

# Stock Returns and the Test of the Random Walk Hypothesis

**Objective:** The objective of the report was to test daily and monthly data from the Standard and Poor's 500 with the expectations of the random walk hypothesis and market efficiency.

**Data:** All tests were performed using 200 SP500 observations and Microsoft Excel. The monthly data ran from 7/84 to 2/01. The daily data ran from 6/21/00 to 4/5/01. The source of the data was **http://finance.yahoo.com**

**Background:** The idea of efficient markets inherently states that it is impossible to beat the market because prices already incorporate and reflect all relevant information. The random walk hypothesis is a variant of the efficient market idea and essentially states that stock prices follow a random walk pattern and thus historic prices are of no value when predicting future prices.

**Methodology:** After collecting the data, we standardized the data to run the various tests, plot the evolution graphs, run necessary regressions, and summarize the data. Again, all of this was done using Microsoft Excel 2000.

**Assumptions:** While there are various classifications for how a market can be classified in terms of market efficiency, the focus of our study will assume that markets are weak efficient, or that all past market prices and data are fully reflected in securities prices.

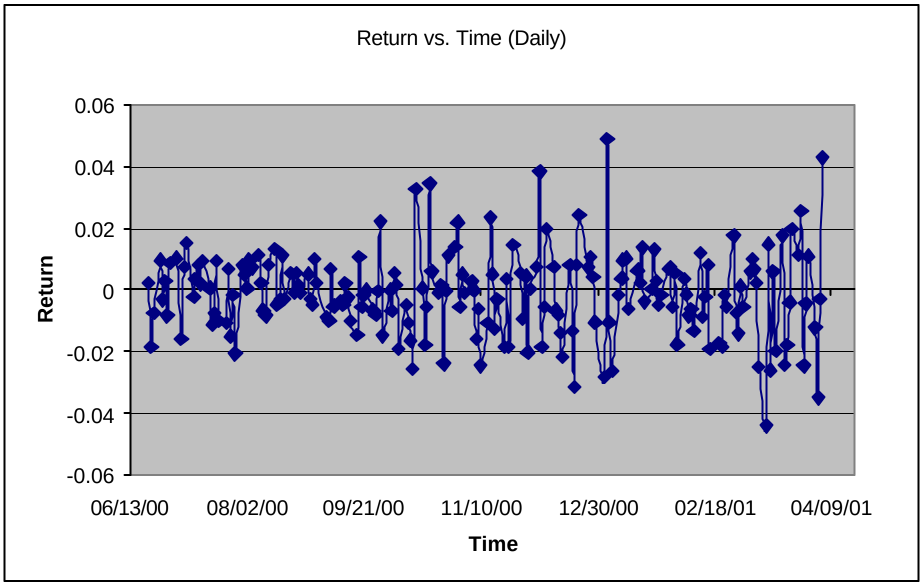
**Expectations of the Random Walk Hypothesis (RWH):**

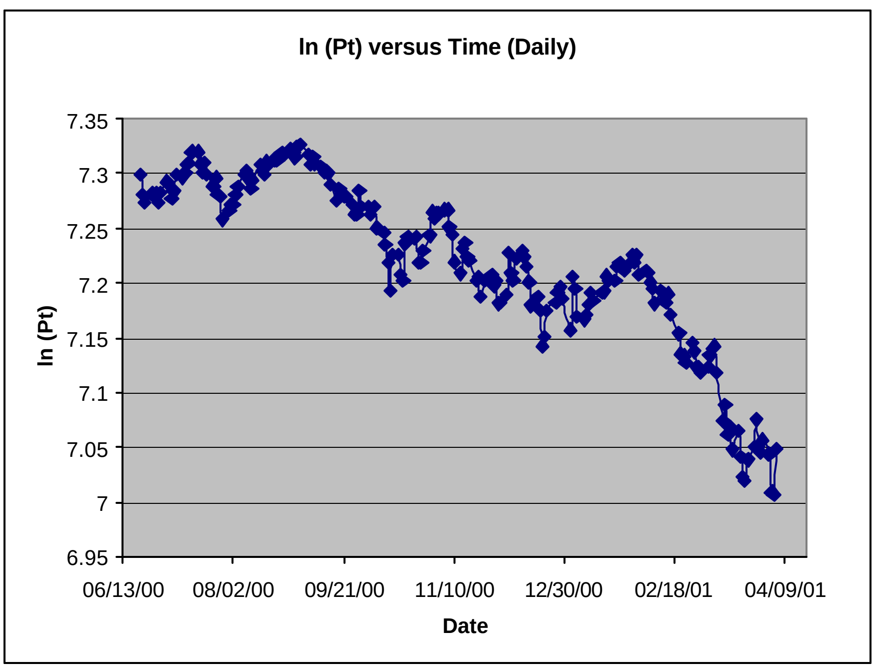
- The $E(R_{t-1,t}) = 0$ and $E(R_{t-1,t} \mid R_{t-2,t-1}) = 0$
- The $VAR(R_{t-1,t}) = s^2$ and $VAR(R_{t-1,t} \mid R_{t-2,t-1}) = s^2$
- The Skewness = 0
- The Kurtosis = 3
- Statistically satisfies the "Runs Test," which measures the sequence of returns
- The correlation coefficient when regressing ($R_{t-1,t}$ and $R_{t-2,t-1}$) is not statistically significant from 0.
- The variance ratio test is satisfied and thus the ratio $VAR(R_{t-2,t}) / 2\ VAR(R_{t-1,t})$ is not statistically significant from 1.

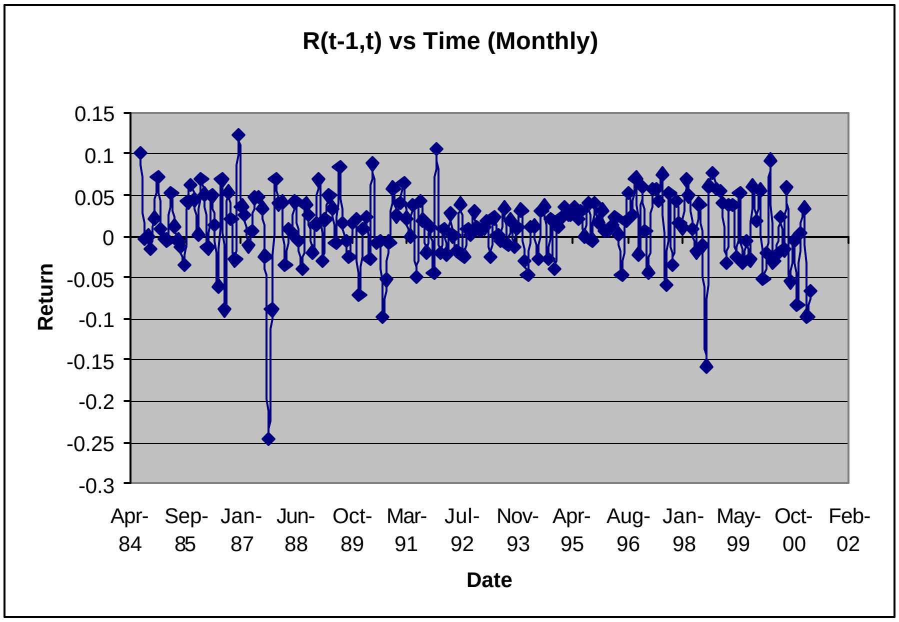
**Summary of Findings and Results:** Upon performing the various tests and analyzing the data for both monthly and daily SP500 data, we found the RWH to be a more than adequate method of formulizing the market structure. Just by glancing at the graphical plots of Return and ln Pt over time, it is clear that the hypothesis of hovering around a mean of 0 is true for this experiment. We have also found that the null hypotheses are satisfied for both monthly and daily data, and hence we can assume the RWH to hold true.

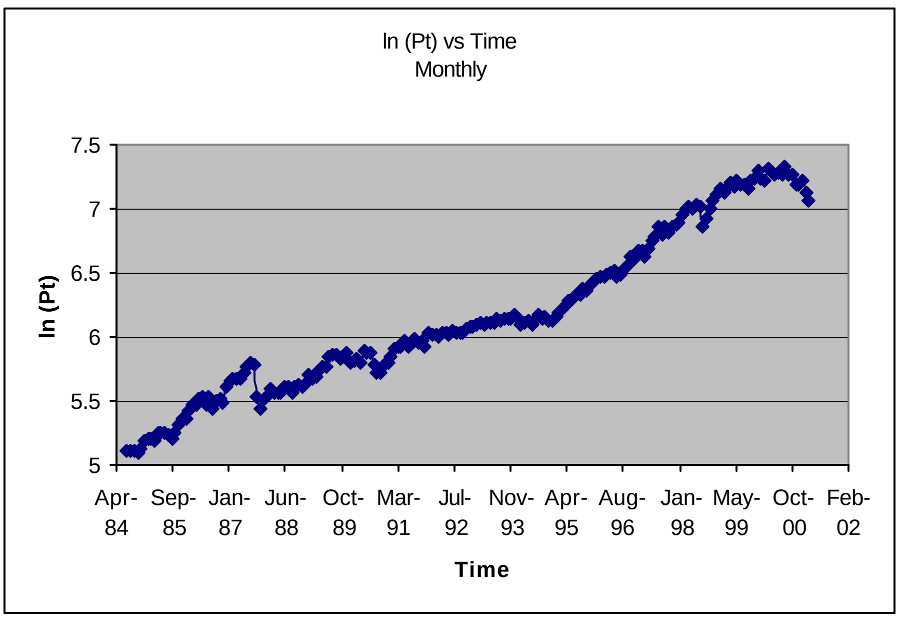
## Experimental Results of the SP500 Data:

1. The first task involved plotting the evolution of ln $P(_t)$ and $R(_{t-1,t})$ vs. Time. Recall that we would expect that ln Pt - $\ln(P_{t-1}) = (R_{t-1,t})$ and the RWH holds that $E(R_{t-1,t}) = 0$






One thing we do see in the graphs of the return is that there are a few significant outliers in the graph that have returns greater than 0. These are the shocks that can occur over a series. These shocks are derived from the following equation:

$$\mathbf{ln(P_t) = ln(P_{t-1}) + ln(e_t)}$$

where the Error Term = $\ln(e_t)$
and intuitively this error term = $(R_{t-1,t})$ with an expected value of 0.

Using the formula for returns over consecutive periods, we got the following data

| | Daily (6/21/00 – 4/5/01) | Monthly (7/84 – 2/01) |
|---|---|---|
| **$E(R_{t,t-1})$** | -.00125 or -.125% | .010532 or 1.0532 % |
| **Standard Deviation** | .01364 | .044678 |
| **Variance** | .0001860496 | .00199612 |
| **Test Statistic** | -.09164 | .2357 |

The test statistic is used as a measure of statistical significance against the null hypothesis. Its value is computed by using the following formula:

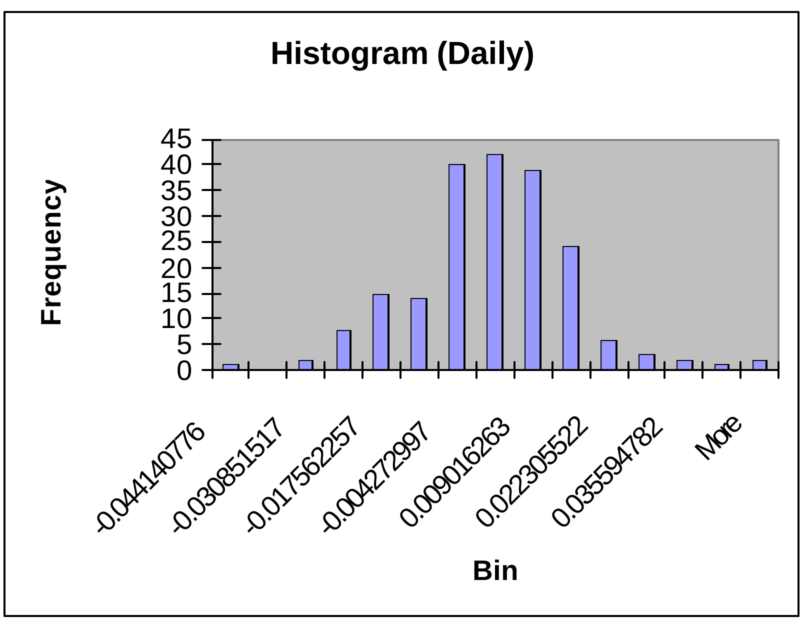
**t-stat = $(((R_{t-1,t}) - H_0) / se(R_{t-1,t}))$**

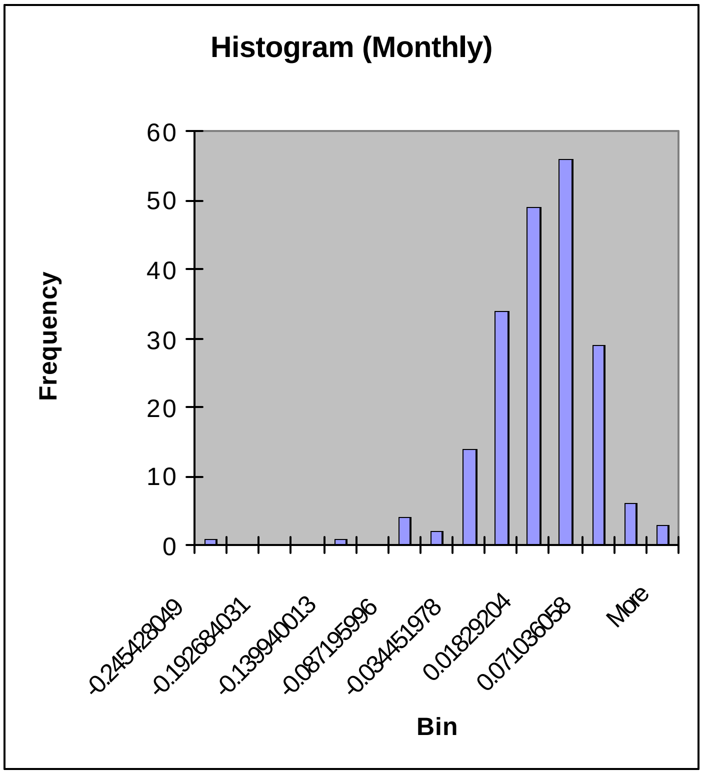
We know that the Null Hypothesis is 0, and thus we find two test statistics that are below 1.96 and greater than –1.96. Thus we accept the Null Hypothesis that $E(R_{t-1,t}) = 0$ for both the monthly and daily data.

What is interesting to observe is the different values of variance and standard deviation between the daily and monthly data of the experiment. The monthly data has a higher expected return, standard deviation, and variance than daily. The rationale for this is due to the intervals that both were measured under. When daily price changes are measured, there is necessarily a lesser chance of significant variance or fluctuation when compared to the amount of variance that can occur over an entire month's period. It is much more likely for shocks in the price to compound over a month's time rather than simply on a day-to-day basis. This accounts for the slight differences between the two series.

2. Another good way to measure the effectiveness of experimental data is to compare it to a normally distribution.

   We know that $(R_{t-1,t})$ is normally distributed with mean = 0 and variance = $\sigma^2$






- Based on the histograms, it clear that while both are relatively close to a normal distribution, they are not completely normal.

Under a normal distribution, we would expect the Kurtosis to be 3 and the Skewness to be 0. Kurtosis is a measure of the fatness of the tails. The skewness is a representation of the general direction that the frequency is skewed toward. In the first histogram (Daily) it seems that the data is slightly skewed to the right, but seems generally pretty normally distributed. The monthly histogram is a little different and has a skewness to the left due to a few outliers that make it less normally distributed.

To calculate Skewness and Kurtosis:

**Skewness(X) = E((X-mean)^3/$s^3$)**

**Kurtosis(X) = E((X-mean)^4/$s^4$)**

Experimental Results

| | **Daily** | **Monthly** |
|---|---|---|
| **Skewness** | .2505 | -1.27379 |
| **Kurtosis** | 1.36 | 5.500775 |

Recall that the E(Skewness) = 0 and E(Kurtosis) = 3 under the RWH. With the experimental skewnesses, it is clear that a the daily histogram is relatively close to 0, while the monthly data, as stated before is skewed to the left and thus has a negative skewness. The kurtosis of the daily values is 1.36, which could be said to be significantly less than 3. A skewness lower than 3implies that the data has too much concentration near the mean of the data and thus the tails are not "fat enough." Looking at the monthly kurtosis of 5.5, it is clear that the tails are fatter than one would expect with the null hypothesis and thus there is too much data at the ends rather than near the center.

In order to account for the differing skewness and kurtosis values for monthly and daily values, it's essential to recall the idea that the longer time series of monthly data can result in greater variation and the presence of random outliers that can skew data.

## 3. <u>Runs Test</u>

The runs test is another important element of the RWH. We know that $\ln(P_t) = \ln(P_{t-1}) + \ln(e_t)$.

The Runs Test says that testing the sequence of successive runs where

$(R_{t-1,t} > \text{mean})$ (defined by 1)
$(R_{t-1,t} < \text{mean})$ (defined by 0)

The RWH has that $P(R_{t-1,t} > \text{mean}) = P(R_{t-1,t} < \text{mean}) = .5$

We know that $E(\#\ \text{Runs}) = 2(n)(.5)(1-.5) + (.5)^2 + (1-.5)^2$

Experimental Results

| | **Daily** | **Monthly** |
|---|---|---|
| **E(Runs) (Based on Null)** | 100.5 | 100.5 |
| **Experimental Runs** | 99 | 108 |
| **Test Statistic** | -.142 | 1.1314 |

Looking at the test statistic values (both less than 1.96 and greater than –1.96), it's clear that we would accept the null hypothesis for the Runs Test and thus consider this consistent with the random walk hypothesis.

Consistent with the previous experiments, it seems that the monthly data tends to vary more from the null hypothesis than the daily data. While both pass the null test, it is clear that since the monthly series covers over 16 years, there is a possibility of greater variation, which accounts for more runs. Again, the daily series is simply covering the past six months, so the economy has not changed nearly as much as it did in the monthly analysis.

## 4. <u>Test of Serial Correlation</u>

As stated in the definitions of the RWH, an analysis of past prices and returns cannot be used to predict future prices. Essentially, the returns of all previous periods should be independent of the current period. A general way of testing for independence is measuring how two returns are correlated.

According to the null hypothesis, the correlation between suppose $R(_{t-1,t})$ and $R(_{t-2,t-1})$ should be 0. In theory, no matter how much we know about the expected returns from the past 50 periods, trying to use this to predict future returns should be futile.

We analyzed the regression $R(_{t-1,t}) = a + b\ R(_{t-2,t-1}) + v(t)$ (where $v(t)$ = error term with $E(v(t)) = 0$). Under the null hypothesis, b = correlation coefficient = 0.

**Experimental Results**

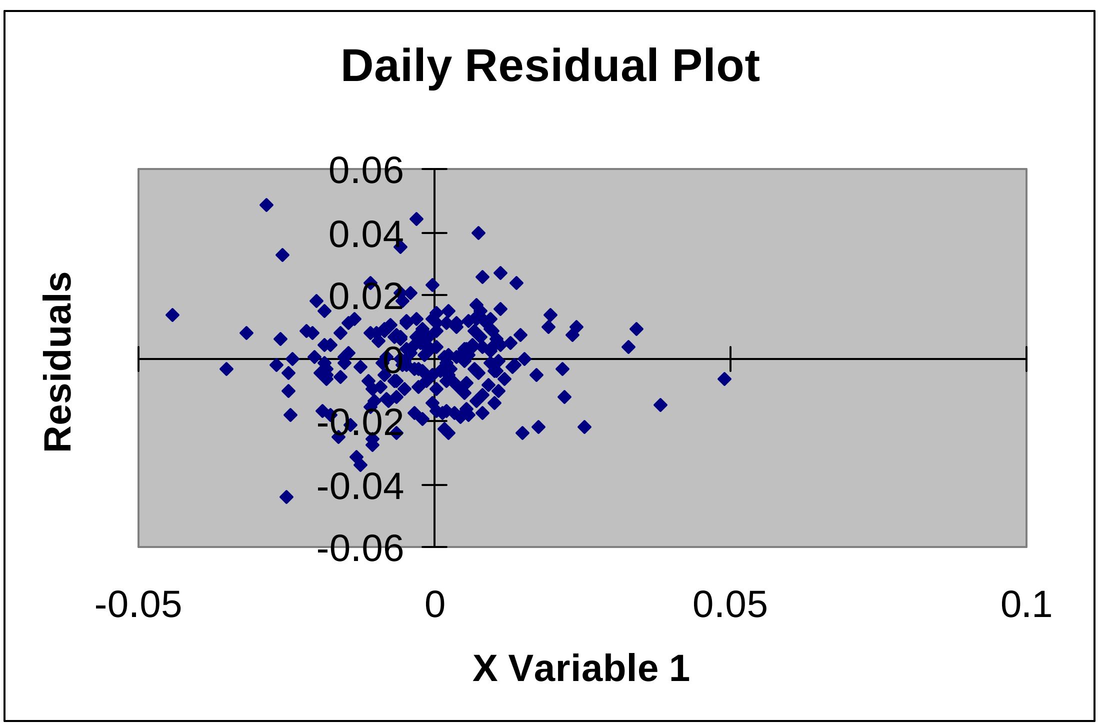
| | Daily | Monthly |
|---|---|---|
| b (correlation coefficient) | -.05 | -.04563 |
| a (intercept) | -.00134 | .011157 |
| Test statistic | -.6849 | -.6324 |

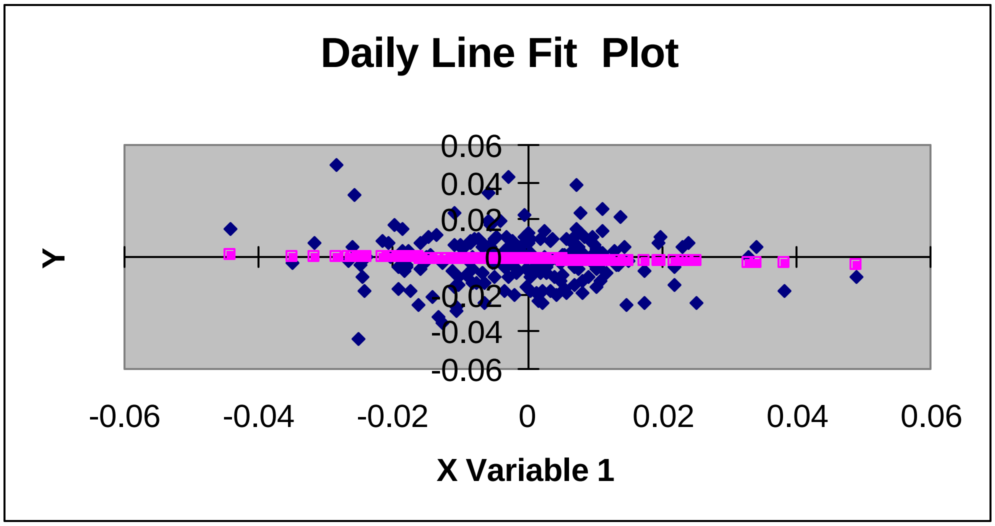
Based on these results, we see that the test statistic satisfies the conditions of the null hypothesis and thus we accept this test for the RWH.

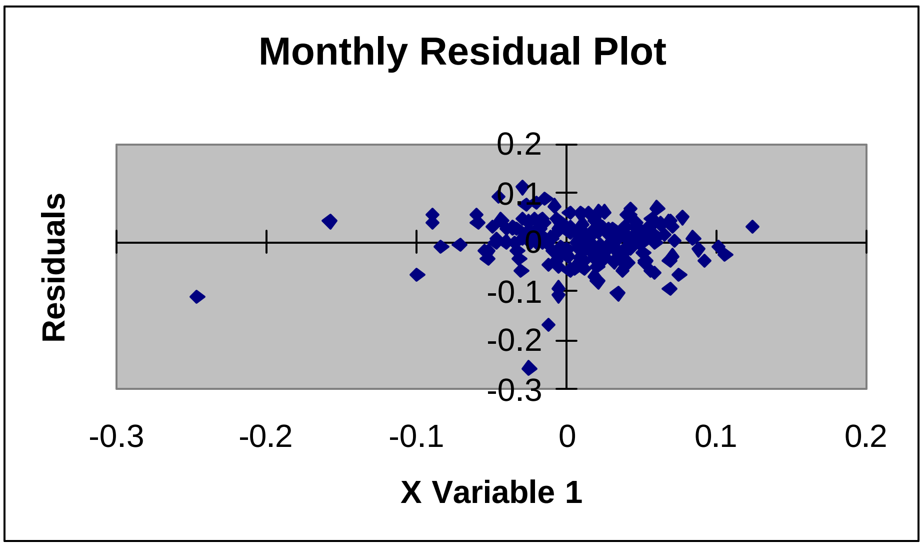
Daily Regression Model: $R(_{t-1,t}) = -.00134 + (-.05)(R(_{t-2,t-1}))$

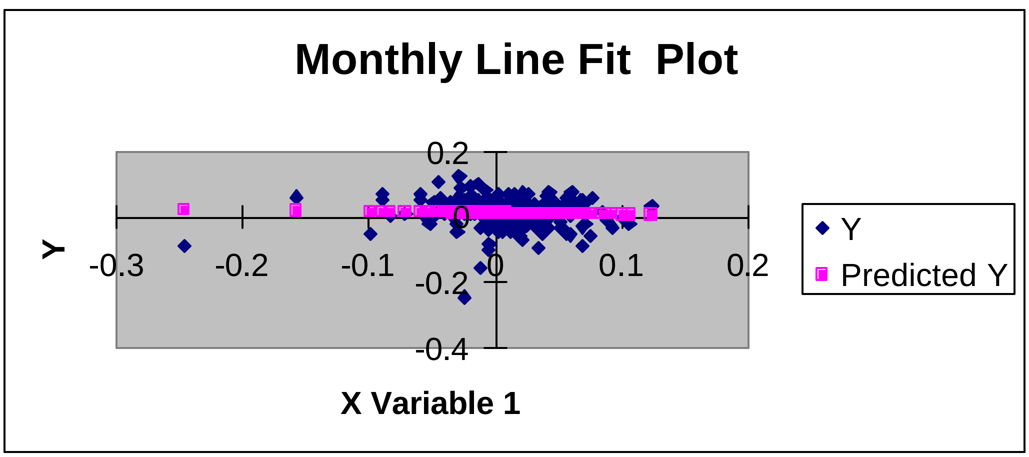
Monthly Regression Model: $R(_{t-1,t}) = .011157 + (-.04563)(\ R(_{t-2,t-1}))$

Its key to notice that intercepts of both regression models correspond to the mean experimental returns stated above.




Daily Line Fit Plot
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
Y
-0.06
-0.04
-0.02
0
0.02
0.04
0.06
X Variable 1


Monthly Residual Plot
0.2
0.1
0
-0.1
-0.2
-0.3
Residuals
-0.3
-0.2
-0.1
0
0.1
0.2
X Variable 1


Monthly Line Fit Plot
0.2
0
-0.2
-0.4
Y
-0.3
-0.2
-0.1
0
0.1
0.2
X Variable 1
Y
Predicted Y

A good way of testing this correlation test is simply by looking at the residual and line fit plots. Aside from a few outliers that are inevitable in most statistical experiments, it is clear that the residuals are clustered around 0, rather than floating around. The line plots reiterate that the null hypothesis is true in this case and that past prices do virtually nothing in predicting future results.

**5. Variance Ratio Tests**

It is rather intuitive to understand that: $\mathbf{R(_{t-2,t}) = R(_{t-1,t}) + R(_{t-2,t-1})}$

With this in mind, we from stats classes that:

$$\mathbf{VAR\ [R(_{t-2,t})] = VAR[R(_{t-1,t})] + VAR[R(_{t-2,t-1})] + 2COV[R(_{t-1,t}),R(_{t-2,t-1})]}$$

Under the null hypothesis of the RWH, future returns are entirely independent of past returns and thus the covariance of $R(_{t-1,t})$ and $R(_{t-2,t-1})$ should be 0.

Hence under null of RWH we should get that

$$\mathbf{VAR\ [R(_{t-2,t})] = VAR[R(_{t-1,t})] + VAR[R(_{t-2,t-1})]}$$

$$\mathbf{(VAR\ [R(_{t-2,t})] / 2VAR[R(_{t-1,t})]\ ) = 1 + ?\ \ (we\ know\ that\ E(?) = 0)}$$

**Experimental Results**

| | **Daily** | **Monthly** |
|---|---|---|
| **VAR $R(_{t-1,t})$** | .000186 | .001996 |
| **VAR $R(_{t-2,t-1})$** | .000178 | .001947 |
| **$VAR[R(_{t-1,t})] + VAR[R(_{t-2,t-1})]$** | .000364 | .003943 |
| **$VAR\ [R(_{t-2,t})]$** | .000358 | .003629 |
| **Ratio** | .982197 | .920326 |
| **Test Statistic** | -.25177 | -1.1267 |

The ratio test has produced two test statistics that are consistent with the null hypothesis, so we accept the Variance Ratio Test for both the daily and monthly data.

**Observations**: Although many brokers and analysts will try to convince you that significant technical analysis on certain companies might lead to the ability to predict and garner gains in the future, our results seem to predict otherwise. In summary, there was no compelling evidence to refute the null hypothesis that the returns follow a random walk. It seems to be quite convincing that no matter what we know about yesterday, stock returns will necessarily move randomly and hover around a mean of 0. One interesting observation lies in the results of monthly and daily data. The monthly data covered a much larger range and thus accounted for the higher standard deviation and variance when compared to the daily data

Along with experimentally proving the basis of the RWH, we begin to question the validity and temptation of investing the market. At first glance at the results, it would seem that beating the market is next to impossible, but then again, does that mean that all the money spent on analysis, research, and consulting simply goes to waste? This continues to be a burning question.  On one hand there are RWH supporters who point to the current bearish market and fallout that has turned high gains from recent years into 0 profits.  On the other hand there are critics and financial analysts who cite that time will bring success and that the recent losses are simply temporary.