ABSTRACT

Two overlooked points for the suggested AR(1) model in the research of Brock, Lakonishok, and LeBaron (1992) are discussed in this paper. First, Durbin’s h statistic shows that the error terms from the AR(1) process for the returns of DJIA are correlated. Secondly, the coefficient of determination is low at 0.2% for the returns from DJIA. In other words, the model suggested can only explain 0.2% of the variation of DJIA. Therefore, the AR(1) model is not an ideal representative for the returns from the price index. In fact, our study shows that the DJIA price index itself has a random walk like process with uncorrelated error terms but not necessarily i.i.d. This study concludes the market is not predictable.

KEYWORDS: AR(1) Model, DJIA

INTRODUCTION

In Brock, Lakonishok, and LeBaron (hereafter abbreviated as BLL, 1992), the AR(1) model, \( r_t = \mu + \phi r_{t-1} + \epsilon_t \), is used to describe the relationship of returns for the Dow Jones Industry Average (DJIA) from the first trading day in 1897 to the last trading day in 1986. From their study, \( \mu \) is estimated to be 0.00015 with a T-value of 2.50; this implies that at \( \alpha = 0.05 \), \( \mu \) is significantly different from 0 (see the results listed in Table V, P. 1747, BLL, 1992). The “abnormal” returns from technical trading rules may be due to positive autocorrelation and \( \mu \neq 0 \), (for instance, the simple moving average and trading range break-out rules) (See P. 1745-1746, BLL, 1992). To support their own claim of having an AR(1) model for the returns, BLL (1992) quoted Conrad and Kaul (1990)'s research result that “a first-order autocorrelation of 0.2 for a value-weighted portfolio of the largest companies during the period 1962-1985….. higher order autocorrelation, beyond a lag of one day, is essentially zero.”

Due to the age of the noteworthy amount of data used in BLL, in our study 9,000 DJIA data points from 4/21/1971 to 10/18/2005 are used. Because the lag of the dependent variable is used as the regressor in the AR(1) model, by using the Durbin’s h-statistic in our study, autocorrelations exist among the error terms. Therefore, the T-test should not be used to test the null hypothesis \( H_0: \mu = 0 \) vs the alternative hypothesis \( H_a: \mu \neq 0 \) in the regression model \( r_t = \mu + \phi r_{t-1} + \epsilon_t \) as suggested on P. 1745 in BLL, 1992, where \( \epsilon_t \) is i.i.d. or a white noise. We also address the coefficient of determination, \( R^2 \), which measures the goodness of regression fit, in this paper, which was not presented in BLL, 1992.
The overall average for the R^2's for the returns from DJIA is 0.0020 in our study, indicating that the AR(1) or the Lag 1 regressive model, r_t=μ+ϕ_{t-1}+ε_t, suggested by BLL can only explain a very small amount of variations of returns r_t. Therefore, the AR(1) or the Lag 1 regressive model as suggested by BLL, 1992 should not be used to interpret for "abnormal" returns. There may exist some other factors affect the outcomes of returns in the market.

In the end, when the AR(1) or the autoregressive model Y_t=μ+ϕ_{t-1}+ε_t is applied to the DJIA itself, Durbin's h-statistic shows that no autocorrelation exist among error terms. And at α=0.01, we failed to reject H_0: μ=0, and H_0: ϕ=1. In other words, DJIA follows a random walk process. The profit gain cannot be predictable and the market is efficient.

DATA ANALYSIS

1. Durbin’s h-statistic

   The conventional Durbin-Watson test is not applicable when the regressor contains a lagged dependent variable as in the AR(1) or the Lag 1 regressive model. In Durbin (1970), the h-statistic is defined as $h = \hat{\rho} \sqrt{n} \left[ 1 - n(s_{b_1})^2 \right]$, where $\hat{\rho}$ is the sample autocorrelation between Y_t and Y_{t-1}, n is the sample size, and $s_{b_1}$ is the standard error for the regressor Y_{t-1}. When n is large, we have $\hat{\rho} \approx 1 - \frac{DW}{2}$, where $DW = \sum_{t=2}^{n} (e_t - e_{t-1})^2 / \sum_{t=1}^{n} e_t$ is the conventional Durbin-Watson statistic (Durbin and Watson, 1950 and 1951). When n gets larger and under the null hypothesis $H_0$ that autocorrelation $\rho=0$, Durbin’s h statistic will follow a standard normal distribution (Durbin, 1971).

   Let Y_t be the price index at time t. The return r_t is defined as the difference of the logarithm of the price indices, $\ln(Y_t) - \ln(Y_{t-1}) = \ln(\frac{Y_t}{Y_{t-1}})$. Applying simple (auto)regression analysis on the logarithm of the price indices at Lag 1 shows that all DJIA price indices follow the random-walk-like model $\ln(Y_t) = \ln(Y_{t-1}) + \omega_t$, or $r_t=\omega_t$, but $\omega_t$ does not follow an i.i.d. or a white noise process. From Durbin’s h test, $h^* = 3.4818$ with p-value of 0.0005, positive autocorrelation exists among the error terms. Accordingly, the T-test cannot be used to test the null hypothesis $H_0$: $\mu = 0$ vs the alternative hypothesis $H_a$: $\mu \neq 0$ in the regression model $r_t=\mu+\phi_{t-1}+\epsilon_t$, as suggested on P. 1745 in BLL, 1992.

2. The coefficient of determination, R^2

   While running the regression analysis of returns $r_t=\beta_0+\beta_1 r_{t-1} + \epsilon_t$, the coefficient of determination, R^2 comes to be 0.002. In other words, only 0.2% variations from $r_t$ can be explained by the above AR(1) or the Lag 1 regressive model. So, the AR(1) or the Lag 1 regressive model is not a good representative for the returns from the DJIA index.

3. Uncorrelated but not necessarily equal variance for residuals from regression on Y_t
Interestingly, when we analyze the DJIA price index itself, the regression analysis results for $Y_t = \beta_0 + \beta_1 Y_{t-1} + \epsilon_t$ are listed in Table 1. The Durbin’s h-statistic, $h^* = -0.4428$ with a $p$-value of 0.6578. This indicates that no autocorrelations among residuals. The coefficient of determination $R^2$ for normality plot of residual is 0.8266. This indicates the residuals follow a uncorrelated normal probability distribution.

Table 1: Results from regression analysis $Y_t = \beta_0 + \beta_1 Y_{t-1} + \epsilon_t$

<table>
<thead>
<tr>
<th></th>
<th>$n$</th>
<th>$H_0: \beta_0=0$ (p-value)</th>
<th>$H_0: \beta_1=1$ (p-value)</th>
<th>$R^2$ for regression</th>
<th>$H_0: \mu_\epsilon=0$ (p-value)</th>
<th>$s_\epsilon$</th>
<th>$R^2$ for Normality Plot</th>
<th>$H_0: \rho=0$ Durbin’s h (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DJIA</td>
<td>8,999</td>
<td>0.9688 (0.2658)</td>
<td>1.0000 (0.1090)</td>
<td>0.9998</td>
<td>0.0000 (0.0000)</td>
<td>N/A</td>
<td>N/A</td>
<td>0.8266 (0.6578)</td>
</tr>
</tbody>
</table>

* Significant at $\alpha = 0.05$

Figure 1: Residual plot for the regression analysis $Y_t = \beta_0 + \beta_1 Y_{t-1} + \epsilon_t$, where $\epsilon_t$ is a white noise

From the residual plot shown in Figure 1, the variance seemed increased after time period 5,000. In other words, heteroskedasticity might occur. Instead of using the visual inspection, the Breusch-Pagan test (1979) is applied to check for heteroskedasticity. The Breusch-Pagan test follows the simple three-step procedure:

Step 1: Regress $Y_t$ on $Y_{t-1}$ by Ordinary Least Squares (OLS) and obtain the residuals $e_t$;
Step 2: Regress $e_t^2$ on $e_{t-1}^2$;
Step 3: The Breusch-Pagan’s test statistic is the result of the coefficient of determination, $R^2$, from the regression analysis of $e_t^2$ on $e_{t-1}^2$ in Step 2 time the sample size $n$. That is $BP = nR^2$, which follows a chi-square distribution with degrees of freedom 1 in our case. When we reject the null hypothesis $H_0$: homoscedasticity, then conditional heteroskedasticity exists.

In our study, we find $BP = (8,999)(0.103961) = 935.55$; therefore, we reject $H_0$ and conclude that there is conditional heteroskedasticity.

4. Robust standard errors
The existence of heteroscedasticity will cause the estimate of standard errors to be biased. It will underestimate the significance for the test on \( \beta \)'s in the regression analysis due to inflated standard errors. To overcome the heteroscedasticity problem, we can either apply the robust standard errors (White, 1980) to find the new calculated T-test statistics, or to rerun the regression analysis by the weighted least squares method. The use of robust standard errors will not change the coefficient estimates obtained from OLS. But it will change the standard errors and provide more accurate p-values for the significance test for \( \beta \)'s. On the other hand, the general least squares (GLS) method minimizes \( \sum \frac{(Y_t - \hat{Y}_t)^2}{\text{Var}(e_t)} \), instead of minimizing the sum of squared residuals, \( \sum (Y_t - \hat{Y}_t)^2 \), in OLS. The parameter estimation for GLS is complex (Nelder and Wedderburn, 1972). But under the special condition that the variances of error terms vary according to one independent variable (says \( \sigma_i = wY_{t-1} \)), the parameter estimation for GLS can be obtained from the weighted least squares (WLS) method by transforming the original data via a diagonal matrix with reciprocal of square root of the weights (i.e., \( \frac{1}{\sqrt{w_i}} \)).

For WLS, we need to ensure we are using the correct weights; if the weights are wrong, then we will make an incorrect estimation to the parameters and the accordingly significant T-test for \( \beta \)'s. Therefore, in this study, the robust standard errors are used. The robust standard errors can be easily found from EXCEL add-in OLSRegression.xla. The result from OLSRegression for the robust standard errors is listed in Table 2. As shown, the estimates for regression coefficients are not changed. However, the p-values for the significant test for \( \beta \)'s have been changed. Fortunately, the random walk like model \( Y_t = Y_{t-1} + \epsilon_t \) where \( \epsilon_t \)'s are uncorrelated but may not all have equal variance. Since the errors are uncorrelated, the market is unpredictable.

Table 2: Results from regression analysis \( Y_t = \beta_0 + \beta_1 Y_{t-1} + \epsilon_t \) with robust standard errors

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>SE</th>
<th>Robust Standard Error (HC3)</th>
<th>( H_0: \beta_0 = 0 )</th>
<th>( H_0: \beta_1 = 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.9688</td>
<td>0.8705</td>
<td>0.4948</td>
<td>1.9577 (p-value)</td>
<td>0.0710 (p-value)</td>
</tr>
<tr>
<td>( Y_{t-1} )</td>
<td>1.0000</td>
<td>0.0001</td>
<td>0.0003</td>
<td>(0.0503)</td>
<td>(0.9434)</td>
</tr>
</tbody>
</table>

* Significant at \( \alpha = 0.05 \)

5. Homoskedasticity and random walk for the last 500-day \( Y_t \)'s

From Figure 1, the heteroskedasticity of residuals might be caused due to the long period of 9,000 days. To eliminate the time-effect factor, we study the last 500 day DJIA daily prices. The regression analysis results are listed in Table 3. From it, we see the regression model for the DJIA is \( Y_t = \mu + Y_{t-1} + \epsilon_t \) where \( \epsilon_t \)'s are uncorrelated errors.

The normality plot in Figure 2 shows that \( \epsilon_t \)'s follow a normal distribution.

The residual plot in Figure 2 indicates the assumption for constant variance for error terms (i.e., the homoscedasticity) is also true. To further prove it, the Breusch-Pagan test is used
to check the heteroskedasticity. When we regress $e_t^2$'s vs $e_{t-1}^2$'s where $e_t$'s are residuals from regressing $Y_t$'s vs $Y_{t-1}$'s in OLS, we obtain $R^2=0.0012$, and $n=498$; therefore the $BP=(498)(0.0012)=0.6017$ with a $p$-value=0.4379 from the Chi-squared distribution of 1 degree of freedom; we fail to reject that the assumption of homoskedasticity is true. From the above discussion, we know $e_t$ are i.i.d.'s. So, the T-test can be used to test $H_0$:

$$\mu_e=0.$$  With $\bar{\varepsilon} = 0.0000$ and standard error $s_{\varepsilon}=2.9948$, from $t^* = \frac{\bar{\varepsilon}}{s_{\varepsilon}} = 0.0000$, we can claim that $E(\varepsilon)=0$ and the T-tests for testing the null hypotheses $H_0: \beta_0=0$ and $H_0: \beta_1=1$, respectively, in Table 3 are valid too. This confirms that DJIA follows a random walk $Y_t=Y_{t-1}+\nu_t$ where $\nu_t$'s are i.i.d where $E(\nu_t)$ may not be equal to 0. The market is efficient and unpredictable.

Table 3: Results from regression analysis $Y_t=\beta_0+\beta_1 Y_{t-1}+\varepsilon_t$, with robust standard errors

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>$H_0: \beta_0=0$ ($p$-value)</th>
<th>$H_0: \beta_1=1$ ($p$-value)</th>
<th>$R^2$ for regression</th>
<th>$H_0: \mu_\varepsilon=0$ ($p$-value)</th>
<th>$s_{\varepsilon}$</th>
<th>$R^2$ for Normality Plot</th>
<th>$H_0: \rho=0$ Durbin’s h ($p$-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DJIA</td>
<td>500</td>
<td>449.129 (0.002)</td>
<td>0.9568 (1.000)</td>
<td>0.9314</td>
<td>0.0000 (1.0000)</td>
<td>2.9948</td>
<td>0.9518</td>
<td>-0.1059 (0.9157)</td>
</tr>
</tbody>
</table>

* Significant at $\alpha = 0.05$

Figure 2: The normality of error term $\varepsilon_t$'s

![Normal Probability Plot](image)
CONCLUSIONS

In this study, we claim that the return, $r_t$, should not follow an AR(1) model $r_t = \mu + \phi r_{t-1} + \varepsilon_t$, where $\varepsilon_t$ is a white noise as suggested by BLL in 1992. The coefficient of determination, $R^2$, for the above regression model is only 0.2%; therefore, the AR(1) is not a good representative of DJIA returns. From our analysis, the DJIA index itself follows a random walk like process, $Y_t = Y_{t-1} + \varepsilon_t$, where $\varepsilon_t$ is uncorrelated but heteroscedastic. However, the effect of heteroscedasticity can be eliminated if we only look at a short-term DJIA, for instance, the last 500 daily DJIA in our study. This implies that DJIA at that time period follows a random walk process. Therefore, the market is efficient and unpredictable. It will not be possible to obtain a gain from the market as suspected by BLL.

REFERENCES