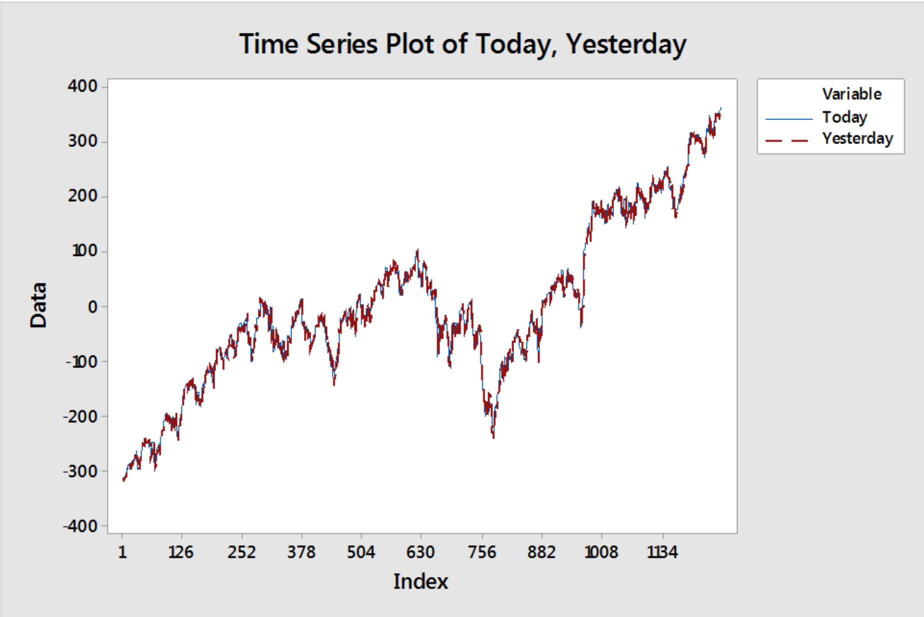
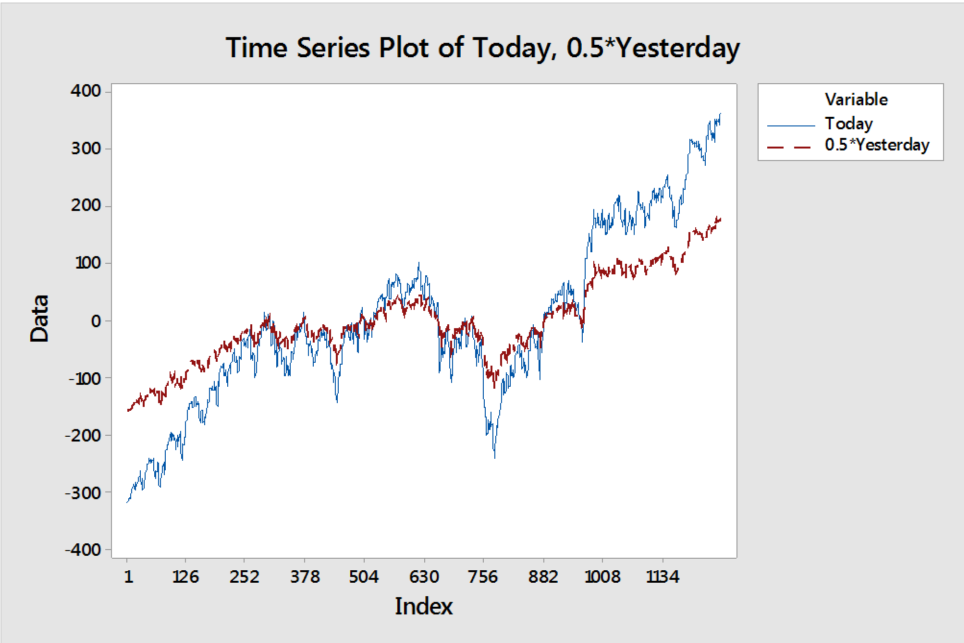
Forecasting Time Series Data HW2 **Tianze Lin tl2718**

Question 1

A)





B) Based on these two plots, Yesterday's Russell is a better forecast of Today's Russell because these two time-series (Today's Russell and Yesterday's Russell) follow more closely to each other and almost coincide.

C)

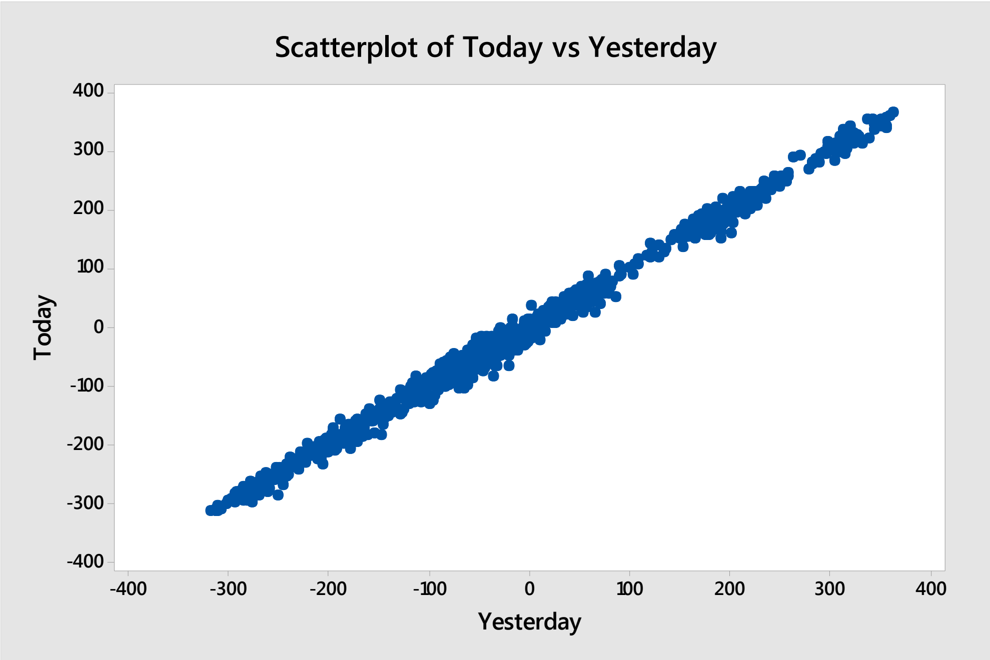
C10	C11
MSFE Yesterday	MSFE 0.5*Yesterday
132.961	362414

Using the formula MSFE = $(\Sigma(\text{forecast-actual})^2)/(n-1)$, we get the above results.

From above table, we see Yesterday and 0.5*Yesterday have a smaller mean squared forecast errors, which means Yesterday's Russell is a better model for forecast.

Question 2

A)



From the plot, we can see a linear relationship between Today and Yesterday, and the correlation should be positive.

B)

Regression Analysis: Today versus Yesterday

Method

Rows unused 1

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	28241141	28241141	212573.82	0.000
Yesterday	1	28241141	28241141	212573.82	0.000
Error	1256	166864	133		
Lack-of-Fit	1234	163466	132	0.86	0.730
Pure Error	22	3398	154		
Total	1257	28408005			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
11.5262	99.41%	99.41%	99.41%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.544	0.325	1.67	0.094	
Yesterday	0.99762	0.00216	461.06	0.000	1.00

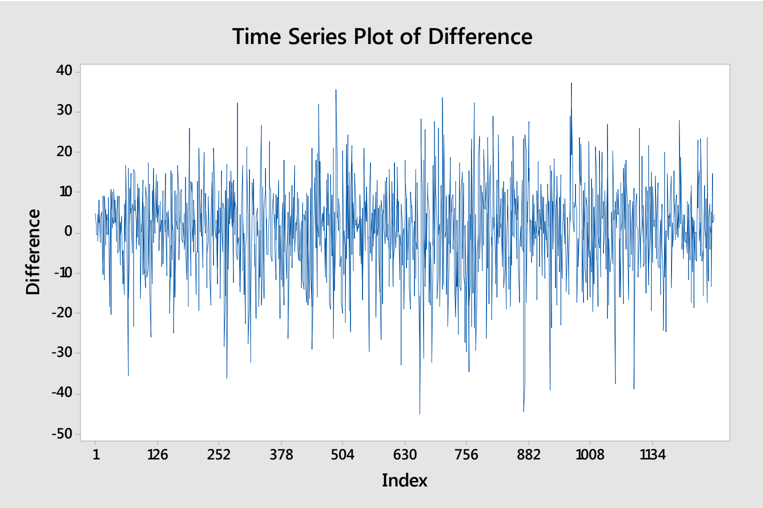
The coefficients of regression model imply that Today's Russell is expected to increase 0.9976 when Yesterday's Russell increases by one, holding all the other independent variables constant. The R^2 is 99.41%, which means 99.41% of variability of our data can be explained by our regression model, so the data fits the model very well. The p-value of coefficient is approximately 0, so we have a strong evidence that the slope coefficient is not equal to zero. Therefore, this regression model seems good, and it is consistent with my answer to 1b.

C)

No, the slope is not significantly different from 1. The intercept is 0.554, and this is the expected mean value of Today's Russell when Yesterday's Russell = 0. The p-value is 0.094, which is greater than 0.05, so there's no strong evidence that the intercept is different from zero at the significance level of 0.05.

D)

Based on everything you have done so far, I don't see any strong evidence that Russell is not a random walk. We can also verify it by differencing. The difference looks like white noise.

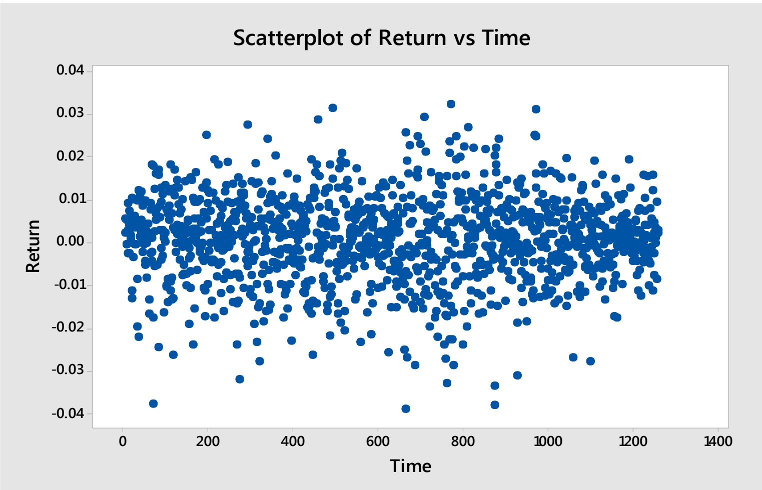


E)

r = sqrt (99.41%) = 0.9970 The correlation is very close to 1, so there is an almost perfect linear association between Today's Russell and Yesterday's Russell.

Question 3

A)



Choose α=0.05. Null hypothesis: mean = 0.

One-Sample T: Return

Descriptive Statistics

N	Mean	StDev	SE Mean	95% CI for μ
1258	0.000509	0.009911	0.000279	(-0.000039, 0.001057)

μ: mean of Return

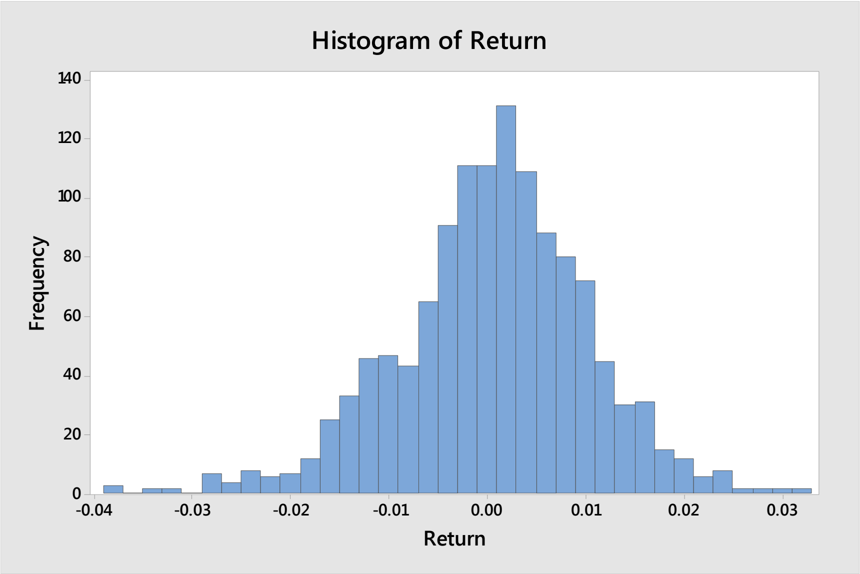
Test

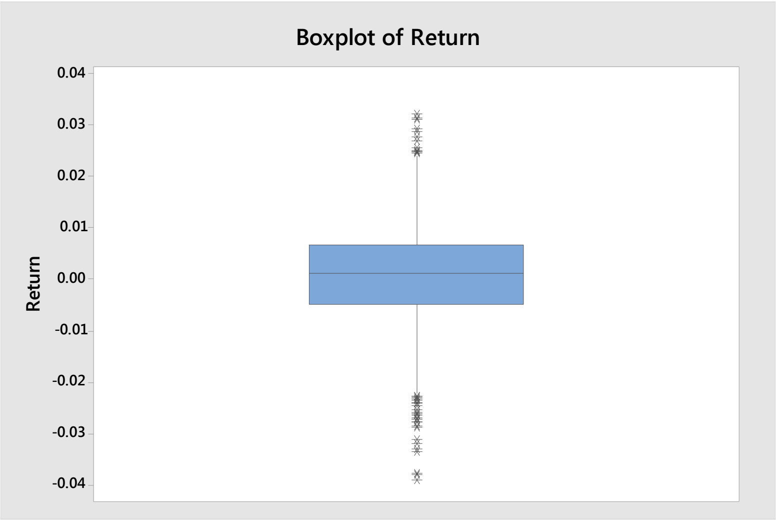
Null hypothesis	H_0: μ = 0
Alternative hypothesis	H_1: μ ≠ 0

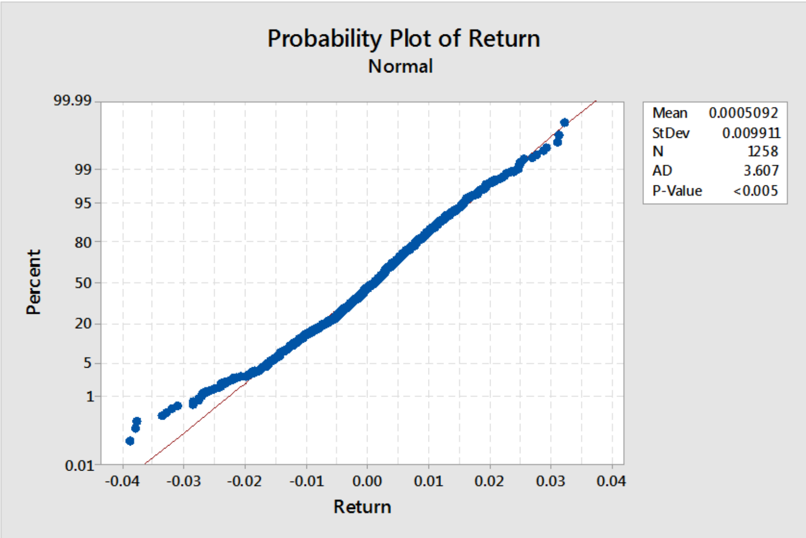
T-Value	P-Value
1.82	0.069

The mean and standard deviation are respectively 0.000509 and 0.009911. The p-value (0.069) is greater than α, so at the significance level of 0.05, we cannot reject the null hypothesis. There's no strong evidence that the mean of return is significantly different from zero.

B)

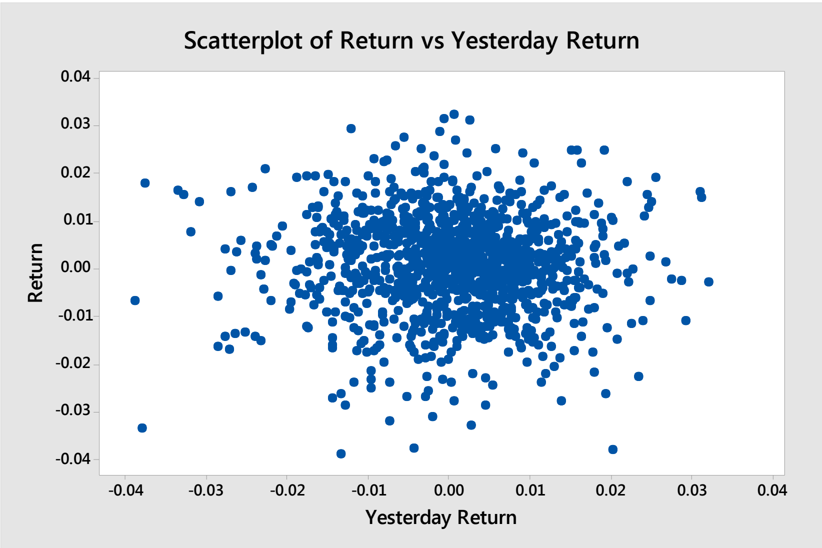






From the histogram and the probability plot, we can see the data is left-skewed. Choose α=0.05. From our normality test, we see the p-value is smaller than α, so we reject the null hypothesis. Therefore, the data do not follow a normal distribution.

C)



This plot is very different from the plot in 2a. Today's Russell is much easier to predict than Today's Return.

D)

Regression Analysis: Return versus Yesterday Return

Method

Rows unused 2

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	0.000101	0.000101	1.03	0.310
Yesterday Return	1	0.000101	0.000101	1.03	0.310
Error	1255	0.123350	0.000098		
Total	1256	0.123451			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.0099140	0.08%	0.00%	0.00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.000520	0.000280	1.86	0.064	
Yesterday Return	-0.0287	0.0282	-1.02	0.310	1.00

Choose α=0.05. The slope coefficient of regression model implies that Today's Return is expected to decrease 0.0287 when Yesterday's return increases by one, holding all the other independent variables constant. The intercept is 0.00052, and this is the expected mean value of Today's Return when Yesterday's Return = 0. However, the p-values of both coefficients are both greater than 0.05, so we fail to reject null hypotheses of both coefficients. Therefore, there is no strong evidence that the slope coefficient or the intercept are significantly different from zero.

Question 4

proof:

$$MSE = E[Y-\hat{Y}]^2 = E[Y-(a+bX)]^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - a - bX_{i-1})^2$$

$$\frac{d}{da}MSE = \frac{d}{da}\left[\frac{1}{n}\sum_{i=1}^{n}(Y_i - a - bX_{i-1})^2\right] = \frac{1}{n}\sum_{i=1}^{n}\frac{d}{da}(Y_i - a - bX_{i-1})^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}2(Y_i - a - bX_{i-1})(-1) = \frac{2}{n}\sum_{i=1}^{n}(a + bX_{i-1} - Y_i)$$

$$\frac{d}{db}MSE = \frac{d}{db}\left[\frac{1}{n}\sum_{i=1}^{n}(Y_i - a - bX_{i-1})^2\right] = \frac{1}{n}\sum_{i=1}^{n}\frac{d}{db}(Y_i - a - bX_{i-1})^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}2(Y_i - a - bX_{i-1})(-X_{i-1}) = \frac{2}{n}\sum_{i=1}^{n}(a + bX_{i-1} - Y_i)(X_{i-1})$$

$$① \ \frac{d}{da}MSE = 2\cdot\frac{1}{n}\sum_{i=1}^{n}(a + bX_{i-1} - Y_i) = 2a + 2bE(X) - 2E(Y)$$

$$② \ \frac{d}{db}MSE = \frac{2}{n}\sum_{i=1}^{n}(aX_{i-1} + bX_{i-1}^2 - Y_iX_{i-1}) = 2aE(X) + 2bE(X^2) - 2E(XY)$$

$$= 2aE(X) + 2b[Var(X) + (E(X))^2] - 2[Cov(X,Y) + E(X)E(Y)]$$

$$= 2aE(X) + 2bVar(X) + 2bE(X)^2 - 2Cov(X,Y) - 2E(X)E(Y)$$

To find minimum MSE, we need $① = 0 \Rightarrow a = E(Y) - bE(X)$

$E[X_t] = 0$ and X_t is stationary, so $E(Y) = E(X) = 0 \Rightarrow a = 0$.

We need $② = 0. \Rightarrow 2aE(X) + 2bVar(X) + 2bE(X)^2 - 2Cov(X,Y)$

$-2E(X)E(Y) = 0 \Rightarrow 2bVar(X) + 2bE(X)E(Y) - 2Cov(X,Y)$

$-2E(X)E(Y) = 0 \Rightarrow bVar(X) - Cov(X,Y) = 0 \Rightarrow b = \frac{Cov(X,Y)}{Var(X)}$

↳ because $E(X) = E(Y) = 0$.

X_t is stationary, so $Var(X) = Var(Y)$.

$$\rho_1 = Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{Cov(X,Y)}{\sqrt{(Var(X))^2}} = \frac{Cov(X,Y)}{Var(X)} = b.$$

Q.E.D.