Forecasting Time Series Homework 2

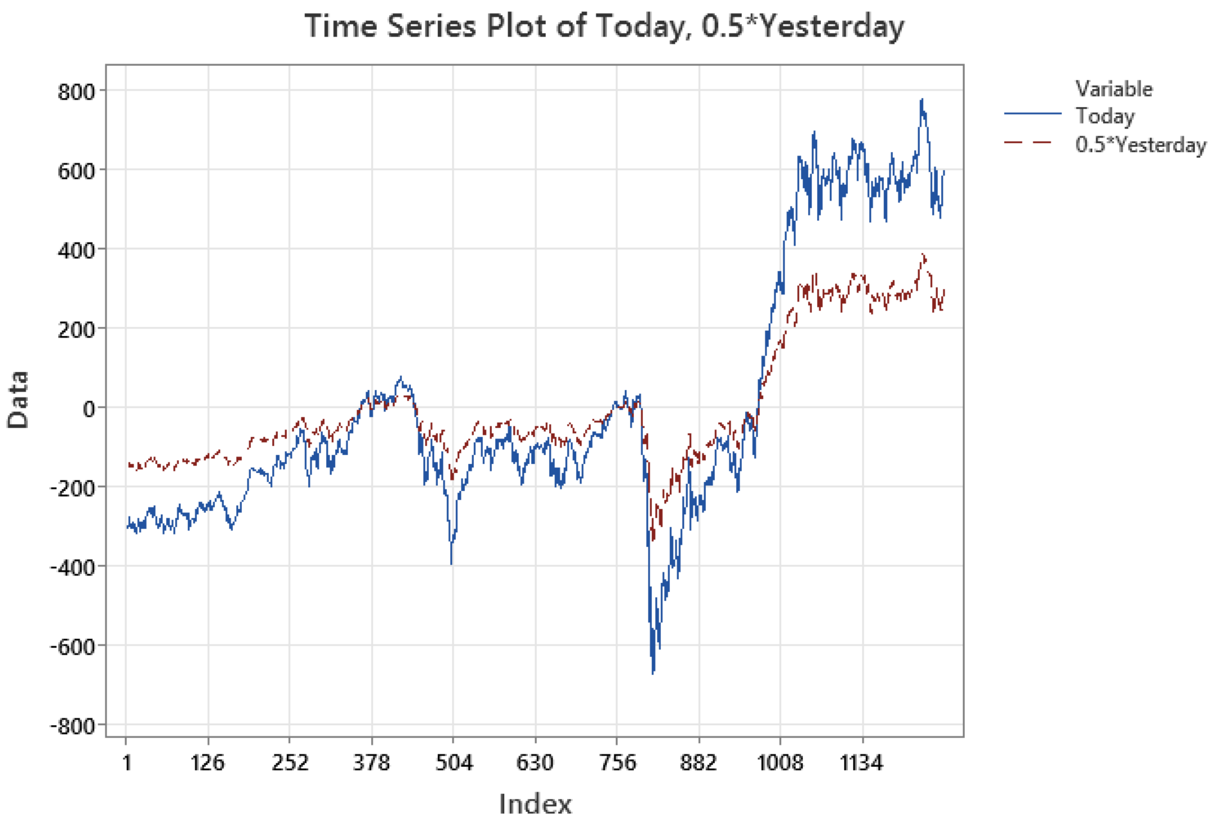
Yoon Tae Park[yp2201], Yeong Koh[yk2678], Yoobin Cheong[yc5206]

In Problems 1-2, we consider the mean-adjusted Russell 2000 data set, $x_t = \text{Russell}_t - \overline{\text{Russell}}$. "Today's Russell" is $x_2, \ldots, x_n$, and "Yesterday's Russell" is $x_1, \ldots, x_{n-1}$. To create x_t in Minitab, use Calc $\rightarrow$ Calculator, Store result in variable: Today, Expression: Russell - mean(Russell)

1)

A) On a single plot, draw Today's Russell versus time, as well as Yesterday's Russell versus time. (To create Yesterday's Russell, use Calc $\rightarrow$ Calculator, Store result in variable: Yesterday, Expression: lag(Today). To create the plot, use Graph $\rightarrow$ Time Series Plot $\rightarrow$ Multiple.) Next, on a single plot, draw Today's Russell versus time, as well as (0.5)(Yesterday's Russell) versus time.

Ans)



B) Based on these two plots, which seems to be a better forecast of Today's Russell: Yesterday's Russell, or (0.5)(Yesterday's Russell)?

Ans) Based on these two plots, Yesterday's Russell seems to be a better forecast of Today's Russell.

C) Calculate the average squared forecast errors for the two forecasts. Based on this, which one was better?

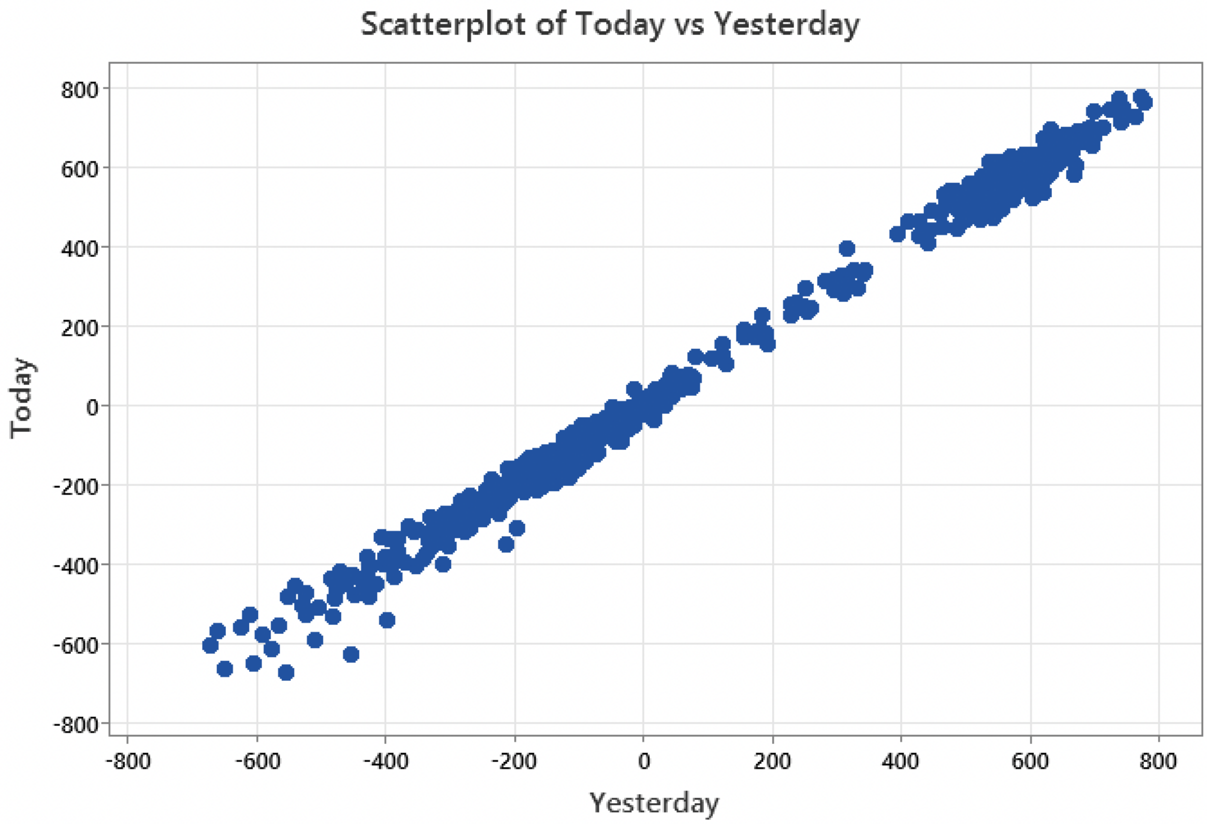
Ans) Based on the average squared forecast errors for the two forecasts, Yesterday's Russell is the better forecast.

C11	C12
MSFE(Yesterday)	MSFE(0.5*Yesterday)
562.632	25228.0

2)

A) Plot Today's Russell versus Yesterday's Russell. Describe any patterns you see.

Ans) In the plot, we see a positive correlation between Today's Russell and Yesterday's Russel.



B) Run a linear regression of Today's Russell (dependent variable) on Yesterday's Russell (independent variable). What is the prediction of Today's Russell implied by the regression coefficients? Is this consistent with your answers to Problem 1, parts B and C?

Ans) The prediction of Today's Russell (based on Adjusted Close) is 1256.40. This prediction is consistent with Problem 1 as it uses the regression with Today vs Yesterday with lower average squared forecast errors.

Regression Analysis: Today versus Yesterday

Method

Rows unused 1

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
23.7213	99.44%	99.44%	99.43%

Regression Equation

Today = 0.702 + 0.99817 Yesterday

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	124476481	124476481	221212.41	0.000
Yesterday	1	124476481	124476481	221212.41	0.000
Error	1255	706190	563		
Lack-of-Fit	1245	703046	565	1.80	0.150
Pure Error	10	3144	314		
Total	1256	125182671			

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.702	0.669	1.05	0.294	
Yesterday	0.99817	0.00212	470.33	0.000	1.00

Prediction for Today

Regression Equation

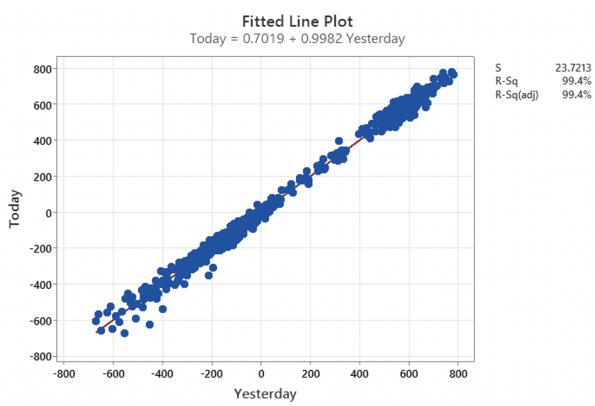
Today = 0.702 + 0.99817 Yesterday

Settings

Variable	Setting
Yesterday	1258

Prediction

Fit	SE Fit	95% CI	95% PI	
1256.40	2.75333	(1251.00, 1261.80)	(1209.55, 1303.25)	XX

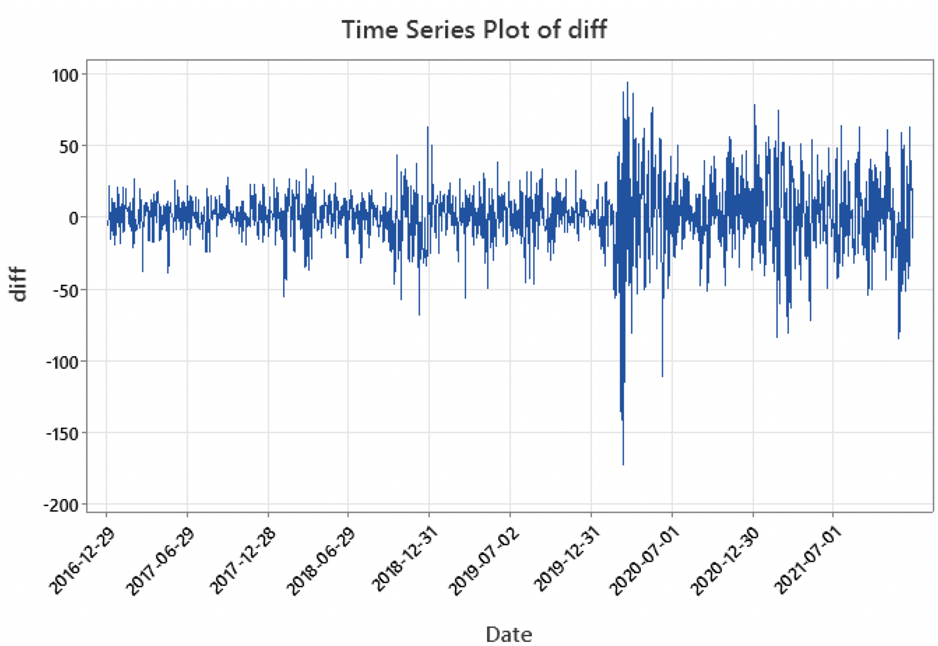


C) Is the slope in your fitted regression significantly different from 1? Briefly comment on the intercept as well. (Unfortunately, as we will see later, the p-values for the slope and intercept cannot necessarily be trusted when we regress a time series on a lagged version of itself, that is, $\{x_t\}$ on $\{x_{t-1}\}$. Furthermore, the p-values cannot necessarily be trusted when we regress $\{x_t\}$ on t.)

Ans) No, the slope is not significantly different from 1. Based on the t-test, if we set up our null hypothesis as μ = 1, we get $\frac{\beta-\mu}{\epsilon} = \frac{0.99817-1}{0.00212}$ = - 0.863 as a t-value. If we convert t-value into p-value, we get 0.388302, which is greater than the alpha value of 0.05. Therefore, we cannot conclude that the slope is significantly different from 1 at the significance level of 0.05.

D) Based on everything you have done so far, do you see any strong evidence that the Russell is *not* a random walk?

Ans) We do not see any strong evidence that Russell is not a random walk. By taking the difference once to remove the linear trend, we can observe that the difference looks like white noise. Also, by 2-C, we may say that the slope for yesterday's russell to today's russell is not significantly different from 1. Therefore, we cannot conclude that Russell is not a random walk.



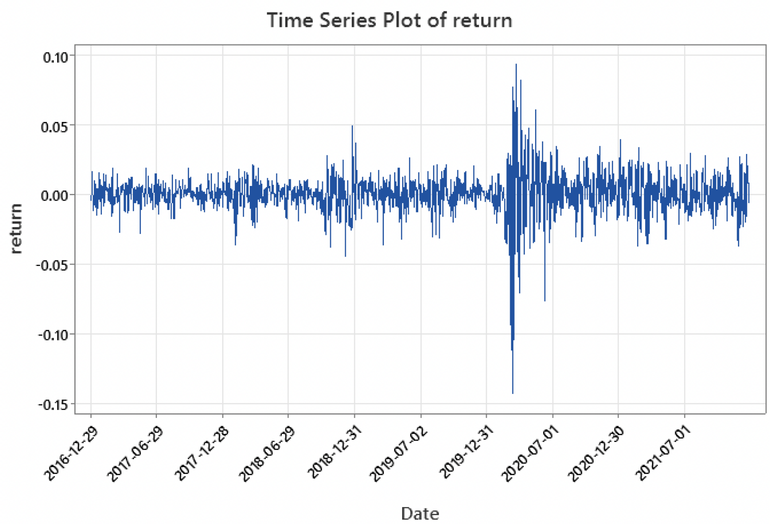
E) Compute the correlation coefficient between Today's Russell and Yesterday's Russell. (This is the square root of R^2 if the slope in the fitted regression is positive. It is $-\sqrt{R^2}$ if the slope is negative). Based on this, how strong is the linear association between Today's Russell and Yesterday's Russell? (Note: The correlation coefficient you got here should be quite close to the value of the slope you got in part C.)

Ans) The correlation coefficient between Today's Russell and Yesterday's Russell is $\sqrt{R^2} = \sqrt{0.9944} \approx 0.9972$ which is close to 1. This indicates that the relationship between Today's Russell and Yesterday's Russell is very close to being perfectly linear.

3)

A) Returning now to the non-mean-adjusted data, compute and plot the Russell returns = $(\text{Russell}_t - \text{Russell}_{t-1})/(\text{Russell}_{t-1})$, versus time. Compute the sample average and standard deviation of the returns. Based on an ordinary t-test, are the mean returns significantly different from zero? Interpret your findings.

Ans) The sample average and standard deviation of the returns are 0.000519 and 0.015480, respectively. The p-value is 0.235, which is greater than a=0.05. At the significance level of 0.05, we cannot reject the null hypothesis. There is no strong evidence that the mean of return is significantly different from zero.



One-Sample T: return

Descriptive Statistics

N	Mean	StDev	SE Mean	95% CI for μ
1257	0.000519	0.015480	0.000437	(-0.000338, 0.001375)

μ: population mean of return

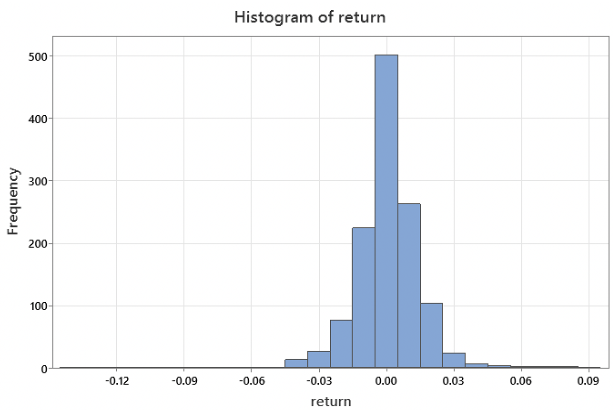
Test

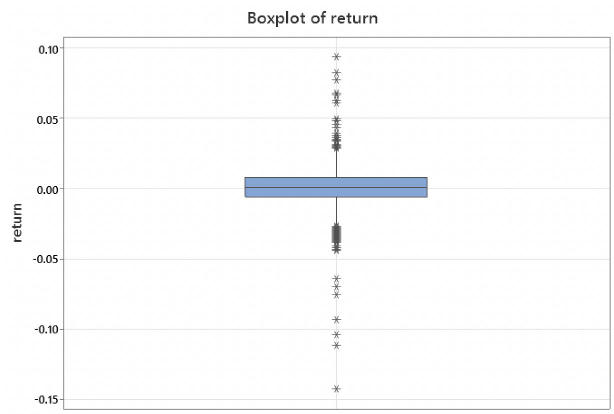
Null hypothesis	H_0: μ = 0
Alternative hypothesis	H_1: μ ≠ 0

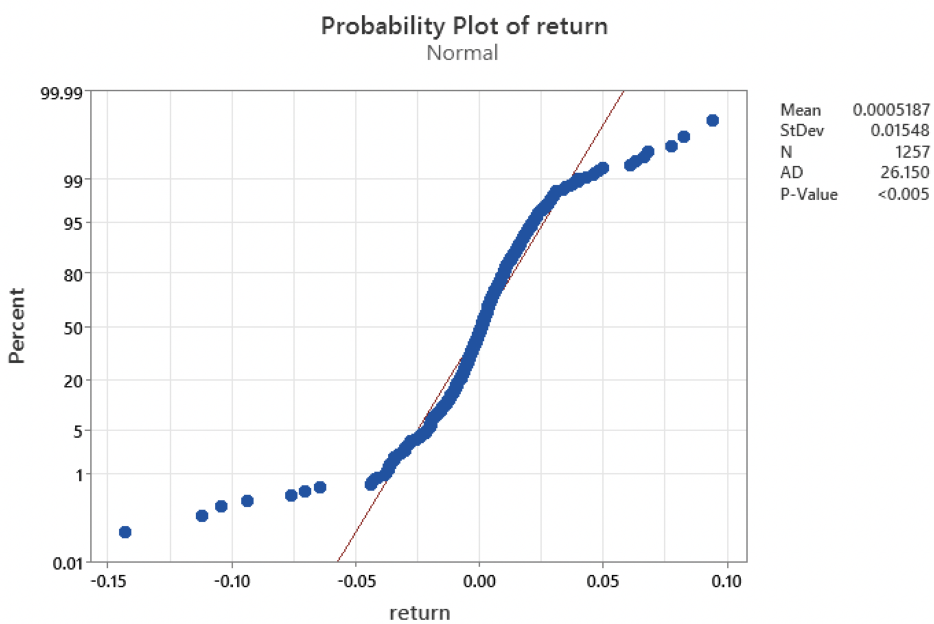
T-Value	P-Value
1.19	0.235

B) Plot a histogram and boxplot of the Russell returns. Also try a normal probability plot (Stat $\rightarrow$ Basic Statistics $\rightarrow$ Normality Test in Minitab), which should reveal an approximately straight-line pattern under normality. Do you think that the Russell returns are normally distributed? Explain.

Ans) From the histogram and the probability plot, we can see that the data is left-skewed. From our normality test, we see the p-value is smaller than a=0.05, so we reject the null hypothesis. Therefore, the data does not support a normal distribution.

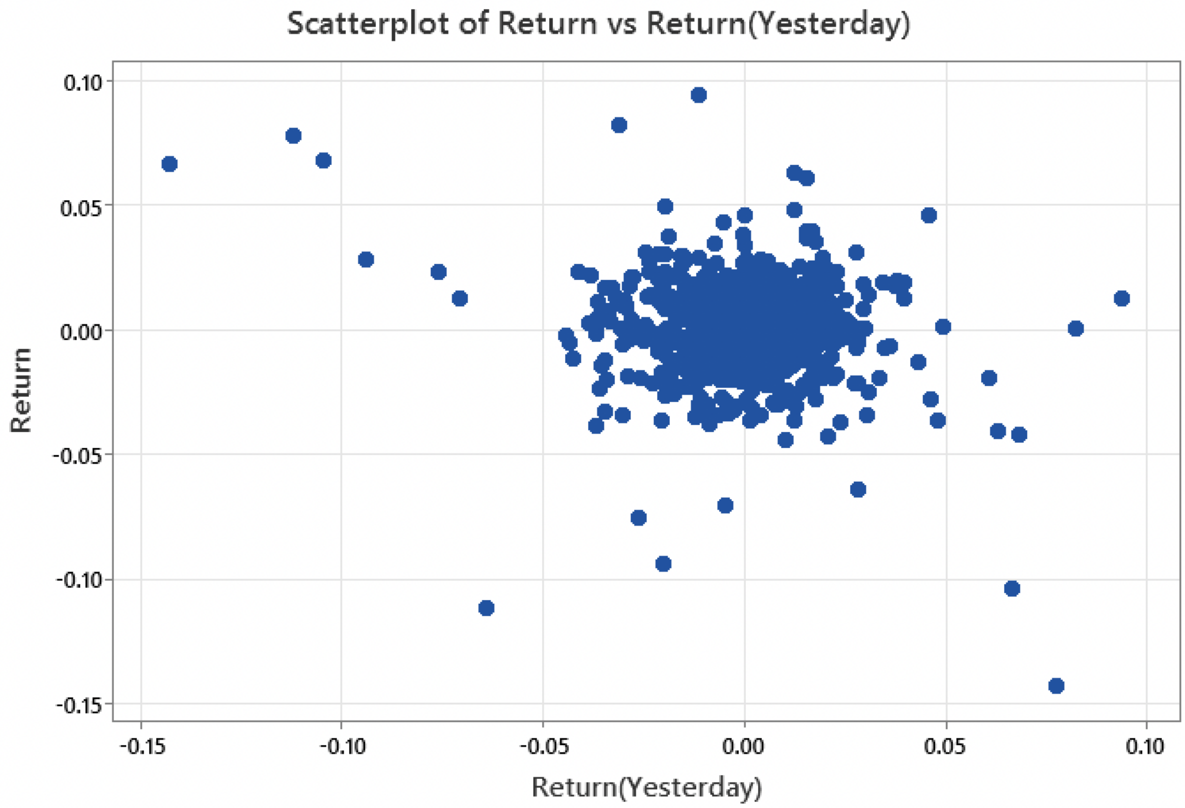






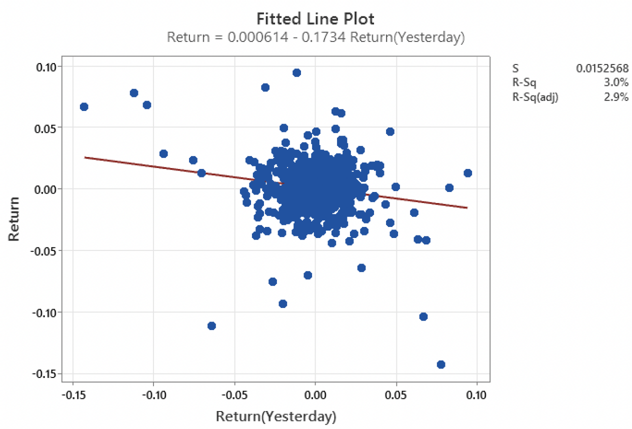
C) Plot today's returns versus yesterday's returns. Does this plot appear very different from the one in 2A)? Which seems to be easier to predict: Today's Russell, or Today's returns?

Ans) Yes. This plot appears very different from the one in 2A. Today's Russell is easier to predict than Today's returns.



D) Run a linear regression of today's returns (dependent variable) on yesterday's returns (independent variable). What is the prediction of today's returns implied by the regression coefficients? Are the coefficients statistically significantly different from zero?

Ans) Today's return can be predicted by multiplying -0.1734 by yesterday's return and then adding a constant term, 0.000614.
The coefficient of constant term is not statistically significantly different from zero because the p-value for constant is greater than the .05 threshold. However, the coefficient of the slope is statistically significantly different from zero because the p-value for slope is less than 0.05.



Regression Analysis: return versus lag(return)

Regression Equation

return = 0.000614 - 0.1734 lag(return)

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.000614	0.000431	1.42	0.155	
lag(return)	-0.1734	0.0278	-6.23	0.000	1.00

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.0152568	3.01%	2.93%	1.94%

4) If $\{x_t\}$ is stationary with $E[x_t]=0$ and $corr(x_t, x_{t-1})=\rho_1$, show that the best linear predictor of x_t based on x_{t-1} is $\rho_1 x_{t-1}$. (You will need to use calculus to do this problem. Here are some hints. First, define the random variables $Y=x_t$, $X=x_{t-1}$. Consider any linear predictor $\hat{Y}=a+bX$, where a and b are any numbers. Consider the mean squared forecasting error, $MSE=E[Y-\hat{Y}]^2=E[Y-(a+bX)]^2$.

Take the derivative of MSE with respect to a and set it equal to zero. Take the derivative of MSE with respect to b and set it equal to zero. Let's assume that the solution to these two equations for a and b gives us the coefficients which *minimize* MSE. By solving these two equations, you should conclude that the best a and b are given by $a=0$, and $b=E[XY]/Var[X]$. Now, use the fact that $\{x_t\}$ is stationary with $E[x_t]=0$ to show that the above expression for b is the same as ρ_1 in this case.)

Ans) Please see the written response on the next page.

$$MSE = E[Y-\hat{Y}]^2 = E[Y-(a+bx)]^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i-(a+bx_{i-1}))^2$$

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

$$= \frac{1}{n}\sum_{i=1}^{n}[2(y_i-a-bx_{i-1})(-1)] = \frac{2}{n}\sum_{i=1}^{n}(a+bx_{i-1}-y_i)$$

$$= 2a + 2bE(X) - 2E(Y)$$

$$\frac{d}{da}MSE = 2a+2bE(X)-2E(Y) \overset{set}{=} 0 \implies a = E(Y) - bE(X)$$

$E[X_t]=0$ and X_t is stationary $\longrightarrow$ $E(Y)=E(X)=0$ $\longrightarrow$ $a=0$

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

$$= \frac{1}{n}\sum_{i=1}^{n}[2(y_i-a-bx_{i-1})(-x_{i-1})] = \frac{2}{n}\sum_{i=1}^{n}[ax_{i-1}+bx_{i-1}^2-y_ix_{i-1}]$$

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

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

$a=0$

$-2E[X]E[Y]=0$ $\longrightarrow$ $2b\,Var(X) - 2\,Cov(X,Y) \overset{set}{=} 0$

$E[X]=0$

$E[Y]=0$

$$b = \frac{Cov(X,Y)}{Var(X)}$$

$Var(X) = Var(Y)$ since X_t is stationary

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

$$Cov(X,Y) = E[(X-E(X))(Y-E(Y))]$$

$X = X_{t-1}$ $\qquad$ $E[X_t]=0$

$y = X_t$ $\qquad$ $\{X_t\}$ is stationary

$E[X_{t-1}]=0$

$E[X]=0$ $\qquad$ $E[Y]=0$

$\Rightarrow E[X\cdot Y] = Cov(X,Y)$

$$b = \frac{E[X\cdot Y]}{Var(X)}$$