

FORECASTING HOMEWORK 2

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, \dots, x_n , and "Yesterday's Russell" is x_1, \dots, 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 $(0.5)(\text{Yesterday's Russell})$ versus time.

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

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

2)

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

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?

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 .)

D) Based on everything you have done so far, do you see any strong evidence that the 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.)

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.

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.

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?

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?

4) If $\{x_t\}$ is stationary with $E[x_t]=0$ and $\text{corr}(x_t, x_{t-1})=\rho_1$, 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.)