Ec 5043 Prof. Le Wang
Managerial Economics II

Homework #2

February 26, 2018

Instruction: Do all the following empirical exercises using R. Turn in your answer with tables and graphs, if any, (along with your program and output files appended at the end of document). Refer to the output file whenever appropriate when discussing your results.

Note that for all simulation exercises, set the seed number to 123456 to ensure the reproducibility of your results.

1. Question 1. [Covariance and Correlation]

Use the data **wage2.csv** for this question.

1. Using R, calculate the covariance and correlation between individual wages (**wage**) and his/her education (**educ**). **This question is corresponding to the definition and the implementation of covariance and correlation on the slides.**
2. Is the correlation in 1) statistically different from zero? Interpret the results. Does this result make sense? **This question is corresponding to test statistics of zero correlation.**
3. Suppose that individual wages (**wage**) are determined by education (**educ**), inate ability (**IQ**), and some other factors (ϵ) as follows

$$wage = a \cdot education + b \cdot IQ$$

Using the covariance formula that we learned from class, can you write the down the covariance between education and wages in terms of the following: **This question is corresponding to the properties of covariance, specifically Property 5.**

(a) The direct relationship between wages and education
(b) The indirect relationship between wages and education through ability (IQ) .

2. Question 2. [Combined Forecast]

Use the data **forecast.csv** for this question. There are threee variables of interset. The original object of interest, y, and forecasts of $\boldsymbol{y}$ from two different models, **forecast1** and **forecast2**. Use what we learned from class (**Application 3: Optimal Forecast Combination**), create a combined forecat that could minimize the variance of forecast error. Does it work? Comment on the reason why it may or may not work.

3. **Question 3. [Portfolio Diversification and Nonlinear Dependence]**

Pick any 3 stocks that you are intersted in

1. You can download them using **quantmod in R.** Pick the data after 2015.
2. Create the best portfolio with only two stocks with equal weights that could minimize the risks. **This question is corresponding to Application 2: Portfolio Diversification on the slides.**
3. Using entropy, calculate the dependence between two stocks that you pick in 2. **This question is corresponding to Entropy and Predictability on the slides.**

4. **Question . [Conditional Distribution, Transition Matrix, and Income Mobility]**

Policymakers and economists are often interested in measuring and understanding income inequality in society. However, as discussed in class, a snapshot of income dispersion in a given time period t, which ignores potential movement over time, does not necessarily give us a good sense of income inequality. A better measure would be to take into account income dynamics for each individual and capture how his/her income moves up and down over his/her life cycle. Transition matrix serves this purpose. Specifically, transition matrix is nothing but a variant of conditional distribution, with each cell being

$$\Pr[Y_t|Y_{t-1}]$$

where Y_t is one's income class in t (defined as lower, middle, and upper classes;); and Y_{t-1} is one's income class in $t-1$. Let's use the following simulated data (remember to set the seed as instructed above). `y.tm1` and `y.t` are incomes in time periods $t-1$ and t, respectively. **This question is corresponding to the definition of various concepts of joint, marginal and conditional distributions. It is also corresponding to the application to income mobility, as discussed in class.**

```
library(MASS)
Sigma <- matrix(c(10,3,3,2),2,2)
data<-mvrnorm(n=1000, rep(0, 2), Sigma)
y.tm1<- data[,1]
y.t<- data[,2]
```

1. Discretize the continuous income variables into three categories. Lower class is defined as those with income less than or equal to the first quartile of the income distribution in each period. Middle class is defined as those with income more than the first quartile but less than or equal to the third quartile of the income distribution. Upper class is the rest of the population. **Here I would like you to learn how you can generate groups based on continuous variables. This provides a very convenient way to approximate continuous variables in pratice.**
2. Use R to calculate the joint distribution
3. Use R to calculate the marginal distribution
4. Use R to calculate the conditional distribution, that is, the transition matrix.
5. Is there any income mobility in this society?