Your name: _____

UCLA Anderson School of Management

MGMTMFE 431 – Data Analytics and Machine Learning

Final 2018

You are allowed one cheat sheet and a calculator.

If a question is unclear to you, please make assumptions as you see fit to answer the question.

Questions start on page 2 and ends on page 13.

Good luck!

Prof. Lars A. Lochstoer

1. Consider the following panel regression:

$$rv_{i,t} = \mu_i + \delta_1 rv_{i,t-1} + \varepsilon_{i,t},$$

where $i = 1, ..., N$ refers to firm and $t = 1, ..., T$ refers to time. The left hand side variable, $rv_{i,t}$, is realized monthly variance based on log daily returns.

(a) What is the interpretation of μ_i? Why would one want to include this in the regression specification?

(b) Given estimates $\hat{\mu}_i$ and $\hat{\delta}_1$ and $rv_{i,t}$, what is the expected realized variance for the next two months, $E_t\left[rv_{i,t+1} + rv_{i,t+2}\right]$? Give an analytical expression.

(c) Assume you also have a panel of implied option volatilities (squared to make variances), $iv^2_{i,t}$, for 2-month options at the end of each month t for each firm i. How would you use the model and the option IV's to create a cross-sectional options trading-strategy? In your answer, make sure you discuss the following issues.

i. How do you implement you strategy in real-time?

ii. How would you construct an option portfolio that takes advantage of your strategy?

iii. How would you test if your strategy is working (i.e., is economically valuable)?

2. Shrinkage and regularization. Consider a ridge regression of an outcome variable, y_{t+1}, on a predictor, x_t.

   (a) Give the objective function of the ridge regression assuming you have a time series with T observations of both variables. Make sure you clearly define the constraint.

   (b) Derive the resulting regression coefficients, β. Show your math clearly.

(c) Explain how K-fold cross-validation works in the ridge regression setting. How does the cross-validation affect the ridge β estimate?

(d) Assume the regression residuals are normally distributed and assume the following prior for β: $\beta \sim N(0, A)$ where N() denotes the normal distribution. Derive analytically the posterior distribution for β and relate the prior variance to the ridge regression estimate.

(e) Assume instead that your prior is $\beta \sim N(1, A)$. E.g., you are estimating a market beta. How can you modify the ridge regression to reflect this prior?

(f) Next, give instead the objective function for a LASSO regression. How is the LASSO constraint different from the Ridge constraint. How does this different objective function tend to change the beta estimate? Are there any reasons to prefer one over the other?

3. You want to estimate the probability of firm-level monthly volatility spikes. In particular, you want the time t probability that realized variance $rv_{i,t}$ (see question 1) is greater than 0.01. The variables you want to use are firm-level lagged realized variance, $rv_{i,t}$, firm book-to-market ratios, $bm_{i,t}$, as well as industry dummies. You choose a logistic regression for this exercise.

   (a) Give the logistic regression specification, where you take care to define all terms.

   (b) Assuming a balanced panel with T time-series observations of all variables and N firms, give the log likelihood function you will estimate.

(c) Explain in this context what "null deviance" and "residual deviance" refer to.

(d) Given a series of predictions $\hat{p}_{i,t}$ for each firm and time. Explain in detail how you would construct an ROC curve to assess model fit. Draw an illustrative graph, label axis, and define all relevant variables. In the process, also explain concept of AUC.

4. Below is a data set of alpha's for different fund managers, as well as the percentage management fee each fund charges and the size (Net Asset Value) of the fund. You want to use a decision tree to predict alpha based on the management fee and fund size. Your tree is to have two terminal nodes.

Fund	Alpha (in %)	Fee (in %)	NAV ($ million)
A	-1.5	0.9	200
B	0.7	0.8	40
C	0.9	0.4	150

   (a) Using Recursive Binary Splitting, create the decision tree. Draw the tree below. Give the intermediate node and its breakpoint, as well as the two terminal node values. Show your calculations.

(b) Give the qualitative intuition for the decision tree. I.e., what predicts mutual fund alpha? Does it make sense relative to what you know about mutual fund performance?

(c) Briefly explain how *bagging* works to improve the mean squared error of the decision tree's prediction error.

(d) Briefly explain the main difference between linear regression models and decision trees. Use equations to illustrate your logic.

(extra space)