Last Name: ______________________ First Name: ______________________

M.I.T. ID# ____

15.450 Final Exam

Professor Hui Chen

Spring 2018

Please read the following instructions carefully.

- Please write your name and MIT ID number on the first two pages.
- The exam lasts 180 minutes. Please try to answer each of the questions.
- You are allowed two $8\frac{1}{2}$”$\times$11” sheets of formulas and one calculator.
- Answer these questions without consulting anyone.
- Always explain your answers and show your work, but try to be concise. Answers without explanations will receive no credit. Wrong answers with partially correct work may receive partial credit.

Good luck!

Name: ____________________ MIT ID#: ____________________

15.450 Final Exam Grade Sheet

1.	________	/	60
2.	________	/	10
3.	________	/	20
4.	________	/	10
Total	________	/	100

1. (60 points) **SHORT QUESTIONS** (Please briefly explain your answers.)

   (a) (7 points) Explain intuitively why a GARCH model generates heavy tails in the unconditional return distribution even though returns are conditionally normal.

   A GARCH model has the feature of high persistence of volatility. As a result, conditional volatility could cluster and can sometimes become much higher than the unconditional (i.e., average) volatility. In other words, GARCH will be able to generate large variation in the conditional volatility, which will contribute to the fat tail in the unconditional return distribution.

   (b) (7 points) You are building a logistic model for corporate defaults. The response variable $Y_{it} = 1$ if firm i defaults in year t, and $Y_{it} = 0$ otherwise. Among the list of predictors is $TLTA$, which is a leverage ratio measure based on firm i's total liability divided by total assets in year $t - 1$. The MLE coefficient for $TLTA$ is 3.36. Please interpret the meaning of this coefficent.

   This tells us that the probability of default is increasing in $TLTA$. In terms of economic magnitude, for example, a change in leverage ratio from 0 to 0.5 increases the log odds of corporate default by 1.68 (3.36×0.5), holding everything else fixed. Unlike a linear model, we cannot interpret the coefficient as the partial derivative or the exact marginal effect in terms of the probability of default.

(c) (15 points) For each of the scenarios below, indicate whether we would generally expect the performance of a linear model fitted using OLS to be better or worse than using a shrinkage method such as ridge regression or LASSO. Justify your answer.

i. The sample size n is extremely large, and the number of predictors p is small.

OLS will work well here. Since $X'X$ and $X'Y$ can be interpreted as sample moments, we might expect that with large n these are close to their population counterparts and hence the OLS coefficient is close to the true projection coefficient.

ii. The number of predictors p is extremely large, and the number of observations n is small.

This is the perfect time for LASSO or Ridge. Note that, in the extreme case where $p > n$, the OLS coefficient is not even uniquely defined!

iii. The variance of the error terms, i.e., $\sigma^2 = \mathrm{Var}(\epsilon)$, is extremely high.

This is unrelated to the choice between OLS and shrinkage methods. Everything else being equal, I suppose we should go with OLS, since it is BLUE under the Gauss Markov Theorem.

(d) (8 points) In an event study of stock returns around earnings announcements, how should we deal with the fact that multiple firms announce their earnings on the same days?

We could group the stocks into a portfolio. The problem here is that this implicitly assumes the stocks will have (near) perfect correlation in their event window movements. Alternatively, we can run a bunch of linear regressions with dummy variables for the event dates. This can be done easily with GMM, running one regression for each stock.

(e) (8 points) You have developed a classification model to detect credit card frauds. Based on the various information of a card applicant, the model produces the probability that the application is fraudulent. Explain how you would use this model to make the decision to approve or deny an application.

We could run a logit model with all our predictors. We would pick a cutoff, say, $\bar{p}$. Then, using each applicant's characterisitcs, we could get an estimate $\hat{p}$ of his or her probability of being fraudulent. If $\hat{p} > \bar{p}$, we deny, otherwise approve.

(f) (7 points) In order to better advise clients on designing compensation packages, you are studying the determinants of CEO compensation across firms. You are quite aware of the fact that both firm size and CEO's managerial ability should be important determinants of compensation, but unfortunately you do not have data to measure the abilities. Instead, you estimated the following model,

$$y_i = \beta_0 + \beta_1 SIZE_i + \epsilon_i,$$

where y_i measure the compensation for the CEO from firm i, and $SIZE_i$ measures firm i's size (market cap). How does the omission of managerial ability from the above model affect the coefficient estimate $\widehat{\beta}_1$?

We have omitted variable bias. If ability is uncorrelated with $SIZE$, then $\hat{\beta}_1$ is unaffected. If, for example, ability and size are positively correlated, and, as we might expect, ability increases compensation, then $\hat{\beta}_1$ is larger (positive) than the true coefficient.

(g) (7 points) Outline the steps you would take to do cross-validation for a decision-tree model that predicts house prices using house characteristics.

To do a K-fold cross-validation, we can: 1) Split the dataset into K "folds" of equal size. 2) Pick a value for the tuning parameter, for example, the number of terminal nodes in a decision-tree model. 3) Use each fold as the testing set, and the K-1 folds as training data to train the decision-tree model that predicts house prices using house characteristics. 3) Average testing performance over K tests, and take it as the estimate of out-of-sample performance. 4) Based on the out-of-sample performance, decide the optimal value for the tuning parameter.

2. (10 points) In a study published in the Journal of Finance, Lucca and Moench (2015) find that U.S. equities earn large excess returns in anticipation of monetary policy decisions made at scheduled meetings of the Federal Open Market Committee. Design a study to check whether this claim is indeed true in the data. Explain all the steps and specify what data you would need.

   To do an even study, we can: 1) Get some measure of excess equity returns (presumably S&P 500), and the calendar of past and scheduled FOMC meetings over the sample period such as 1993–2016. 2) Define an FOMC meeting to be the event and define the day of an FOMC meeting as the event window ($T_1 = 0, T_2 = 1$). 3) Specify the reference model for "normal" returns, for example, the constant mean model, and estimate model over the estimation window of past one year($T_0 = -249$, $T_1 = 0$). 4) State the null hypothesis: the mean excess return is unchanged on FOMC announcement dates. 5) Average the cumulative abnormal returns on event days and characterize its distribution under the null; 6) perform hypothesis testing.

3. (20 points) **An interest rate model with stochastic volatility**

Suppose you observe two time series, the interest rates Y_t and unemployment rate X_t. You have a model for Y_t:

$$Y_{t+1} = \rho Y_t + (a_0 + a_1 X_t)\,\epsilon_{t+1},\ \ t = 1, ..., T$$

where $\epsilon_{t+1} \sim \mathcal{N}(0, 1)$ and are independent of the shocks to unemployment X_t and the lagged values of Y_t. There is no model for X_t.

(a) Show that it is valid to estimate ρ using an OLS regression of Y_{t+1} on Y_t.

We are running:

$$Y_{t+1} = \rho Y_t + v_{t+1}$$

for some error v_{t+1}. OLS assumes:

$$E_t[Y_t v_{t+1}] = 0$$

In this case, we know, $v_{t+1} = (a_0 + a_1 X_t)\epsilon_{t+1}$. This implies:

$$E_t[Y_t(a_0 + a_1 X_t)\epsilon_{t+1}] = Y_t(a_0 + a_1 X_t)E_t[\epsilon_{t+1}] = 0$$

where we used the fact that time-t variables are known and the that ϵ is mean 0. Thus, OLS is valid (asymptotically the estimated $\hat{\rho}$ will converge to the true ρ).

(b) Suppose that the variance of the estimator $\widehat{\rho}$ is $(1/T)\sigma_\rho^2$. Describe how you would test the hypothesis that $\rho = 0$.

With a large-enough sample, by the standard GMM properties, $\hat{\rho}$ will be asymptotically normal, so

$$\frac{\hat{\rho}}{\sqrt{(1/T)\sigma_\rho^2}} \sim N(0, 1)$$

We could compare this to, say, the 95% critical value.

(c) Write down the conditional log-likelihood function of interest rates:

$$\mathcal{L}(Y_2, \cdots, Y_T | Y_1, \rho, a_0, a_1)$$

Since $\epsilon_{t+1} \sim N(0, 1)$, we see that $Y_{t+1} \sim N(\rho Y_t, [a_0 + a_1 X_t]^2)$ Therefore:

$$\ln \mathcal{L} = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum \ln([a_0 + a_1 X_t]^2) - \frac{1}{2} \sum \frac{(Y_{t+1} - \rho Y_t)^2}{[a_0 + a_1 X_t]^2}$$

(d) Suppose that the parameters a_0 and a_1 are known. Derive the maximum-likelihood estimate for ρ.

The first order condition gives (and therefore the MLE estimate of $\hat{\rho}$ solves):

$$\frac{1}{T} \sum \frac{Y_{t+1} - \hat{\rho} Y_t}{[a_0 + a_1 X_t]^2} Y_t = 0$$

This is weighted least squares, where we know the variance of the error.

4. (10 points) The following is an excerpt from a Bloomberg article in 2017:

> *"It's been said that the stock-picker's market is back ... Equity correlations plunged to a record low, as economic and policy optimism drew distinct leadership on U.S. exchanges. It was a welcome development for fund managers, whose efforts to pick winners [and beat their benchmarks] have been thwarted for years as shares swung in unison."*

The purpose of this question is to more rigorously examine the above statement.

(a) Interpret the statement. Specifically, if the statement is true, when is an active manager (who does stock-picking) more likely to out-perform the market — when the average correlation among stocks is high or low?

This is basically saying that in order for stock-pickers to succeed they must be able to find stocks that will perform differently from one another. That is, when correlations are low, stock-pickers do better, presumably because they (think they) can differentiate which stocks will go up or down, which becomes much harder when they all move together.

(b) Next, you want to do your own investigation. You have the data of quarterly returns for a group of active managers and the returns of the market index from 1991 to 2015. In addition, you have individual returns of all publicly traded stocks during the same period. List the steps to rigorously assess whether the statement above is true or not.

(**Hint:** You might need a way to measure how high equity correlation is in a particular year. For this you can use the cross-sectional standard deviation of returns across all stocks in that year. Suppose there are N stocks,

$$V_t = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(r_{it} - \bar{r}_t)^2}, \quad \bar{r}_t = \frac{1}{N}\sum_{i=1}^{N} r_{it}$$

Intuitively, if returns across stocks are highly correlated, their values should be closer to each other, leading to lower V_t.)

The simplest way would be to treat V_t as a factor, as in, say FF3 model. Then we would look to see if our portfolio of active manager returns has a positive or negative loading on V_t. We may want to control for other variables, like the other Fama French factors, as well.

Extra space