MF 840 – Spring 2021
Study Guide for Exam 1

- The exam is from 8:15am to 10:15am. Use the 8:00am section zoom link to log in. Read these instructions now so you will not be surprised. The extra 15 minutes is to give you time to scan your answers and upload your solved exam.

- You will log in on **Gradescope** (not Questrom Tools) to download the exam. It will be available at 8:15am. You will upload your completed exam on Gradescope. Uploading must be over at 10:30am with no recourse. Make sure to upload each question separately in the allocated space on the template.

- You may consult anything posted on the MF840 and MF793 Questrom Tools site or the recommended readiings. You are encouraged to use R for computations. **You should not use the "*internet*":** anything not in line with the notation or the assumptions of the course will be considered wrong with no recourse. If your answer appears to have been copied from an internet web site, you will get a zero and be referred to the Graduate Office for violation of the academic integrity rules.

- The exam is zoom proctored: **You must never leave zoom and must have your video on at all times otherwise points will be taken off your total grade.** In brief, if we don't see you, you lose points.

- You **cannot communicate with anybody by any means** during the exam. So if you are at home, make it clear that you cannot be disturbed or talked to during the exam. So you can not use any device to call, email, text, chat, zoom, etc.. with anybody. Put your phone in Do Not Disturb mode during the duration of the exam.

- You can not be in the same room as another student taking the exam.

- The exam has a number of independent questions: algebraic or number calculations as in class or homework, or a bit different, discussion questions involving a concise but complete justification of the answer to check that you understood the discussions in class.

- Discussion questions may have a True / False feature: If part of a statement is correct and part is False, you must label the statement as False. Then in your discussion, write clearly what is correct, what is false. and why. If you say True, explain why the entire statement is true. Saying True with no explanation gets zero point even if it is correct.

- The topics to review include what we did in class, the two articles, and the two problem sets. If you make sure that you understand, can do the proofs, and can use all the material in the lecture notes and homework, you are in very good shape.

- Correct numerical answers without justification or starting theoretical formula get zero point.

General guidelines

- Formulas to know: every formula or definition in blue or red or bold face. Specific guidelines may override this either way, read the specific guide below.

- The outline below reinforces the most important concepts to check during your studying. But you should be comfortable with all we discussed in class.

- Able to reproduce the proofs that we did in class and in the notes. Some proofs may be excluded in the specific guidelines below, otherwise be able to prove everything.

- R Code: You may need to use R code directly related to the homework and the R files in the Lecture note folder. You need to review the R files in the Lecture Notes Folder, they are lecture notes designed to implement and learn the concepts, as we used them in class. I want to know that you can tell an interviewer how to code a Wald test, a SUR estimate or covariance matrix, a GLS or MLE.

- Understand all the problems and solutions in the problem sets, including the discussion questions.

Specific Topics guidelines

- Review of Inference
  In this note, you should know how to prove everything in the note or slight variations thereof.
  - OLS
  - GLS: theory, kitchen recipe for feasible GLS, specific of AR(1)
  - SUR: general results, writing big **Y,X,€** with Kronecker, proving results when all Xs are equal.

- Computing and Maximizing the likelihood: MLE, basically everything in the Notes, know how to prove and how to use.
  - Writing a likelihood, maximizing it to compute ML estimates, their covariance matrix by the information matrix using Cramer-Rao.
  - Properties of the MLE: mean, variance, efficiency
  - Proving the equality between the "second derivative" and the "squared first derivative" of the log-likelihood
  - Using Delta method for functions of parameters, make sure you can do it for multivariate functions
  - Examples: sample mean and variance by MLE with i.id. and non i.i.d errors
  - Regression with i.i.d or non i.i.d. errors by MLE. (GLS situation by MLE)
  - Hessian

- Multivariate Normal Distribution
  - Anything on P 1-7, know, understand, prove
  - P8: Be able to explain and interpret (7)(8). Know proof for bivariate and multiple regression case.
  - P9: Be able to write (9) and work the proof until (10) explaining the steps of "completing the square".
  - P10: will not ask to prove that (10) is (7), too tedious and not very interesting.
  - P11: Know and understand every detail of these results.
  - MLE: Know proof until P15. Step 1: know how to introduce the sufficient statistics. Then find MLE of the mean vector. Be able to explain the intuition for $\widehat{\Sigma}_{MLE}$, no need to know the J&W result on P16, I will give it, but know how to use it to find $\widehat{\Sigma}_{MLE}$.
  - Wishart: understand the properties.

- Power of the Test: Any power of the test computation.

- Classical Tests and data mining corrections
  - Be able to compute probabilities of Type I and type II errors, power of the test. Discuss issues of size vs. Type II. The R file is crucial for your review, practice with it to gain understanding of the effect of α, sample size.
  - Consider the movie and the R file as teaching notes !
  - Hotelling T^2: Exact F distribution: Know, understand, know how to use. Proof of asymptotic distribution.
  - Wald, LRT, LM: know general formulas, graphical proofs of Wald≈LRT, LM≈LRT, proof of Wald is like T^2 for linear restrictions. Homework on implementing Wald.
  - Data Mining: Be able to explain data mining. Computing exact size of the naïve t-test for multiple tests assuming independence. Computing exact size adjustments for multiple tests assuming independence. Computing Bonferroni and Hotelling size adjustment for multiple tests. Proof of Bonferroni correction, be able to explain the rationale behind Hotelling correction,.

- Sorting by Estimates
  - Understand everything about it !
  - Understand the two R lecture notes, and be able to discuss A2 Problem 3 results.

- Fama-McBeth.

  Know the outline of the methodology. These key specific questions discussed in class in this order, mostly have a one or two sentence answer. If you can answer them, you are doing great. Practice
  - Why group the stocks into portfolios?
  - When making portfolios, why not randomly allocate the (say 2000) stocks to (say 40) random portfolios?
  - What problem do we resolve when we make portfolios on the basis of sorted betas?
  - What problem do we cause when we make portfolios on the basis of sorted betas?
  - Be able to describe the two methods seen to alleviate the problem causee above. Method 1 seen in class and Method 2 seen in the homework
  - Is the second pass regression time series or cross sectional? Is it predictive?
  - What should the coefficients of the second pass regression be if CAPM is correct?
  - For practitioners who use a cross-sectional regression to predict cross-sectional returns differences, why is it important to also include risk measures (such as factor betas)?
  - You ran the cross sectional regression for many years (or quarters) in a row. How do you estimate the variance of $\bar{\gamma}$ the average of the slope coefficients $\hat{\gamma}_t s$ when you want to do a t-test of H_0: $\gamma = 0$?

- You should be able to operate all the R files given (A1 and A2 solutions as well as R lecture notes) with parameter modifications if needed to produce results during the exam

Additional Exercises (No solution will be given)

Consider the regression $Y = X\beta + \epsilon$, where $\epsilon \sim N(0, \Omega)$ (no σ^2 in this notation). Ω is known. You collect T observations of Y and X.

a) Write the joint density of the vector ϵ , then of the vector Y.

b) Find the MLE estimator of β.

c) Find its covariance matrix.

In the multiperiod i.i.d. log-normal model, $r_t = \log(1 + R_t) \sim N(\mu, \sigma^2)$, you know and should be able to prove that the sample mean can be written as $\mu + (1/T) \sum_1^T \sigma\epsilon_t$, where $\epsilon_t \sim N(0,1)$.

a) If returns are autocorrelated with correlation matrix C_T (dimension TxT for a sample of T observations), rewrite $\sigma \sum_1^T \epsilon_t$ to allow for this autocorrelation. Then show that $\hat{\mu} = \mu + \omega\, \sigma\sqrt{i' C_T\, i}\,/T$, with $\omega \sim N(0,1)$.
b) Use this result to find the variance of the sample mean when the errors follow an MA(1) with first order autocorrelation θ. What is the % increase in sample mean variance over the i.i.d case if $\theta = 0.2$ and T=120?

Additional Voluntary Readings

If you feel you want to read more about the topics above, I have flagged below some appropriate sections of Greene. This is not required. I am showing it so that you don't start reading unnecessary sections !

GLS: Greene 9.3.1 Note how theorem 9.5 is basically Gauss-Markov for GLS. Greene's proofs use asymptotics for X as well, we did not do that. Ignore testing (material after Theorem 9.5 box)

Feasible GLS: Green 9.3.1. See the formal requirement 9.16 on the estimate of Ω in order to get the feasible GLS to converge to the "true" GLS. It is just a consistency requirement, note Theorem 9.6

> Interesting for background: Greene 9.4.4, eqn. 9.27 contains the formal proof of the OLS HAC standard errors in case of heteroskedasticity which we did in MF793. One of you asked for more background on robust error. This is it.

Weighted Least Square: Greene 9.6

SUR: Greene 10.2 – 10.3

MLE: Greene 14.1-14.3, 14.4.1 until Theorem 14.1 included, 14.4.2-14.4.4, Example 14.3, only skim 14.4.5, 14.9