vrije Universiteit amsterdam

School of Business and Economics

Exam:	Multivariate Statistics
Code:	E_EOR2_MS
Examinator:	Andre Lucas
Co-reader:	xxx
Date:	March 26, 2019
Time:	12:00
Duration:	2 hours
Calculator allowed:	Yes
Graphical calculator allowed:	Yes
Number of questions:	4
Type of questions:	Open
Answer in:	English

Remarks: This is a trial exam to give an indication of types of questions. It is too long for a 2h span.

Credit score:	
Grades:	The grades will be made public on: April 9, 2019.
Inspection:	Wednesday, April 10, 2019 at 10am. Room tba via CANVAS.
Number of pages:	6 (including front page)

Good luck!

Question 1.
You model European firms and have observations on the following 4 variables:
X_1: Profits (EBIT)
X_2: Asset Size (A)
X_3: Long Term Debt (D).

You assume these variables come from a multivariate normal distribution.
a) Give a reason why this may *not* be a good modeling assumption.

You proceed with your multivariate normality assumption nonetheless.
b) Provide the analytic form of the normal pdf for this particular case. Carefully define your variables, parameters, and the dimensions/formats of those.

You find out that people typically do not use the variables above in an economic analysis. Rather, they use the following variables:
Y_1: Return on Assets (ROA) = EBIT / Assets
Y_2: Size (S) = natural logarithm of Assets
Y_3: Leverage (LEV) = Debt over Assets

c) Given your modeling assumption for the X_i variables, derive the joint density of your Y_i variables. Carefully provide your steps.

Assume you have the following relationships:

$$X_3 \sim N(3,3)$$
$$X_1 | X_3 \sim N(X_3, 2),$$
$$X_2 | X_1, X_3 \sim N(X_1 - X_3, 4).$$

d) Derive the distributions of $X = (X_1, X_2, X_3)^\top$, and of $(X_1 + X_3) | (X_1 + X_2)$.

e) You try to approximate X_1 by the best linear function in X_2, X_3. What is the form of this function, and what is its precision?

Question 2.
You consider a bivariate copula $C(u_1, u_2)$. Based on your copula, your fellow students suggest alternative copulas $\tilde{C}(u_1, u_2)$.

a) Which of their suggestions will yield a proper copula function, and which will not (and why)?
a.1) $\tilde{C}(u_1, u_2) = 1 - C(u_1, u_2)$.
a.2) $\tilde{C}(u_1, u_2) = 1 - C(1 - u_1, 1 - u_2)$.
a.3) $\tilde{C}(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2)$.

Consider the copula $C(u_1, u_2) = u_1 u_2 (1 - 0.5(1 - u_1)(1 - u_2))^{-1}$. Also consider two random variables X_1 and X_2 with cdf $F_{X_1}(x_1) = 1 - \exp(-0.1x_1)$ and $F_{X_2}(x_2) = 1 - \exp(-0.2x_2)$, respectively, and with the same copula structure.

b) Compute $P[X_1 \leq 2,\ X_2 \leq 1]$ as well as $P[X_1 \leq 2 \mid X_2 = 1]$, as well as $f_X(2,1)$, where $f_X(x_1, x_2)$ is the pdf of $X = (X_1, X_2)$.

You observe 5 stock index returns and you want to model them using a Student's t copula, while also using Student's t marginals. For the marginals, you impose that the degrees of freedom parameter for each of the marginals exceeds 2.

c) Why would you impose that the degree of freedom is larger than 2?

d) What (if anything) does this assumption imply for the degrees of freedom of the copula? Explain your answer.

Question 3.

You consider a sample from a bivariate normal. The means of these normals should be *negative*, and the variables should be positively correlated. Also the first variable should have a higher standard deviation. You friend parameterizes the distribution using the parameters $\theta_1, \dots, \theta_5$:

$$\mu_1 = -\exp(\theta_1), \quad \mu_2 = -\exp(\theta_2),$$

$$\sigma_{11} = \sigma_1^2 = \exp(\theta_3) + \exp(\theta_4), \quad \sigma_{22} = \sigma_2^2 = \exp(\theta_4), \quad \rho_{12} = (1 + \exp(-\theta_5))^{-1},$$

where μ_1 and μ_2 are the means of the two coordinates, σ_{11} and σ_{22} the variances, and ρ_{12} the correlation.

a) Does this parameterization satisfy all the constraints that you want to impose? Explain your answer.

b) Write down the likelihood function for your model using your parameterization.

Using your maximum likelihood estimation routine, you obtain (with standard errors in parentheses)

$$\hat{\theta}_1 = -1.00\ (0.2),$$
$$\hat{\theta}_2 = -2.00\ (0.3),$$
$$\hat{\theta}_3 = -0.50\ (0.4),$$
$$\hat{\theta}_4 = -1.00\ (0.1),$$
$$\hat{\theta}_5 = +1.00\ (0.5),$$

and inverse negative Hessian of the likelihood function equal to the matrix

$$-\mathcal{I}_{\hat{\theta}}^{-1} = \begin{pmatrix} 0.04 & 0.05 & 0.07 & 0.02 & 0.09 \\ 0.05 & 0.09 & 0.09 & 0.02 & 0.14 \\ 0.07 & 0.09 & 0.16 & 0.03 & 0.19 \\ 0.02 & 0.02 & 0.03 & 0.01 & 0.04 \\ 0.09 & 0.14 & 0.19 & 0.04 & 0.25 \end{pmatrix}$$

c) Provide the estimate and standard error of $\hat{\rho}_{12}$ and of $\hat{\sigma}_{11}$. Include your derivations.

d) Describe how you would test whether the two means μ_1 and μ_2 are equal.

e) How would you build a joint confidence set for the subvector $(\theta_1, \theta_2, \theta_5)$?

Question 4.
You observe 20 firm characteristics of 500,000 companies.
You decide to run a PCA and a factor analysis on your data.

a) Describe the differences between these two.

b) Write down the likelihood function for your factor model.

Using your factor model estimates with 3 factors, you obtain that the first factor loads primarily on the first 5 firm characteristics.

c) Explain why you cannot (without further motivation) conclude from this that the first factor reflects the average of the first 5 characteristics.