vrije Universiteit amsterdam

School of Business and Economics

Exam:	Multivariate Statistics
Code:	E_EOR2_MS
Examinator: Co-reader:	Andre Lucas 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.

The normal has the whole real line as support. This is not sensible for variables like asset size or long term debt, which cannot go negative. [The problem will be limited if the mean is sufficiently high and the variance sufficiently small, as the probability of becoming negative then is negligible in practice.]

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.

Define the mean $\mu = (\mu_1, \mu_2, \mu_3)^\top$ holding the means of the three features $X = (X_1, X_2, X_3)^\top$, respectively, with realization(s) $x = (x_1, x_2, x_3)^\top$. Also define the 3x3 covariance matrix Σ. The pdf is now given by

$$f_X(x) = \frac{1}{|2\pi\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^\top\Sigma^{-1}(x-\mu)\right)$$

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.

Define the transformation $(Y_1, Y_2, Y_3) = (X_1/X_2, \log(X_2), X_3/X_2) \leftrightarrow (X_1, X_2, X_3) = (Y_1 \exp(Y_2), \exp(Y_2), Y_3 \exp(Y_2))$. The jacobian is given by

$$abs(|\partial X/\partial Y^\top|) = abs\begin{vmatrix} \exp(Y_2) & Y_1\exp(Y_2) & 0 \\ 0 & \exp(Y_2) & 0 \\ 0 & Y_3\exp(Y_2) & \exp(Y_2) \end{vmatrix} = \exp(3Y_2).$$

There is no domain change: all variables can be all over the real line given the normality assumption for X. There is a problem for Y_2 as it is not defined for $X_2 < 0$. Economically, there is a problem for Y_3 as well, as it will be between zero and one. We ignore these two problems. We now use the transformation of variables theorem and obtain

$$f_Y(y) = f_X(y_1\exp(y_2), \exp(y_2), y_3\exp(y_2)) \cdot \exp(3y_2) =$$

$$\frac{\exp(3y_2)}{|2\pi\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\left(\begin{pmatrix} y_1\exp(y_2) \\ \exp(y_2) \\ y_3\exp(y_2) \end{pmatrix} - \mu\right)^\top \Sigma^{-1}\left(\begin{pmatrix} y_1\exp(y_2) \\ \exp(y_2) \\ y_3\exp(y_2) \end{pmatrix} - \mu\right)\right).$$

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 $(X_1 + X_3)|(X_1 + X_2)$.

Combining the first two lines using Th5.4, we get

$$\begin{pmatrix} X_1 \\ X_3 \end{pmatrix} \sim N\left(\begin{pmatrix} 3 \\ 3 \end{pmatrix}, \begin{pmatrix} 5 & 3 \\ 3 & 3 \end{pmatrix}\right).$$

Combining this with the third line and using Th5.4, we get

$$\begin{pmatrix} X_2 \\ X_1 \\ X_3 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 3 \\ 3 \end{pmatrix}, \begin{pmatrix} 6 & 2 & 0 \\ 2 & 5 & 3 \\ 0 & 3 & 3 \end{pmatrix}\right), \quad X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N\left(\begin{pmatrix} 3 \\ 0 \\ 3 \end{pmatrix}, \begin{pmatrix} 5 & 2 & 3 \\ 2 & 6 & 0 \\ 3 & 0 & 3 \end{pmatrix}\right).$$

Defining

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} X \sim N\left(\begin{pmatrix} 6 \\ 3 \end{pmatrix}, \begin{pmatrix} 14 & 10 \\ 10 & 15 \end{pmatrix}\right),$$

so

$$(Y_1|Y_2) \sim N\left(6 + \frac{10}{15}(Y_2 - 3), 14 - \frac{100}{15}\right) = N\left(4 + \frac{2}{3}Y_2, 7\frac{1}{3}\right).$$

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?

The answer is the conditional expectation, so

$$E[X_1|X_2, X_3] = 3 + \begin{pmatrix} 2 & 3 \end{pmatrix} \begin{pmatrix} 6 & 0 \\ 0 & 3 \end{pmatrix}^{-1} \begin{pmatrix} X_2 \\ X_3 - 3 \end{pmatrix} = \frac{1}{3}X_2 + X_3.$$

The precision is given by the conditional variance:

$$5 - \begin{pmatrix} 2 & 3 \end{pmatrix} \begin{pmatrix} 6 & 0 \\ 0 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = 5 - \frac{2}{3} - 3 = \frac{4}{3}.$$

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) $\tilde{C}(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2)$.

a.1) not a copula. For instance, $\tilde{C}(1,1) = 1 - C(1,1) = 1 - 1 = 0$, whereas it should be 1.

a.2) not a copula. It is not grounded. $\tilde{C}(u_1, 0) = 1 - C(1 - u_1, 1) = 1 - (1 - u_1) = u_1 \neq 0$.

a.3) is a copula. The copula density is the same, so it is 2-increasing. It is grounded, $\tilde{C}(0, u_2) = u_2 - 1 + C(1, 1 - u_2) = u_2 - 1 + 1 - u_2 = 0$ and similarly for $\tilde{C}(u_1, 0)$. And it has uniform marginals: $\tilde{C}(1, u_2) = u_2 + C(0, 1 - u_2) = u_2$ (because C is grounded, as it is a copula). Similarly, $\tilde{C}(u_1, 1) = u_1$.

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 $X = (X_1, X_2)$.

We first compute the PITs.
$u_1 = F_{X_1}(2) = 1 - \exp(-0.2) \approx 0.181$ and $u_2 = 1 - \exp(-0.2) \approx 0.181$.
Then, using Sklar's theorem, we know

$$P[X_1 \leq 2, X_2 \leq 1] = C(0.181, 0.181) = \frac{0.181^2}{1 - 0.5 \cdot 0.819^2} = 0.049.$$

For the pdf, we get

$$c(u_1, u_2) = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2} = \frac{\partial}{\partial u_2}\left[\frac{u_2}{1 - 0.5(1 - u_1)(1 - u_2)} - \frac{0.5 u_1 u_2 (1 - u_2)}{(1 - 0.5(1 - u_1)(1 - u_2))^2}\right] =$$

$$\frac{1}{1 - 0.5(1 - u_1)(1 - u_2)} - \frac{0.5 u_2 (1 - u_1)}{(1 - 0.5(1 - u_1)(1 - u_2))^2} - \frac{0.5 u_1 (1 - 2u_2)}{(1 - 0.5(1 - u_1)(1 - u_2))^2} + \frac{0.5 u_1 (1 - u_1) u_2 (1 - u_2)}{(1 - 0.5(1 - u_1)(1 - u_2))^3}$$

$$f_X(2,1) = 0.1 \exp(-0.2) \cdot 0.2 \exp(-0.2) \cdot c(0.181, 0.181) =$$

$$= 0.02 \exp(-0.4) \cdot \left[\frac{1}{0.6646} - \frac{0.5 \cdot 0.181 \cdot 0.819}{0.6646^2} - \frac{0.5 \cdot 0.181 \cdot 0.638}{0.6646^2} + \frac{0.5 \cdot 0.181^2 \cdot 0.819^2}{0.6646^3}\right]$$

$$\approx 0.0167.$$

$$P[X_1 \leq 2 | X_2 = 1] = C(u_1 | u_2) = \frac{\partial C(u_1, u_2)}{\partial u_2} = \left[\frac{u_1}{1 - 0.5(1 - u_1)(1 - u_2)} - \frac{0.5 u_1 u_2 (1 - u_1)}{(1 - 0.5(1 - u_1)(1 - u_2))^2}\right]$$

$$= \frac{0.181}{0.6646} - \frac{0.5 \cdot 0.181^2 \cdot 0.819}{0.6646^2} \approx 0.242.$$

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?

Because you want the second moment (or the variance) to exist. For instance, economically it may not make sense to expect infinite variance of for instance stock returns.

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

Nothing. Marginal behavior can be completely separate from the dependence behavior due to Sklar's theorem.

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 $\theta_1, \ldots, \theta_5$:

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

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

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.

Well, the first mean and second mean are now negative by design, so that is fine.
Also, $\sigma_{11} > \sigma_{22}$ by design, so that is also good.
Finally, the correlation is now between 0 and 1. That is in general too restrictive, because a correlation might also be negative. But here we expect the correlation to be positive (as is stated in the exercise), so the transformation does precisely what we want.

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

Define

$$\Sigma(\theta) = \begin{pmatrix} \exp(\theta_3) + \exp(\theta_4) & \sqrt{\exp(\theta_3 + \theta_4) + \exp(2\theta_4)}\,(1 + \exp(-\theta_5))^{-1} \\ \sqrt{\exp(\theta_3 + \theta_4) + \exp(2\theta_4)}\,(1 + \exp(-\theta_5))^{-1} & \exp(\theta_4) \end{pmatrix}$$

and

$$\mu(\theta) = \begin{pmatrix} -\exp(\theta_1) \\ -\exp(\theta_2) \end{pmatrix}.$$

Then

$$\ell(\theta) = -\frac{n}{2}\log|2\pi\Sigma| - \frac{1}{2}\sum_{i=1}^{n}\left(x_i - \mu(\theta)\right)^\top \Sigma(\theta)^{-1}\left(x_i - \mu(\theta)\right) \quad .$$

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

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

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

For $\hat{\theta}_5 = 1$, we have $\hat{\rho}_{12} = (1 + e^{-1})^{-1} \approx 0.731$. For the standard error we have

$$G = \frac{\partial \rho_{12}}{\theta^\top} = \left(0,0,0,0,\rho_{12}(1 - \rho_{12})\right), \quad \hat{G} = (0,0,0,0,0.1966)$$

and thus

$$\sqrt{G\left(-\mathcal{J}_{\hat{\theta}}^{-1}\right)G^\top} = \sqrt{0.25 \cdot 0.1966^2} \approx 0.098.$$

For $\hat{\theta}_3 = -0.5, \hat{\theta}_4 = -1$, we have $\hat{\sigma}_{11} \approx 0.6065 + 0.3679 = 0.9744$. For the standard error we have

$$G = \frac{\partial \sigma_{11}}{\theta^\top} = \left(0,0,\exp(\theta_3), \exp(\theta_4), 0\right), \quad \hat{G} = (0,0,0.6065,0.3679,0)$$

and thus

$$\sqrt{G\left(-\mathcal{J}_{\hat{\theta}}^{-1}\right)G^\top} = \sqrt{0.16 \cdot 0.6065^2 + 0.01 \cdot 0.3679^2 + 2 \cdot 0.03 \cdot 0.6065 \cdot 0.3679} \approx 0.27.$$

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

Two ways.
Way one: test whether $\theta_1 - \theta_2 = \mathcal{A}\theta = 0$ for $\mathcal{A} = (1, -1, 0, 0, 0)$, and use the statistic

$$(\mathcal{A}\hat{\theta})^{\top} \left(\mathcal{A}^{\top}(-\mathcal{I}_{\hat{\theta}}^{-1})\mathcal{A}\right)^{-1} (\mathcal{A}\hat{\theta}) \sim \chi_1^2.$$

Alternatively, you could directly test on μ, but then you would first need to compute the covariance matrix of $\hat{\mu}$ using the delta method.
The result would be of the form

$$(\mathcal{A}\hat{\mu})^{\top} \left(\mathcal{A}^{\top} G(-\mathcal{I}_{\hat{\theta}}^{-1}) G^{\top} \mathcal{A}\right)^{-1} (\mathcal{A}\hat{\mu}) \sim \chi_1^2.$$

For $\mu = (\mu_1, \mu_2)^{\top}$, $\mathcal{A} = (1, -1)$, and $G = \partial\mu / \partial\theta^{\top}$.

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

$$\left\{ (\theta_1, \theta_2, \theta_3) \in \mathbb{R}^3 \,\middle|\, \left((\theta_1, \theta_2, \theta_3) - (-1, -2, 1)\right) \begin{pmatrix} 0.04 & 0.05 & 0.09 \\ 0.05 & 0.09 & 0.14 \\ 0.09 & 0.14 & 0.25 \end{pmatrix} \left(\begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} - \begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix} \right) \leq \chi^2_{3,1-\alpha} \right\}.$$

For significance level α.

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.

PCA is a compression technique, summarizing the 20 characteristics into a smaller set of features for all companies. It is not model based. It has uniquely defined factors.

Factor analysis is model-based, such that we can estimate by maximum likelihood. The parameters, however, are not identified fully. In particular, we can still rotate the factor loadings as we like. Also, we explicitly model the features error terms after taking out the common factors and estimate the error variances.

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

$$X = \mu + QF + E, \quad F \sim N(0, \mathcal{I}_k), \quad E \sim N(0, \Psi),$$

then we have

$$\ell(\mu, Q, \Psi) = -\frac{n}{2} \log|2\pi(QQ^{\top} + \Psi)| - \frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^{\top} (QQ^{\top} + \Psi)^{-1} (x_i - \mu).$$

Normalization restrictions like $Q^{\top}\Psi^{-1}Q = \mathcal{I}_k$ are typically imposed to get a unique likelihood maximum.
(You can also write down the shorter version by picking $\hat{\mu} = \bar{x}$.)

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.

Reason 1: the weights may not be equal 😉
Reason 2: the factors can still be rotated at will; they are only identified up to rotation.