STATISTICAL METHODS FOR FINANCE: EXAMINATION SOLUTIONS, 2014.

Q1. With $\ell(\theta)$ the log-likelihood, the *score function* is

$$s := \ell'; \qquad [\mathbf{2}]$$

the *information per reading* is

$$I(\theta) := E[\{\ell'(\theta)\}^2] = -E[\ell''(\theta)] : \quad I(\theta) = E[s^2(\theta)] = E[-s'(\theta)]. \qquad [\mathbf{2}]$$

In the example given, write $v := \sigma^2$.

$$\ell(v) = \log f = const - \frac{1}{2}\log v - \frac{1}{2}(X-\mu)^2/v,$$

$$s(v) := \ell'(v) = -\frac{1}{2v} + \frac{(X-\mu)^2}{2v^2},$$

$$s'(v) = \frac{1}{2v^2} - \frac{(X-\mu)^2}{v^4}.$$

The information per reading is

$$I = I(v) = E[-s'(v)] = -\frac{1}{2v^2} + \frac{E[(X-\mu)^2]}{v^3} = -\frac{1}{2v^2} + \frac{v}{v^3} = \frac{1}{2v^2}. \qquad [\mathbf{8}]$$

The CR bound is

$$1/(nI) = 2v^2/n. \qquad [\mathbf{2}]$$

Write

$$S_0^2 := \frac{1}{n}\sum_1^n (X_i - \mu)^2. \qquad [\mathbf{2}]$$

Then

$$nS_0^2/\sigma^2 \sim \chi^2(n)$$

(definition of $\chi^2(n)$), so has mean n and variance $2n$ – because $\chi^2(1)$ has mean 1 ('normal variance') and variance 2 (by an MGF calculation or from memory). So S_0^2 has mean σ^2 (so is unbiased for σ^2), and variance $2n.\sigma^4/n^2 = 2v^2/n$, the CR bound above, so is efficient for $v = \sigma^2$. [**4**]
Seen – lectures (bookwork) and problems (example).

Q2. (i) Markowitz' work of 1952 (which led on to CAPM in the 1960s) gave two key insights:
(a). *Think of risk and return together, not separately.* Now return corresponds to mean (= mean rate of return), risk corresponds to variance – hence *mean-variance* analysis, *efficient frontier*, etc. – maximise return for a given level of risk/minimise risk for a given return rate). **[2]**
(b). *Diversify* (don't 'put all your eggs in one basket'). Hold a *balanced portfolio* – a range of risky assets, with lots of *negative correlation* – so that when things change, losses on some assets will be offset by gains on others. **[2]**
Hence the vector-matrix parameter (μ, Σ) is accepted as an essential part of any model in mathematical finance.
(ii) *Elliptical distributions.*

The normal density in higher dimensions is a multiple of $\exp\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\}$, where the matrices Σ, Σ^{-1} are *positive definite* (PD), so the contours $(x-\mu)^T\Sigma^{-1}(x-\mu) = \text{const.}$ are *ellipsoids*. The general *elliptically contoured* distribution has a density

$$f(x) = const.g(x-\mu)^T\Sigma^{-1}(x-\mu)).$$

This is a *semi-parametric* model, where $\theta := (\mu, \sigma)$ is the parametric part and the *density generator* g is the non-parametric part. **[4]**
(iii) *Normal (Gaussian) model*: elliptically contoured ($g(.) = e^{-\frac{1}{2}\cdot}$). Though very useful, it has various deficiencies, e.g.:
(a) It is *symmetric*. Many financial data sets show asymmetry, or *skew*. This reflects the asymmetry between profit and loss. Big profits are nice; big losses can be lethal (to the firm – bankruptcy). **[3]**
(b) It has extremely thin tails. Most financial data sets have tails that are *much fatter* than the ultra-thin normal tails. **[3]**
(iv) For asset returns (= profit/loss over initial asset price) over a period, the *return period*: matters vary dramatically with the return period.
(a) For *long* return periods (monthly, say – the Rule of Thumb is that 16 trading days suffice), the CLT applies, and asset returns are approximately *normal* ('aggregational Gaussianity). **[2]**
(ib) For *intermediate* return periods (daily, say), a commonly used model is the *generalised hyperbolic (GH)* – log-density a hyperbola, with linear asymptotes, so density decays like the exponential of a linear function). **[2]**
(c) For *high-frequency* returns ('tick data', say – every few seconds), the density typically decays like a power (as with the Student t distribution). **[2]**
Seen – lectures.

Q3. For the multivariate normal $N_p(\mu, \Sigma)$, $\bar{x}$ and S are the maximum likelihood estimators for μ, Σ.

Proof. Write $V = (v_{ij}) := \Sigma^{-1}$. The likelihood is given as

$$L = const.|V|^{n/2} \exp\{-\frac{1}{2}n\ trace(VS) - \frac{1}{2}n(\bar{x}-\mu)^T V(\bar{x}-\mu)\},$$

so the log-likelihood is

$$\ell = c + \frac{1}{2}n \log|V| - \frac{1}{2}n\ trace(VS) - \frac{1}{2}n(\bar{x}-\mu)^T V(\bar{x}-\mu). \qquad [\mathbf{2}]$$

The MLE $\hat{\mu}$ for μ is $\bar{x}$, as this reduces the last term (the only one involving μ) to its minimum value, 0. **[3]**

For a square matrix $A = (a_{ij})$, its determinant is

$$|A| = \sum_j a_{ij}A_{ij}\ \forall i, \quad \text{or} \quad |A| = \sum_i a_{ij}A_{ij}\ \forall j, \qquad [\mathbf{2}]$$

expanding by the ith row or jth column, where A_{ij} is the *cofactor* (signed minor) of a_{ij}. From either,

$$\partial|A|/\partial a_{ij} = A_{ij}, \quad \text{so} \quad \partial \log|A|/\partial a_{ij} = A_{ij}/|A| = (A^{-1})_{ji},$$

the (j,i) element of A^{-1}, recalling the formula for the matrix inverse (or $(A^{-1})_{ij}$ if A is symmetric). **[2]**

Also, if B is symmetric,

$$trace(AB) = \sum_i \sum_j a_{ij}b_{ji} = \sum_{i,j} a_{ij}b_{ij} : \qquad \partial\ trace(AB)/\partial a_{ij} = b_{ij}. \quad [\mathbf{2}]$$

Using these, and writing $S = (s_{ij})$,

$$\partial \log|V|/\partial v_{ij} = (V^{-1})_{ij} = (\Sigma)_{ij} = \sigma_{ij} \qquad (V := \Sigma^{-1}), \qquad [\mathbf{2}]$$

$$\partial\ trace(VS)/\partial v_{ij} = s_{ij}. \qquad [\mathbf{2}]$$

So

$$\partial\ell/\partial v_{ij} = \frac{1}{2}n(\sigma_{ij} - s_{ij}), \qquad [\mathbf{2}]$$

which is 0 for all i and j iff $\Sigma = S$. This says that S is the MLE for Σ: $\hat{\Sigma} = S$, as required. // **[3]**

Seen – lectures.

Q4. From the model equation

$$y_i = \sum\nolimits_{j=1}^{p} a_{ij}\beta_j + \epsilon_i, \qquad \epsilon_i \quad iid \quad N(0, \sigma^2),$$

the likelihood and log-likelihood are

$$\begin{aligned} L &= \frac{1}{\sigma^n 2\pi^{\frac{1}{2}n}} . \prod\nolimits_{i=1}^{n} \exp\{-\frac{1}{2}(y_i - \sum\nolimits_{j=1}^{p} a_{ij}\beta_j)^2/\sigma^2\} \\ &= \frac{1}{\sigma^n 2\pi^{\frac{1}{2}n}} . \exp\{-\frac{1}{2}\sum\nolimits_{i=1}^{n}(y_i - \sum\nolimits_{j=1}^{p} a_{ij}\beta_j)^2/\sigma^2\}, \end{aligned}$$

$$\ell := \log L = const - n\log\sigma - \frac{1}{2}[\sum\nolimits_{i=1}^{n}(y_i - \sum\nolimits_{j=1}^{p} a_{ij}\beta_j)^2]/\sigma^2. \qquad (*) \quad [\mathbf{5}]$$

Maximise w.r.t. β_r in $(*)$ (Fisher, MLE) – equivalently, minimise [...]: $\partial\ell/\partial\beta_r = 0$ (Least Squares):

$$\sum\nolimits_{i=1}^{n} a_{ir}(y_i - \sum\nolimits_{j=1}^{p} a_{ij}\beta_j) = 0 \qquad (r = 1, \ldots, p):$$

$$\sum\nolimits_{j=1}^{p}(\sum\nolimits_{i=1}^{n} a_{ir}a_{ij})\beta_j = \sum\nolimits_{i=1}^{n} a_{ir}y_i.$$

Write $C = (c_{ij})$ for the $p \times p$ matrix $C := A^T A$, which we note is *symmetric*: $C^T = C$. Then

$$c_{ij} = \sum\nolimits_{k=1}^{n}(A^T)_{ik}A_{kj} = \sum\nolimits_{k=1}^{n} a_{ki}a_{kj}. \qquad [\mathbf{5}]$$

So this says

$$\sum\nolimits_{j=1}^{p} c_{rj}\beta_j = \sum\nolimits_{i=1}^{n} a_{ir}y_i = \sum\nolimits_{i=1}^{n}(A^T)_{ri}y_i.$$

In matrix notation, this is

$$(C\beta)_r = (A^T y)_r \qquad (r = 1, \ldots, p): \qquad C\beta = A^T y, \qquad C := A^T A. \qquad (NE) \quad [\mathbf{4}]$$

These are the *normal equations*. As A ($n \times p$, with $p << n$) has full rank, A has rank p, so $C := A^T A$ has rank p, so is non-singular. So the normal equations have solution

$$\hat{\beta} = C^{-1}A^T y = (A^T A)^{-1}A^T y. \qquad [\mathbf{3}]$$

Multiplying both sides by A, with P the *projection matrix* $P := AC^{-1}A^T = A(A^T A)^{-1}A^T$,

$$Py = A(A^T A)^{-1}A^T y = A\hat{\beta}. \qquad [\mathbf{3}]$$

Seen – lectures.

Q5. *ARMA(1,1)*: $X_t = \phi X_{t-1} + \epsilon_t + \theta\epsilon_{t-1}$: $(1-\phi B)X_t = (1+\theta B)\epsilon_t$.
Condition for Stationarity: $|\phi| < 1$ (assumed). **[2]**
Condition for Invertibility: $|\theta| < 1$ (assumed). **[2]**

$$X_t = (1-\phi B)^{-1}(1+\theta B)\epsilon_t = (1+\theta B)(\sum\nolimits_0^\infty \phi^i B^i)\epsilon_t$$

$$= \epsilon_t + \sum\nolimits_1^\infty \phi^i B^i \epsilon_t + \theta \sum\nolimits_0^\infty \phi^i B^{i+1}\epsilon_t = \epsilon_t + (\theta+\phi)\sum\nolimits_1^\infty \phi^{i-1}B^i\epsilon_t :$$

$$X_t = \epsilon_t + (\phi+\theta)\sum\nolimits_{i=1}^\infty \phi^{i-1}\epsilon_{t-i}. \qquad \textbf{[4]}$$

Variance: lag $\tau = 0$. Square and take expectations: ϵs iid $N(0,\sigma^2)$, so

$$\gamma_0 = var X_t = E[X_t^2] = \sigma^2 + (\phi+\theta)^2 \sum\nolimits_1^\infty \phi^{2(i-1)}\sigma^2$$

$$= \sigma^2 + \frac{(\phi+\theta)^2\sigma^2}{(1-\phi^2)} = \sigma^2(1-\phi^2+\phi^2+2\phi\theta+\theta^2)/(1-\phi^2) :$$

$$\gamma_0 = \sigma^2(1+2\phi\theta+\theta^2)/(1-\phi^2). \qquad \textbf{[5]}$$

Covariance: lag $\tau \geq 1$. $\quad X_{t-\tau} = \epsilon_{t-\tau} + (\phi+\theta)\sum_{j=1}^\infty \phi^{j-1}\epsilon_{t-\tau-j}$.
Multiply the series for X_t and $X_{t-\tau}$ and take expectations:

$$\gamma_\tau = cov(X_t, X_{t-\tau}) = E[X_t X_{t-\tau}],$$

$$= E\{[\epsilon_t + (\phi+\theta)\sum\nolimits_{i=1}^\infty \phi^{i-1}\epsilon_{t-i}].[\epsilon_{t-\tau} + (\phi+\theta)\sum\nolimits_{j=1}^\infty \phi^{j-1}\epsilon_{t-\tau-j}]\}.$$

The ϵ_t-term in the first [.] gives no contribution. The i-term in the first [.] for $i = \tau$ and the $\epsilon_{t-\tau}$ in the second [.] give $(\phi+\theta)\phi^{\tau-1}\sigma^2$. The product of the i term in the first sum and the j term in the second contributes for $i = \tau + j$; for $j \geq 1$ it gives $(\phi+\theta)^2\phi^{\tau+j-1}.\phi^{j-1}.\sigma^2$. So

$$\gamma_\tau = (\phi+\theta)\phi^{\tau-1}\sigma^2 + (\phi+\theta)^2\phi^\tau\sigma^2\sum\nolimits_{j=1}^\infty \phi^{2(j-1)}.$$

The geometric series is $1/(1-\phi^2)$ as before, so for $\tau \geq 1$

$$\gamma_\tau = \frac{(\phi+\theta)\phi^{\tau-1}\sigma^2}{(1-\phi^2)}.[1-\phi^2+\phi(\phi+\theta)] : \qquad \gamma_\tau = \sigma^2(\phi+\theta)(1+\phi\theta)\phi^{\tau-1}/(1-\phi^2).$$

[5]

Autocorrelation. The autocorrelation $\rho_\tau := \gamma_\tau/\gamma_0$ is thus

$$\rho_0 = 1, \qquad \rho_\tau = \frac{(\phi+\theta)(1+\phi\theta)}{(1+2\phi\theta+\theta^2)}.\phi^{\tau-1} \qquad (\tau \geq 1). \qquad \textbf{[2]}$$

Seen – lectures.

Q6. The *multivariate normal distribution* $N_n(\mu, \Sigma)$ in n dimensions with mean vector μ and covariance matrix Σ has characteristic function $\phi(t) = \exp\{-\mu^T t - \frac{1}{2} t^T \Sigma t\}$, and by Edgeworth's theorem, when Σ is positive definite (invertible), has density

$$f(\mathbf{x}) := \frac{1}{(2\pi)^{\frac{1}{2}n} |\mathbf{\Sigma}|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\mathbf{x} - \mu)^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \mu)\}. \qquad \textbf{[3]}$$

If

$$y|u \sim N(X\beta + Zu, R), \qquad u \sim N(0, D),$$

$$f(y,u) = f(y|u)f(u) = const. \exp\{-\frac{1}{2}(y - X\beta - Zu)^T R^{-1} (y - X\beta - Zu)\}. \exp\{-\frac{1}{2} u^T D^{-1} u\}.$$

This is multinormal (has the functional form in u of Edgeworth's theorem). So $u|y$ is also multinormal (conditioning a multinormal on a subvector gives a multinormal):

$$f(u|y) \sim N(\Sigma, \nu), \qquad \textbf{[3]}$$

say. As

$$f(u|y) = f(y,u)/f(y)$$

by Bayes' theorem, we can identify *which* multinormal $u|y$ is by picking out the quadratic term in u and the linear term in u (in $f(y,u)$: $f(y) = \int f(y|u) du$ does not involve u) and using Edgeworth's theorem. **[2]**

The quadratic term in u is

$$-\frac{1}{2}[u^T Z^T R^{-1} Z u + u^T D^{-1} u], = -\frac{1}{2} u^T \Sigma^{-1} u, \qquad \Sigma := (Z^T R^{-1} Z + D^{-1})^{-1}.$$

[3]

The linear term in u is

$$-u^T Z^T R^{-1} (y - X\beta), = -u^T \Sigma^{-1} \nu,$$

$$\nu := \Sigma Z^T R^{-1} (y - X\beta) = (Z^T R^{-1} Z + D^{-1})^{-1} Z^T R^{-1} (y - X\beta): \qquad \textbf{[3]}$$

$$u|y \sim N(\Sigma, \nu), \quad \Sigma = (Z^T R^{-1} Z + D^{-1})^{-1}, \quad \nu = (Z^T R^{-1} Z + D^{-1})^{-1} Z^T R^{-1} (y - X\beta).$$

[3]

Application: 1. *Financial.* Here y is the response variable (output, profit, market share etc.); $X\beta$ represents the *fixed effects* (macro-economic variables – interest rates, trade figures etc.); Zu represents the *random effects* (firm-specific aspects – firms differ). **[3]**

2. *Educational.* Response: performance; fixed effects: teaching methods, etc.; random effects: the pupils – pupils differ.

Seen – lectures.