SMF EXAMINATION SOLUTIONS 2015-16

Q1. *Markowitz; elliptical distributions; return periods.*
(i) Markowitz' work of 1952 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]**
So (μ, Σ) is accepted as necessary.
(ii) *Elliptical distributions.*

The normal density $f(x)$ in higher dimensions is a multiple of $\exp\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\}$, with Σ, Σ^{-1} *positive definite* (PD), so the contours $f =$ const. *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]**
(b) 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.]

Q2. *Lognormal distribution; normal means.*

X has the *log-normal* distribution with parameters μ and σ, $X \sim LN(\mu, \sigma)$, if $Y := \log X \sim N(\mu, \sigma)$. **[2]**
The MGF of Y is $M_Y(t) := E[e^{tY}] = \exp\{\mu t + \frac{1}{2}\sigma^2 t^2\}$: $M_Y(1) = E[e^Y] = \exp\{\mu + \frac{1}{2}\sigma^2\}$.
But $e^Y = X$: $E[X] = \exp\{\mu + \frac{1}{2}\sigma^2\}$: $LN(\mu, \sigma)$ has mean $\exp\{\mu + \frac{1}{2}\sigma^2\}$. **[3]**

In *geometric Brownian motion (GBM)*, as in the *Black-Scholes model*, the price process $S = (S_t)$ of a risky asset is driven by the SDE

$$dS_t / S_t = \mu dt + \sigma dW_t, \qquad (GBM)$$

with $W = (W_t)$ Brownian motion/the Wiener process. This has solution

$$S_t = S_0 \exp\{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t\} :$$

$\log S_t$ is lognormally distributed. **[5]**

For a normal population $N(\mu, \sigma)$ with σ known: to test $H_0 : \mu = \mu_0$ v. $H_1 : \mu < \mu_0$. First, take any $\mu_1 < \mu_0$. To test H_0 v. $\mu = \mu_1$, by the Neyman-Pearson Lemma (NP), the best (most powerful) test uses test statistic the likelihood ratio (LR) $\lambda := L_0/L_1 = L(\mu_0)/L(\mu_1)$, where with data $x_1, \ldots, x_n$

$$L(\mu) = \sigma^{-n}(2\pi)^{-\frac{1}{2}n} \exp\{-\frac{1}{2}\sum_1^n (x_i - \mu)^2/\sigma^2\},$$

and critical region R of the form $\lambda \leq$ const: *reject* H_0 if λ is *too small*. Here $\lambda = \exp\{-\frac{1}{2}[\sum(x_i - \mu_0)^2 - \sum(x_i - \mu_1)^2]\}$. Forming the LR λ, the constants cancel, so R has the form $\log\lambda \leq$ const, or $-2\log\lambda \geq$ const. Expanding the squares, the $\sum x_i^2$ terms cancel, so (as $\sum x_i = n\bar{x}$) this is

$$-2\mu_0 n\bar{x} + n\mu_0^2 + 2\mu_1 n\bar{x} - n\mu_1^2 \geq \text{const} : \quad 2(\mu_1 - \mu_0)\bar{x} + (\mu_0^2 - \mu_1^2) \geq \text{const}.$$

As $\mu_1 < \mu_0$, this is $\bar{x} \leq c$. At significance level α, c is the lower α-point of the distribution of $\bar{x}$ under H_0. Then $\bar{x} \sim N(\mu_0, \sigma/\sqrt{n})$, so
$Z := (\bar{x} - \mu_0)\sqrt{n}/\sigma \sim \Phi = N(0, 1)$. If c_α is the lower σ-point of $\Phi = N(0, 1)$, i.e. of $Z := (\bar{x} - \mu_0)\sqrt{n}/\sigma$, $c_\alpha = (c - \mu_0)\sqrt{n}/\sigma$: $c = \mu_0 + \sigma c_\alpha/\sqrt{n}$. **[7]**

But this holds for *all* $\mu_1 < \mu_0$. So R is *uniformly most powerful* (UMP) for $H_0 : \mu = \mu_0$ (simple null) v. $H_1 : \mu < \mu_0$ (composite alternative). **[3]**
[Seen – lectures]

Q3. *Regression: towers.*
$y_A = a + \epsilon_1$, $y_B = b + \epsilon_2$, $y_{B-A} = -a + b + \epsilon_3$. So the regression model is $y = A\beta + \epsilon$, with y is the 3-vector of readings, ϵ the 3-vector of errors, $\beta := (a, b)^T$ the 2-vector of parameters and A the design matrix.
(a)

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 1 \end{pmatrix}; \qquad [\mathbf{3}]$$

$$C := A^T A = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}; \quad |C| = 3;$$

$$C^{-1} = \frac{1}{3}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, C^{-1}A^T = \frac{1}{3}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \end{pmatrix}.$$

[**3**]

(b)

$$P = \frac{1}{3}\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix}, \quad I - P = \frac{1}{3}\begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix}.$$

[**4**]

(c) The parameter estimates are

$$C^{-1}A^T y = \frac{1}{3}\begin{pmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \end{pmatrix}\begin{pmatrix} y_A \\ y_B \\ y_{B-A} \end{pmatrix} = \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix}:$$

$$\hat{a} = (2y_A + y_B - y_{B-A})/3, \quad \hat{b} = (y_A + 2y_B + y_{B-A})/3. \qquad [\mathbf{3}]$$

(d) The fitted values are $\hat{y} = Py$, which can be written in two ways:

$$\begin{pmatrix} (2y_A + y_B - y_{B-A})/3 \\ (y_A + 2y_B + y_{B-A})/3 \\ (-y_A + y_B + 2y_{B-A})/3) \end{pmatrix} = \begin{pmatrix} \hat{a} \\ \hat{b} \\ \hat{b} - \hat{a} \end{pmatrix}. \qquad [\mathbf{3}]$$

(e) As $n = 3$, $p = 2$ here, the ranks of P, $I - P$ are 2 and 1. [**2**]
(f) Write SSE, the *sum of squares for error*, for $y^T(I - P)y$. As $n - p = 1$, $\hat{\sigma}^2 = SSE/(n - p) = y^T(I - P)y$, which can be calculated by above. With numerical data, SSE is the sum of squared residuals, $\sum(y_i - \hat{y}_i)^2$. [**2**]
[Unseen; similar seen in lectures and problems]

Q4. *Regression plane.*

With two regressors u and v and response variable y, given a sample of size n of points $(u_1, v_1, y_1), \dots, (u_n, v_n, y_n)$ we have to fit a least-squares *plane* – that is, choose parameters a, b, c to minimise the sum of squares

$$SS := \sum_{i=1}^{n} (y_i - c - au_i - bv_i)^2.$$

Taking $\partial SS/\partial c = 0$ gives

$$\sum_{i=1}^{n} (y_i - c - au_i - bv_i) = 0 : \qquad c = \bar{y} - a\bar{u} - b\bar{v}.$$

$$SS = \sum_{i=1}^{n} [(y_i - \bar{y}) - a(u_i - \bar{u}) - b(v_i - \bar{v})]^2.$$

Then $\partial SS/\partial a = 0$ and $\partial SS/\partial b = 0$ give

$$\sum_{i=1}^{n} (u_i - \bar{u})[(y_i - \bar{y}) - a(u_i - \bar{u}) - b(v_i - \bar{v})],$$

$$\sum_{i=1}^{n} (v_i - \bar{v})[(y_i - \bar{y}) - a(u_i - \bar{u}) - b(v_i - \bar{v})].$$

Multiply out, divide by n to turn the sums into averages, and re-arrange:

$$as_{uu} + bs_{uv} = s_{yu},$$

$$as_{uv} + bs_{vv} = s_{yv}.$$

These are the *normal equations* (NE) for a and b. **[10]**

Condition for non-degeneracy. The determinant is

$$s_{uu}s_{vv} - s_{uv}^2 = s_{uu}s_{vv}(1 - r_{uv}^2)$$

(as $r_{uv} := s_{uv}/(s_u.s_v)$), $\neq 0$ iff $r_{uv} \neq \pm 1$, i.e., iff the (u_i, v_i) are not collinear, and this is the condition for (NE) to have a unique solution. **[4]**

Application: Grain futures.

The two principal factors affecting grain yields (apart from the weather near harvest – unpredictable!) are *sunshine* (in hours) and rainfall (in mm) during the *spring* growing season (known in advance). Using these as *predictor* variables u, v gives a best (linear unbiased) estimator of grain yield y.

The volumes of grain traded yearly are enormous. So, the ability to predict as accurately as possible the size of the summer harvest (and so, by supply and demand, its price), given information available in the spring, is very valuable. Such predictions can be used to form trading strategies for grain futures and grain options, etc. (example: the Great Grain Steal of 1972, by the then USSR, on the USA and Canada). **[6]**

[Seen: problem sheets]

Q5. *Yule-Walker equations and* $AR(2)$.

The $AR(p)$ model is (with (ϵ_t) white noise $WN(\sigma)$)

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t. \qquad \textbf{[2]}$$

Multiply by X_{t-k} and take E: as $E[X_{t-k}X_{t-i}] = \rho(|k-i|) = \rho(k-i)$,

$$\rho(k) = \phi_1 \rho(k-1) + \cdots + \phi_p \rho(k-p) \qquad (k > 0). \qquad (YW)$$

These are the *Yule-Walker equations.* **[4]**

They give a *difference equation* of *order* p, with *characteristic polynomial*

$$\lambda^p - \phi_1 \lambda^{p-1} - \cdots - \phi_p = 0.$$

If the roots are $\lambda_1, \cdots, \lambda_p$, the trial solution $\rho(k) = \lambda^k$ is a solution iff λ is one of the roots λ_i. Since the equation is linear,

$$\rho(k) = c_1 \lambda_1^k + \cdots + c_p \lambda_p^k$$

(for $k \geq 0$ and use $\rho(-k) = \rho(k)$ for $k < 0$) is a solution for all choices of constants c_i – the *general solution* of (YW) if all the roots λ_i are distinct. **[4]**

Example of an $AR(2)$ *process.*

$$X_t = \frac{1}{3} X_{t-1} + \frac{2}{9} X_{t-2} + \epsilon_t, \qquad (\epsilon_t) \quad WN. \qquad (1)$$

The Yule-Walker equations here are $\rho(k) = \phi_1 \rho(k-1) + \phi_2 \rho(k-2)$.

The characteristic polynomial is

$$\lambda^2 - \frac{1}{3}\lambda - \frac{2}{9} = 0 : \quad (\lambda - 2/3)(\lambda + 1/3) = 0; \quad \lambda_1 = 2/3, \ \lambda_2 = -1/3.$$

So as the roots are distinct, the autocovariance is $\rho(k) = a\lambda_1^k + b\lambda_2^k$. **[5]**

$k = 0$: $\rho(0) = 1$ gives $a + b = 1$: $b = 1 - a$. So $\rho(k) = a(2/3)^k + (1-a)(-1/3)^k$.

$k = 1$: $\rho(1) = \phi_1 \rho(0) + \phi_2 \rho(-1)$; as $\rho(0) = 1$ and $\rho(-1) = \rho(1)$, $\rho(1) = \phi_1/(1 - \phi_2)$. As here $\phi_1 = 1/3$ and $\phi_2 = 2/9$, this gives $\rho(1) = 3/7$. So

$$\rho(1) = 3/7 = a.(2/3) + (1-a).(-1/3).$$

That is,

$$(\frac{3}{7} + \frac{1}{3}) = a.(\frac{2}{3} + \frac{1}{3}) = a :$$

$a = (9 + 7)/21 = 16/21$. Thus

$$\rho(k) = \frac{16}{21}(\frac{2}{3})^k + \frac{5}{21}(\frac{-1}{3})^k. \qquad \textbf{[5]}$$

[Seen, lectures]

Q6. *Edgeworth's theorem; normal regression.*
(i) Edgeworth's theorem says that if $x \sim N(\mu, \Sigma)$ and $K := \Sigma^{-1}$,

$$f(x) \propto \exp\{-\frac{1}{2}(x-\mu)^T K (x-\mu)\}. \qquad [\mathbf{1}]$$

$$f(x_1, x_2) \propto \exp\{-\frac{1}{2}(x_1^T - \mu_1^T, x_2^T - \mu_2^T) \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}\},$$

giving (as a scalar is its own transpose, so the two cross-terms are the same)

$$\exp\{-\frac{1}{2}[(x_1^T - \mu_1^T)K_{11}(x_1 - \mu_1) + 2(x_1^T - \mu_1^T)K_{12}(x_2 - \mu_2) + (x_2^T - \mu_2^T)K_{22}(x_2 - \mu_2)]\}.$$

So

$$f_{1|2}(x_1|x_2) = f(x_1, x_2)/f_2(x_2)$$

$$\propto \exp\{-\frac{1}{2}[(x_1^T - \mu_1^T)K_{11}(x_1 - \mu_1) + 2(x_1^T - \mu_1^T)K_{12}(x_2 - \mu_2)]\}, \qquad (*)$$

treating x_2 here as a constant and x_1 as the argument of $f_{1|2}$. **[4]**
By Edgeworth's theorem again, if the conditional mean of $x_1|x_2$ is ν_1,

$$f_{1|2}(x_1|x_2) \propto \exp\{-\frac{1}{2}(x_1^T - \nu_1^T)V_{11}(x_1 - \nu_1)\}, \qquad (**)$$

for some matrix V_{11}. So $x_1|x_2$ is multinormal (as may be quoted). **[3]**

Equating coefficients of the quadratic term gives the conditional concentration matrix of $x_1|x_2$ as $V_{11} = K_{11}$:

$$conc(x_1|x_2) = K_{11}. \qquad [\mathbf{3}]$$

So the conditional covariance matrix is K_{11}^{-1}. Then equating linear terms in $(*)$ and $(**)$ gives the conditional mean:

$$x_1^T K_{11} \nu_1 = x_1^T K_{11} \mu_1 - x_1^T K_{12}(x_2 - \mu_2): \quad \nu_1 := E[x_1|x_2] = \mu_1 - K_{11}^{-1} K_{12}(x - \mu_2):$$

[3]

$$x_1|x_2 \sim N(\mu_1 - K_{11}^{-1}K_{12}(x - \mu_2), K_{11}^{-1}). \qquad [\mathbf{2}]$$

Using the quoted result for the inverse of a partitioned matrix gives

$$M = K_{11}, \qquad M^{-1} = K_{11}^{-1} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21},$$

$$K_{11}^{-1}K_{12} = M^{-1}(-MBD^{-1}) = -BD^{-1} = -\Sigma_{12}\Sigma_{22}^{-1}.$$

Combining,

$$x_1|x_2 \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}). \qquad [\mathbf{4}]$$

[Seen – Problems] NHB