Imperial College London

UNIVERSITY OF LONDON

BSc and MSci EXAMINATIONS (MATHEMATICS)

May-June 2006

This paper is also taken for the relevant examination for the Associateship.

M3S2/M4S2

Statistical Modelling

Date: Thursday, 1st June 2006 Time: 2 pm – 4 pm

Credit will be given for all questions attempted but extra credit will be given for complete or nearly complete answers.

Calculators may not be used.

Statistical tables will not be available.

1. The Capital Asset Pricing Model (CAPM) is used to model the return on single asset r_{it}, in terms of the overall market returns r_{mt}, and the return on a risk free security r_{ft}. The model is specified as

$$r_{it} - r_{ft} = \alpha_i + \beta_i \left(r_{mt} - r_{ft} \right) + \epsilon_{it}, \quad i = 1, \ldots, I, \quad t = 0, \ldots, N-1. \tag{1}$$

ϵ_{it} is the unsystematic (unpredictable) risk for security i.

(a) What assumptions are usually made on the expectation and covariance structure of ϵ_{it}? Discuss the validity of these assumptions, in particular in terms of the i and t indexing for the covariance.

(b) How do we interpret β_i? What judgement is usually made on r_{it} if $|\beta_i| > 1$?

(c) Take expectations on both sides of equation (1) and simplify the resulting expression.

(d) Formally write down the Sharpe-Lintner-Markowitz Hypothesis associated with the CAPM, in terms of a null and an alternative hypothesis. Give an interpretation of the hypothesis.

(e) Find the correlation of $r_{it} - r_{ft}$ and $r_{mt} - r_{ft}$, and discuss the role of β_i in this correlation.

(f) Find the ratio

$$\frac{\text{Cov}\left(r_{it} - r_{ft}, r_{mt} - r_{ft} \right)}{\text{Var}\left(r_{mt} - r_{ft} \right)},$$

and discuss its interpretation.

(g) Discuss how to test the Sharpe-Lintner-Markowitz Hypothesis for a given security based on a sample of $y_{it} = r_{it} - r_{ft}, \quad i = 1, \ldots, I, \ t = 0, \ldots, N-1$ and $z_t = r_{mt} - r_{ft}, \ \ t = 0, \ldots, N-1$ defining a least square estimator of $\boldsymbol{\beta} = \left[\alpha_i \quad \beta_i \right]^T$, denoted $\widehat{\boldsymbol{\beta}}$.

Hint: You may assume that an estimate of Cov($\widehat{\boldsymbol{\beta}}$) $= \boldsymbol{V}_N$, denoted $\widehat{\boldsymbol{V}}_N$, is given and that the usual large sample results hold, as long as you state them carefully before usage. Do NOT carry out any matrix calculations when giving the estimator.

2. Rates of return are studied to investigate the efficient market hypothesis (EMH). In particular, Hansen and Hodrick modelled the rate of return to speculation rr_t via a relationship between current rates and future values, and in particular for a fixed value of the lag k, a model for the k steps forward rate of return, is given by:

$$rr_{t+k} = \alpha_0 + \alpha_1 rr_t + \epsilon_{t+k}, \quad t = 0, \ldots, N-1,$$

Assume $\boldsymbol{\alpha} = [\alpha_0 \quad \alpha_1]^T$ is a vector of constants, whilst $\{\epsilon_t\}$ are random variables, with zero mean, and a covariance structure that may be non-negligible. Furthermore the rate of return can also be written as

$$rr_{t+k} = s_{t+k} - g_{t,k},$$

where s_{t+k} is the logarithm of the value of the asset at time $t+k$, and $g_{t,k}$ is the logarithm of the forward rate, set at time t for k steps into the future. Note that rr_t is a stationary process.

(a) Write down the general definition of a return.

(b) Describe, in words, the efficient market hypothesis (EMH).

(c) Express the null hypothesis corresponding to the EMH in the form

$$H_0: \quad \boldsymbol{A}\boldsymbol{\alpha} = \boldsymbol{a}_0$$

for suitably chosen matrix matrix $\boldsymbol{A}$ and vector $\boldsymbol{a}_0$. Hence, describe how to estimate $\boldsymbol{\alpha}$ using ordinary least squares, specifying design matrix $\boldsymbol{\mathcal{X}}$ and response vector $\boldsymbol{\mathcal{Y}}$. Discuss the statistical properties of this estimator under relevant assumptions on $\{\epsilon_t\}$.

Do NOT carry out any detailed matrix calculations.

(d) Suppose that $\widehat{\boldsymbol{V}}$ is an estimator of Cov$\{\boldsymbol{A}\boldsymbol{\alpha}\}$. Describe how to test the EMH.

3. Returns are frequently modelled as non-Gaussian, and various tests have been proposed to test for different non-Gaussian features. In particular we consider *heavy-tailed* data. Assume that marginally, for any realisation X_t from the distribution $F(x)$ we have

$$P\left(X_t > x\right) = C^{(\alpha)} x^{-\alpha}, \; x >> 0,$$

where $C^{(\alpha)}$ is a constant.

(a) Derive an expression for $F(x)$ in terms of $C^{(\alpha)}$ and α.

(b) Construct the empirical estimator, $\widehat{F}(x)$, of $F(x)$ based on an independent sample $X_1, \ldots, X_N$. You may denote the order statistics of the sample as $X_{(1)}, \ldots, X_{(N)}$.

(c) Derive an estimator of α using $\widehat{F}(x)$, the expression found in part (a), and the last s order statistics of the sample. Discuss the appropriate choice of s.

(d) Using the estimator for α and the expression found in part (a), write down an estimator for $C^{(\alpha)}$.

(e) It can be shown that

$$\sqrt{s}\left(\widehat{\alpha} - \alpha\right) \; \stackrel{\sim}{\sim} N(0, \alpha^2).$$

Describe how to construct a 95% confidence interval for α.

4. The log absolute returns for an exchange rate are modelled as

$$R_t \equiv \log\left(|r_t - \mathsf{E}\{r_t\}|\right) = \log(\sigma_t) + \log|s_t| + \xi_t, \quad t = 0, 1, \ldots, N-1,$$

where $\xi_t \sim N(0,1)$ are independent random variables, σ_t are potentially persistent volatility components, and s_t are public regularly scheduled announcements. The objective is to estimate $\log|s_t|$ from the observed R_t.

(a) (i) Identify by name this type of model structure and the components of the model.

(ii) If $\log(\sigma_t)$ is known we may rewrite the model as

$$\widetilde{R}_t = S_t + \xi_t, \quad t = 0, \ldots, N-1,$$

with $\widetilde{R}_t = R_t - \log(\sigma_t)$ and $S_t = \log|s_t|$.

Using a linear model representation with an $N-$dimensional parameter vector and an identity design matrix, find the least squares estimate of $S_t,\ t = 0, \ldots, N-1$.

(b) Now suppose we represent S_t in terms of a set of N orthogonal basis functions $\psi_k(t),\ k = 0, \ldots, N-1$, writing

$$S_t = \sum_{k=0}^{N-1} d_k \psi_k(t).$$

We can write

$$\widetilde{\boldsymbol{W}}_R = \boldsymbol{W}_S + \boldsymbol{W}_\xi,$$

where

$$\widetilde{\boldsymbol{W}}_R = \mathcal{W}\widetilde{\boldsymbol{R}} \qquad \boldsymbol{W}_S = \mathcal{W}\boldsymbol{S} \qquad \boldsymbol{W}_\xi = \mathcal{W}\boldsymbol{\xi},$$

$\mathcal{W}$ is an $N \times N$ matrix with kth row

$$[\psi_k(0), \psi_k(1), \ldots, \psi_k(N-1)]$$

and $\boldsymbol{S} = [S_0, \ldots, S_{N-1}]^T$, $\boldsymbol{\xi} = [\xi_0, \ldots, \xi_{N-1}]^T$ and $\widetilde{\boldsymbol{R}} = [\widetilde{R}_0, \ldots, \widetilde{R}_{N-1}]^T$. Denote the kth element of vector $\widetilde{\boldsymbol{W}}_R$ by $\widetilde{W}_{R,k}$.

(i) Find the expectation and variance of vector $\boldsymbol{W}_\xi$.

(ii) Find the expectation and variance of estimator $\widehat{W}_{S,k}$ defined by $\widehat{W}_{S,k} = \widetilde{W}_{R,k}$.

(iii) An alternative estimator is proposed by

$$\widehat{W}^{(t)}_{S,k} = \begin{cases} \widetilde{W}_{R,k} & \text{if } \ |\widetilde{W}_{R,k}| > \lambda \\ 0 & \text{if } \ |\widetilde{W}_{R,k}| < \lambda \end{cases}.$$

If $\lambda = \lambda(N) = \sqrt{2\log(N)}$, show that

$$P\left(\max|\widehat{W}^{(t)}_{\xi,k}| > \lambda(N)\right) \to 0.$$

Discuss the properties of $\widehat{W}^{(t)}_{S,k}$ for this choice of λ.

You may assume $1 - (2\Phi(\lambda(N)) - 1)^N \to 0$, *as* $N \to \infty$.

5. For $i = 1, \ldots, m$, let the loss of obligor i, denoted by L'_i, given a random rate variable Λ_i, be modelled as a Poisson random variable; that is given $\Lambda_i = \lambda_i,\ i = 1,\ \ldots, m$,

$$L'_i|_{\Lambda_i = \lambda_i} \sim \text{Poisson}\left(\lambda_i\right)$$

independently. We are interested in properties of the total loss

$$L_{\text{tot}} = \sum_{i=1}^{m} L'_i,$$

independently over i.

(a) Write down the form of the distribution of L_{tot}, conditional on the $\{\Lambda_i\}$ taking a set of fixed values $\{\lambda_i\}$.

(b) What is the probability of default for obligor i, i.e. $P\left(L'_i > 0\right)$, given $\Lambda_i = \lambda_i << 1$?

(c) Calculate the mean and variance of vector $\boldsymbol{L}' = \left[L'_1, \ldots, L'_m\right]^T$, in terms of the mean and variance of $\boldsymbol{\Lambda} = \left[\Lambda_1, \ldots, \Lambda_m\right]^T$.

(d) Denote the joint distribution of $\boldsymbol{\Lambda}$ by $p_{\boldsymbol{\Lambda}}(\boldsymbol{\lambda})$, and write down the form of the marginal distribution of $\boldsymbol{L}'$, in terms of this density. Are the components of $\boldsymbol{L}'$ independent?

(e) In the special case of $\Lambda_1 = \Lambda_2 = \cdots = \Lambda_m = \Lambda$, where $\Lambda \sim$ *Exponential*(β) find the distribution of the vector $\boldsymbol{L}'$.

(f) Find the distribution of L_{tot}, leaving the expression in a summation form.