Spring 1997 John Rust
Economics 551b 37 Hillhouse, Rm. 27

MIDTERM EXAM

(Due: March 31, 1997)

QUESTION 1 The zero-beta version of the CAPM (capital asset pricing model) implies the following equation for security returns:

$$R_{it} = \gamma + \beta_i(R_{mt} - \gamma) + \epsilon_{it}, \quad i = 1, \dots, K, \quad t = 1, \dots, T, \tag{1}$$

where R_{it} is the realized return on security i at time t, R_{mt} is the return on the market portfolio at time t, γ is the expected return on a zero-beta portfolio, and ϵ_{it} is an error term reflecting idiosyncratic risks satisfying the condition $E\{\epsilon_{it}|R_{mt}\} = 0$. Assume that security returns are serially independent, but contemporaneously correlated with $K \times K$ covariance matrix Σ, i.e.

$$\Sigma = [\sigma_{ij}], \tag{2}$$

where $\sigma_{ij} = E\{\epsilon_{it}\epsilon_{jt}\}$ (note that serial independence implies that $E\{\epsilon_{it}\epsilon_{js}\} = 0$ if $t \neq s$). Suppose we estimate the parameters of the K regression equations

$$R_{it} = \alpha_i + \beta_i R_{mt} + \epsilon_{it}, \quad i = 1, \dots, K, \quad t = 1, \dots, T \tag{3}$$

so that the full vector of parameters is (Σ, α, β) where Σ is the $K \times K$ contemporaneous covariance matrix, α is the $K \times 1$ vector of intercepts, and β is the $K \times 1$ vector of slope coefficients.

1. What testable restrictions does the Zero Beta CAPM place on the α and β parameters?

2. Derive an efficient unbiased estimator for (α, β).

3. Describe how to estimate Σ.

4. Using daily stock return data on the 63 securities available via anonymous ftp to `gemini.econ.yale.edu` and cd to the subdirectory

   `pub/John_Rust/courses/econ161/stats/stocks/dat`

   and say `binary` and `get stockdat.dat` and `get stockdat.dht` to transfer the Gauss data file `stockdat.dat` containing the stock return data (a 754×65 matrix for the 63 securities and two market return indices, `EQLAV` and `VALAV`) and its associated header file `stockdat.dht` (containing the names of the securities in each column of `stockdat.dat`). Load the data into gauss using the commands `open f1=stockdat` and `dat=readr(f1,rowsf(f1))` and put the names of the securities (in 1 to 1 correspondence with the columns of `dat`) in the vector `stocknms` using the Gauss command `stocknms=getname("stockdat").` (**Note:** if you are using Gauss on a Unix system, you must first convert the `stockdat.dat` and `stockdat.dht` files to Unix format. Do this by issuing the command `transdat stockdat.dat`). Now compute estimates of (α, β) using the methods you described in parts 2 and 3 above.

5. Suppose you are a Bayesian with independent prior beliefs about the 126×1 parameter vector $(\alpha', \beta')'$ and the 63×63 covariance matrix Σ. Suppose your prior beliefs about $(\alpha', \beta')'$ are given by $N(e, I)$ where I is a 126×126 identity matrix and e is a 126×1 vector with the first 63 components equal to 0's (corresponding to an prior expectation of 0 for all the α's) and the remaining 63 components equal to 1's (corresponding to a prior expectation that all β's are 1). Let the prior for $[\Sigma]^{-1}$ be given by $W(\rho, \Omega)$, a Wishart distribution with degrees of freedom $\rho = 70$ and parameter matrix $\Omega = I$, where I is the 63×63 identity matrix. Use Gibbs sampling techniques to compute the posterior means and standard deviations of (α, β) using these prior beliefs.

QUESTION 2 Let $\{\tilde{X}_1, \ldots, \tilde{X}_N\}$ be *IID* draws from a multinomial distribution with density

$$f(x_i|\theta) = \prod_{k=1}^{K} \theta_k^{I\{x_i=k\}}, \tag{4}$$

where $\theta \in \Delta^{K-1}$, the $K-1$-dimensional simplex (i.e. the set of $\theta \in R^K$ satisfying $\theta_k \geq 0$ and $\sum_{k=1}^{K} \theta_k = 1$).

1. Derive the maximum likelihood estimator for θ.

2. Is the maximum likelihood estimator biased or unbiased?

3. Derive the covariance matrix for θ.

4. Derive the information matrix.

5. Is the maximum likelihood estimator efficient? If not, find a more efficient estimator for θ.

6. Suppose you were a Bayesian with a Dirichlet prior distribution for θ with hyperparameters $\alpha \in R_+^K$ (i.e. we restrict $\alpha_k > 0$ for $k = 1, \ldots, K$). If $K = 4$ and $\theta^* = (.1, .2, .4, .3)$ and $\alpha = (3, 1, 4, 4)$, compute the prior probability that θ lies in the set Γ, where Γ is a ball of radius ..01 about θ^*. Then generate 1000 random draws $\{\tilde{X}_1, \ldots, \tilde{X}_{1000}\}$ from the multinomial distribution with parameter θ^* and compute the posterior probability that θ lies in Γ, using A/R simulation methods if necessary. Finally compare this posterior probability with the probability computed from the normal approximation to the posterior density discussed in class.

7. Suppose we re-parameterized the model by the $K \times 1$ vector of parameters λ writing $\theta = p(\lambda)$, where θ_k is defined by the multinomial logit formula:

$$\theta_k = p_k(\lambda) = \frac{\exp\{\lambda_k\}}{\sum_{k'=1}^{K} \exp\{\lambda_{k'}\}}. \tag{5}$$

   What is the relationship between the MLE for θ and the MLE for λ? Is the MLE for λ unbiased? Compute the information matrix for the MLE of λ. Is the MLE efficient?

Hint 1: In parts 2 and 3 you might find the following matrix result useful: let the $(K-1) \times (K-1)$ matrix Σ be given by:

$$\Sigma = \begin{bmatrix} p_1(1-p_1) & -p_1p_2 & \cdots & -p_1p_{K-1} \\ -p_1p_2 & p_2(1-p_2) & \cdots & -p_2p_{K-1} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ -p_1p_{K-1} & p_2p_{K-1} & \cdots & p_{K-1}(1-p_{K-1}) \end{bmatrix},$$

where the $\{p_1, \ldots, p_K\}$ are positive numbers satisfying

$$\sum_{k=1}^{K} p_k = 1.$$

Then verify the Σ is invertible with inverse given by:

$$[\Sigma]^{-1} = \begin{bmatrix} \frac{1}{p_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{p_2} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \frac{1}{p_{K-2}} & 0 \\ 0 & \cdots & \cdots & \frac{1}{p_{K-1}} \end{bmatrix} + \frac{1}{(1-p_1-p_2-\cdots-p_{K-1})} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ 1 & \cdots & \cdots & 1 \\ 1 & \cdots & \cdots & 1 \\ 1 & 1 & \cdots & 1 \end{bmatrix}.$$

Hint 2: In part 6 you are required to compute the posterior probability of being within a ball of radius .01 about θ^*. I suggest computing this probability by monte carlo simulations, drawing say 10,000 draws from the Dirichlet posterior and computing the fraction that lie in the ball of radius .01 about θ^*. How do you draw from a Dirichlet distribution? Use the following result:

Lemma: *Let $(\tilde{G}_1, \ldots, \tilde{G}_K)$ be independent random variables, where each $\tilde{G}_i$ has a gamma distribution with parameter (α_i, β). Then the random vector $\tilde{\boldsymbol{D}} = (\tilde{D}_1, \ldots, \tilde{D}_K)$ has a Dirichlet distribution with parameter $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$ where each $\tilde{D}_i$ is defined by:*

$$\tilde{D}_i = \frac{\tilde{G}_i}{(\tilde{G}_1 + \cdots + \tilde{G}_K)}$$

How do you draw from a Gamma distribution with parameters (α, β)? Use the fact that when α is an integer, the Gamma distribution with parameters (α, β) has the representation as the sum of α *IID* exponentials with parameter β, i.e.

$$\tilde{G}_i = \tilde{E}_1 + \cdots \tilde{E}_\alpha$$

where $(\tilde{E}_1, \ldots, \tilde{E}_\alpha)$ are *IID* exponentials with density $f(x) = \beta \exp\{-\beta x\}$, $x \geq 0$, $\beta > 0$. Note that given the prior hyperparameters α in this problem, both the prior and posterior α hyperparameters will always be integers, making it easy for you to simulate from a Gamma an therefore from a Dirichlet.

QUESTION 3: Prove that if $\sup_{\theta\in\Theta} |H_N(\theta) - H(\theta)| \to 0$ with probability 1, H continuous, then if $\hat{\theta} \to \theta^*$ with probability 1, then $H_N(\hat{\theta}) \to H(\theta^*)$ with probability 1.

QUESTION 4. Access simulated data from a trinomial choice model with latent utilities v_j given by

$$v_j = x\theta_j + \epsilon_j, \quad j = 1, 2, 3 \tag{6}$$

available via anonymous ftp at `gemini.econ.yale.edu`. Use the `cd` command to go to `pub/John_Rust` and then `cd` again to the subdirectory `course/econ551/dat` and use the `get` command to retrieve the files `data1.asc` and `data2.asc`. These files contain ASCII data sets consisting of three columns, where the first column is an indicator of the chosen alternative and the remaining columns are the x covariates.

A. Estimate the unknown 6×1 vector θ by logit maximum likelihood using the first 1500 observations of both data sets. Note that $\theta = (\theta_2, \theta_3)$ consists of two 3×1 vectors affecting choice of alternatives 2 and 3 in equation (6) and θ_1 is normalized to a 3×1 vector of zeros. Note also that in addition to the two columns in x there is a third component representing the constant term. Thus, θ_{21} is the constant term for alternative 2, θ_{32} is the first slope coefficient for alternative 3, etc. Use the estimated model to compare predicted versus actual outcomes in rows 1501 to 2000 of the data set. Can you think of some sort of a "pseudo R^2" measure or any simple way of summarizing how well the estimated choice model predicts actual outcomes? (**Hint:** Some *Gauss* code for estimating logit models via maximum likelihood is available on the `econ551` Web page.)

B. Retrieve `data2.asc`. This contains data for a binomial choice model. Compute the probit maximum likelihood estimates of θ (now a 3×1 vector – the intercept and two slop terms for alternative 2 since the coefficients for alternative 1 have been normalized to zero) and the identified terms of the 2×2 matrix of Σ, the covariance matrix of (ϵ_1, ϵ_2). Compute binary logit estimates using the same data set. Which model, the probit or logit, does a better job of predicting the observations in rows 1501 to 2000 of this data set? Which of the two models, logit or probit, do you think was used as the true data generating process for this problem?

C. **(OPTIONAL EXTRA CREDIT QUESTION)** Retreive `data3.asc`. This contains data for a trinomial probit model. Unlike the binomial probit model, numerical integration is required to compute the likelihood function. However a Bayesian approach using Gibbs sampling is feasible using the data augmentation methods outlined in the Rossi and McCulloch *Journal of Econometrics* paper. Specify a Wishart prior for the inverse covariance matrix of the probit error terms and a normal prior for the θ coefficients in the utility function and compute posterior means of these quantities by Gibbs sampling. Students who do this question will received substantial extra credit, but **nobody will be penalized for not doing this part.**

Hints: Load the data sets `data1.asc` and `data2.asc` by anonymous ftp following the same instructions as above, but `cd /pub/John_Rust/courses/econ551/dat` and `get data1.asc` and `get data2.asc` (there is no need to switch to binary mode since these two data sets are in ASCII format). You will use `data1.asc` to estimate a trinomial logit model with 6 unknown alternative-specific coefficients by maximum likelihood.

To compute the maximum likelihood estimates you first need to write down the likelihood function and derivatives. Only first derivatives are required if you use BHHH but second derivatives are required if you use Newton's method. There are two ways to write down these formulae: 1) write down formulae that are specific to the case at hand, 2) write down formulae that will be useful for estimating any type of multinomial logit model with any number of alternatives.

1. **Alternative-specific trinomial code:** You will need to write a procedure `EVALMNL1.G` that codes the likelihood and first and second derivatives of the trinomial logit model with alternative-specific coefficients. This procedure will then be passed in the setup program that reads in the data and runs the main maximization procedure, `MAX.G`. Partition the unknown 6×1 parameter vector θ into two components, $\theta = (q_2, q_3)$ where q_2 is the 3×1 parameter vector for alternative 2 and q_3 is the 3×1 parameter vector for alternative 3 with q_1 normalized to 0 since it is unidentified. Next write down the log-likelihood function for this model as follows:

$$\begin{aligned}\log\{P(1|x)\} &= -\log\{1+\exp(xq_2)+\exp(xq_3)\} \\ \log\{P(2|x)\} &= xq_2 - \log\{1+\exp(xq_2)+\exp(xq_3)\} \\ \log\{P(3|x)\} &= xq_3 - \log\{1+\exp(xq_2)+\exp(xq_3)\}.\end{aligned}$$

2. **General multinomial logit code:** you can write a general multinomial logit procedure `EVALMNL.G` that codes the likelihood and first and second derivatives for a general multinomial logit model with fixed coefficients and alternative-specific covariates. This procedure is then passed in the setup program that reads in the data and runs the main maximization procedure, `MAX.G`. This code will estimate an arbitrary MNL model with an arbitrary number of alternatives. The general model has a covariate vector $x = (x_1, \ldots, x_J)$ which is alternative-specific (with x_j being the covariate vector for alternative j) and the parameter vector θ is not alternative-specific. The likelihood function for the general MNL model has the following form:

$$\log\{P(d|x)\} = x_d\theta - \log\left\{\sum_{j=1}^{J}\exp\{x_j\theta\}\right\}.$$

You can "trick" the general MNL code into estimating the alternative-specific trinomial model of this assignment by defining an 18×1 covariate vector $x = (x_1, x_2, x_3)$ as follows:

$$\begin{aligned}x_1 &= (0_{1\times3}, 0_{1\times3}) \\ x_2 &= (x_{1\times3}, 0_{1\times3}) \\ x_3 &= (0_{1\times3}, x_{1\times3})\end{aligned}$$

where $x_{1\times3}$ is the original non-alternative-specific covariate vector (dependence on observation i, $i = 1, \ldots, 2000$ is supressed for clarity). It is straightforward to verify that when the alternative-specific covariate vector (x_1, x_2, x_3) defined above is substituted into the general formula for the MNL log-likelihood we obtain the special case for the trinomial logit model with alternative-specific coefficients. We now complete the assignment by specifying the first and second derivatives of the log-likelihood function. We will do this for both the alternative-specific and general forms of the MNL model.

First derivatives for trinomial logit model with alternative-specific θ's

$$\begin{aligned}
\partial \log \{P(1|x)\} / \partial \theta &= (-xP(2|x), -xP(3|x)) \\
\partial \log \{P(2|x)\} / \partial \theta &= (x[1-P(2|x)], -xP(3|x)) \\
\partial \log \{P(3|x)\} / \partial \theta &= (-xP(2|x), x[1-P(3|x)])
\end{aligned}$$

First derivatives for general MNL model

$$\partial \log \{P(d|x)\} / \partial \theta = \sum_{j=1}^{J} \left[I\{j=d\} - P(j|x) \right] x_j$$

Second derivatives for trinomial logit model with alternative-specific θ's

$$\begin{aligned}
\partial^2 \log \{P(d|x)\} / \partial q_2 \partial q_2' &= -x'xP(2|x)[1-P(2|x)] \\
\partial^2 \log \{P(d|x)\} / \partial q_2 \partial q_2' &= -x'xP(3|x)[1-P(3|x)] \\
\partial^2 \log \{P(d|x)\} / \partial q_2 \partial q_3' &= x'xP(2|x)P(3|x)
\end{aligned}$$

Second derivatives for general MNL model

$$\partial^2 \log \{P(d|x)\} / \partial \theta \partial \theta' = -\sum_{j=1}^{J} x_j x_j' P(j|x) + \left[\sum_{j=1}^{J} x_j P(j|x) \right] \left[\sum_{j=1}^{J} x_j P(j|x) \right]'$$

QUESTION 5 Consider the linear regression model

$$y_t = X_t \beta^* + \epsilon_t, \quad t = 1, \ldots, T \tag{7}$$

where $\mathrm{med}(\epsilon_t | X_t) = 0$. Prove that the LAD estimator defined by

$$\hat{\beta} = \underset{\beta}{argmin} \frac{1}{T} \sum_{t=1}^{T} |y_t - X_t \beta| \tag{8}$$

converges with probability 1 to β^*.