Spring 2001 | John Rust
Economics 551b | 37 Hillhouse, Rm. 27

FINAL EXAM
April 27, 2001

INSTRUCTIONS: Do all Parts I, II and III below. You are required to answer **all** questions in Part I, 2 out of 6 questions in Part II, and 1 out of 4 questions from Part III. Total points for the final exam is 100. Part I should take about 15 minutes and is worth 15 points. Part II should take about 30 minutes and is worth 30 points. Part III should take about 60 minutes and is worth 55 points. You have 3 hours for the exam, but my expectation that almost all students will complete it in two hours.

Part I: 15 minutes, 15 points. Answer all questions below:

1. Suppose $\{\tilde{X}_1,\ldots,\tilde{X}_N\}$ are *IID* draws from a $N(\mu,\sigma^2)$ distribution (i.e. a normal distribution with mean μ and variance σ^2). Consider the estimator $\hat{\theta}_N$ defined by:

$$\hat{\theta}_N = \left(\frac{1}{N}\sum_{i=1}^{N}\tilde{X}_i\right)^2 \tag{1}$$

Which of the following statements are true and which are false?

A. $\hat{\theta}_N$ is a consistent estimator of σ^2.

B. $\hat{\theta}_N$ is an unbiased estimator of σ^2.

C. $\hat{\theta}_N$ is a consistent estimator of μ.

D. $\hat{\theta}_N$ is an unbiased estimator of μ.

E. $\hat{\theta}_N$ is a consistent estimator or μ^2.

F. $\hat{\theta}_N$ is an unbiased estimator of μ^2.

G. $\hat{\theta}_N$ is an upward biased estimator of μ^2.

H. $\hat{\theta}_N$ is a downward biased estimator of μ^2.

2. Consider estimation of the linear model

$$y = X\beta + \epsilon \tag{2}$$

based on N *IID* observations $\{y_i, X_i\}$ where X_i is a $K\times 1$ vector of independent variables and y_i is a 1×1 scalar independent variable. Mark each of the following statements as true or false:

A. The Gauss-Markov Theorem proves that the ordinary least squares estimator (OLS) it BLUE (Best Linear Unbiased Estimator).

B. The Gauss-Markov Theorem requires that the error term in the regression ϵ be normally distributed with mean 0 and variance σ^2.

C. The Gauss-Markov Theorem does not apply if the true regression function does not equal $X\beta$, i.e. if $E\{y|X\} \neq X\beta$.

D. The Gauss-Markov Theorem does not apply if there is heteroscedasticity.

E. The Gauss-Markov Theorem does not apply if the error term has a non-normal distribution.

F. The maximum likelihood estimator of β is more efficient than the OLS estimator of β.

G. The OLS estimator of β will be unbiased only if the error terms are distributed independently of X and have mean 0.

H. The maximum likelihood estimator of β is the same as OLS only in the case where ϵ is normally distributed.

I. The OLS estimator will be a consistent estimator of β even if the error term ϵ is not normal and even if there is heteroscedasticity.

J. The OLS estimator of the asymptotic covariance matrix for β, $\hat{\sigma}^2(X'X/N)^{-1}$ (where $\hat{\sigma}^2$ is the sample variance of the estimated residuals $\hat{\epsilon}_i = y_i - X_i\hat{\beta}$) is a consistent estimator regardless of whether ϵ is normally distributed or not.

K. The OLS estimator of the asymptotic covariance matrix for β, $\hat{\sigma}^2(X'X/N)^{-1}$ (where $\hat{\sigma}^2$ is the sample variance of the estimated residuals $\hat{\epsilon}_i = y_i - X_i\hat{\beta}$) is a consistent estimator regardless of whether there is heteroscedasticity in ϵ.

L. If the distribution of ϵ is double exponential, i.e. if $f(\epsilon) = \exp\{-|\epsilon|/\sigma\}/(2\sigma)$, the maximum likelihood estimator of β is the Least Absolute Deviations estimator and it is asymptotically efficient relative to the OLS estimator.

M. The OLS estimator cannot be used if the regression function is misspecified, i.e. if the true regression function $E\{y|X\} \neq X\beta$.

N. The OLS estimator will be inconsistent if ϵ and X are correlated.

O. The OLS estimator will be inconsistent if the dependent variable y is truncated, i.e. if the dependent variable is actually determined by the relation

$$y = \max[0, X\beta + \epsilon] \tag{3}$$

P. The OLS estimator is inconsistent if ϵ has a Cauchy distribution, i.e. if the density of ϵ is given by

$$f(\epsilon) = \frac{1}{\pi(1+\epsilon^2)} \tag{4}$$

Q. The 2-stage least squares estimator is a better estimator than the OLS estimator because it has two stages and is therefore twice as efficient.

R. If the set of instrumental variables W and the set of regressors X in the linear model coincide, then 2 stage least squares estimator of β is the same as the OLS estimator of β.

Part II: 30 minutes, 30 points. Answer 2 of the following 6 questions below.

QUESTION 1 (Probability question) Suppose $\tilde{Z}$ is a $K \times 1$ random vector with a multivariate $N(0, I)$ distribution, i.e. $E\{\tilde{Z}\} = 0$ where 0 is a $K \times 1$ vector of zeros and $E\{\tilde{Z}\tilde{Z}'\} = I$ where I is the $K \times K$ identity matrix. Let M be a $K \times K$ idempotent matrix, i.e. a matrix that satisfies

$$M^2 = M * M = M \tag{5}$$

Show that

$$\tilde{Z}' M \tilde{Z} \sim \chi^2(J) \tag{6}$$

where $\chi^2(J)$ denotes a chi-squared random variable with J degrees of freedom and $J = \text{rank}(M)$.
Hint: Use the fact that M has a singular value decomposition, i.e.

$$M = XDX' \tag{7}$$

where $X'X = I$ and D is a diagonal matrix whose diagonal elements are equal to either 1 or 0.

QUESTION 2 (Markov Processes)

A. (10%) Are Markov processes of any use in econometrics? Describe some examples of how Markov processes are used in econometrics such as providing models of serially dependent data, as a framework for establishing convergence of estimators and proving laws of large numbers, central limit theorems, etc. and as computational tool for doing simulations.

B. (10%) What is a random walk? Is a random walk always a Markov process? If not, provide a counter-example.

C. (40%) What is the ergodic or invariant distribution of a Markov process? Do all Markov processes have invariant distributions? If not, provide a counterexample of a Markov process that doesn't have an invariant distribution. Can a Markov process have more than 1 invariant distribution? If so, give an example.

D. (40%) Consider the discrete Markov process $\{X_t\} = \{1, 2, 3\}$ with transition probability

$$\begin{aligned} P\{X_{t+1} = 1 | X_t = 1\} &= \frac{1}{2} \quad P\{X_{t+1} = 2 | X_t = 1\} = \frac{1}{3} \quad P\{X_{t+1} = 3 | X_t = 1\} = \frac{1}{6} \\ P\{X_{t+1} = 1 | X_t = 2\} &= \frac{3}{4} \quad P\{X_{t+1} = 3 | X_t = 2\} = \frac{1}{4} \quad P\{X_{t+1} = 2 | X_t = 3\} = 1 \end{aligned}$$

Does this process have an invariant distribution? If so, find all of them.

QUESTION 3 (Consistency of M-estimator) Consider an M-estimator defined by:

$$\widehat{\theta}_N = \arg\max_{\theta \in \Theta} Q_N(\theta).$$

Suppose following two conditions are given
(i) (Identification) For all $\varepsilon > 0$

$$Q(\theta^*) > \sup_{\theta \notin B(\theta^*, \varepsilon)} Q(\theta)$$

where $B(\theta^*, \varepsilon) = \{\theta \in R^k | \|\theta - \theta^*\| < \epsilon\}$.
(ii) (Uniform Convergence)

$$\sup_{\theta \in \Theta} |Q_N(\theta) - Q(\theta)| \xrightarrow{p} 0.$$

Prove consistency of the estimator by showing

$$P\left(\widehat{\theta}_N \notin B(\theta^*, \varepsilon)\right) \xrightarrow{p} 0.$$

QUESTION 4 (Time series question) Suppose $\{X_t\}$ is an ARMA(p,q) process, i.e.

$$A(L)X_t = B(L)\epsilon_t$$

where $A(L)$ is a q^{th} order lag-polynomial

$$A(L) = \alpha_0 + \alpha_1 L + \alpha_2 L^2 + \cdots + \alpha_q L^q$$

and $B(L)$ is a p^{th} order lag-polynomial

$$B(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + \cdots + \beta_p L^p$$

and the lag-operator L^k is defined by

$$L^k X_t = X_{t-k}$$

and $\{\epsilon_t\}$ is a white-noise process, $E\{\epsilon_t\} = 0$ and ($\text{cov}(\epsilon_t, \epsilon_s)$= 0 if $t \neq s$, $= \sigma^2$ if $t = s$).

A. (30%) Write down the autocovariance and spectral density functions for this process.

B. (30%) Show that if $p = 0$ an autoregression of X_t on q lags of itself provides a consistent estimate of $(\alpha_0/\sigma, \ldots, \alpha_q/\sigma)$. Is the autoregression still consistent if $p > 0$?

C. (40%) Assume that a central limit theorem holds, i.e. the distribution of normalized sums of $\{X_t\}$ to converge in distribution to a normal random variable. Write down an expression for the variance of the limiting normal distribution.

QUESTION 5 (Empirical question) Assume that shoppers always choose only a single brand of canned tuna fish from the available selection of K alternative brands of tuna fish each time they go shopping at a supermarket. Assume initially that the (true) probability that the decision-maker chooses brand k is the same for everybody and is given by θ_k^*, $k = 1, \ldots, K$. Marketing researchers would like to learn more about these choice probabilities, $\theta^* = (\theta_1^*, \ldots, \theta_K^*)$ and spend a great deal of money sampling shoppers' actual tuna fish choices in order to try to estimate these probabilities. Suppose the Chicken of the Sea Tuna company undertook a survey of N shoppers and for each shopper shopping at a particular supermarket with a fixed set of K brands of tuna fish, recorded the brand b_i chosen by shopper i, $i = 1, \ldots, N$. Thus, $b_1 = 2$ denotes the observation that consumer 1 chose tuna brand 2, and $b_4 = K$ denotes the observation that consumer 4 chose tuna brand K, etc.

A. (10%) Without doing any estimation, are there any general restrictions that you can place on the $K \times 1$ parameter vector θ^*?

B. (10%) Is it reasonable to suppose that θ_k^* is the same for everyone? Describe several factors that could lead different people to have different probabilities of purchasing different brands of tuna. If you were a consultant to Chicken of the Sea, what additional data would you recommend that they collect in order to better predict the probabilities that consumers buy various brands of tuna? Describe how you would use this data once it was collected.

C. (20%) Using the observations $\{b_1, \ldots, b_N\}$ on the observed brand choices of the sample of N shoppers, write down an estimator for θ^* (under the assumption that the "true" brand choice probabilities θ^* are the same for everyone). Is your estimator unbiased?

D. (20%) What is the maximum likelihood estimator of θ^*? Is the maximum likelihood estimator unbiased?

E. (40%) Suppose Chicken of the Sea Tuna company also collected data on the prices $\{p_1, \ldots, p_K\}$ that the supermarket charged for each of the K different brands of tuna fish. Suppose someone proposed that the probability of buying brand j was a function of the prices of all the various brands of tuna, $\theta_j^*(p_1, \ldots, p_K)$, given by:

$$\theta_j^*(p_1, \ldots, p_K) = \frac{\exp\{\beta_j + \alpha p_j\}}{\sum_{k=1}^K \exp\{\beta_k + \alpha p_k\}}$$

Describe in general terms how to estimate the parameters $(\alpha, \beta_1, \ldots, \beta_K)$. If $\alpha > 0$, does an increase in p_j decrease or increase the probability that a shopper would buy brand j?

QUESTION 6 (Regression question) Let (y_t, x_t) be *IID* observations from a regression model

$$y_t = \beta x_t + \epsilon_t$$

where y_t, x_t, and ϵ_t are all scalars. Suppose that ϵ_t is normally distributed with $E\{\epsilon_t|x_t\} = 0$, but $\text{var}(\epsilon_t|x_t) = \sigma^2 |x_t|^\theta$. Consider the following two estimators for β^*:

$$\hat{\beta}_T^1 = \frac{\sum_{t=1}^T y_t}{\sum_{t=1}^T x_t}$$

$$\hat{\beta}_T^2 = \frac{\sum_{t=1}^T x_t y_t}{\sum_{t=1}^T x_t^2}$$

A. (20%) Are these two estimators consistent estimators of β^*? Which estimator is more efficient when: 1) if we know *a priori* that $\theta^* = 0$, and 2) we don't know θ^*? Explain your reasoning for full credit.

B. (20%) Write down an asymptotically optimal estimator for β^* if we know the value of θ^* *a priori.*

C. (20%) Write down an asymptotically optimal estimator for (β^*, θ^*) if we don't know the value of θ^* *a priori.*

D. (20%) Describe the feasible GLS estimator for (β^*, θ^*). Is the feasible GLS estimator asymptotically efficient?

E. (20%) How would your answers to parts A to D change if you didn't know the distribution of ϵ_t was normal?

Part III (60 minutes, 55 points). Do 1 out of the 4 questions below.

QUESTION 1 (Hypothesis testing) Consider the GMM estimator with *IID* data, i.e the observations $\{y_i, x_i\}$ are independent and identically distributed using the moment condition $H(\theta) = E\{h(\tilde{y}, \tilde{x}, \theta)\}$, where h is a $J \times 1$ vector of moment conditions and θ is a $K \times 1$ vector of parameters to be estimated. Assume that the moment conditions are correctly specified, i.e. assume there is a unique θ^* such that $H(\theta^*) = 0$. Show that in the overidentified case $(J > K)$ that the minimized value of the GMM criterion function is asymptotically χ^2 with $J - K$ degrees of freedom:

$$N H_N(\hat{\theta}_N)[\hat{\Omega}_N]^{-1} H_N(\hat{\theta}_N) \underset{d}{\Longrightarrow} \chi^2(J-K), \tag{8}$$

where H_N is a $J \times 1$ vector of moment conditions, θ is a $K \times 1$ vector of parameters, $\chi^2(J-K)$ is a Chi-squared random variable with $J - K$ degrees of freedom,

$$\hat{\theta}_N = \underset{\theta \in \Theta}{argmin}\, H_N(\theta)[\hat{\Omega}_N]^{-1} H_N(\theta),$$

$$H_N(\theta) = \frac{1}{N} \sum_{i=1}^{N} h(y_i, x_i, \theta),$$

and $\hat{\Omega}_N$ is a consistent estimator of Ω given by

$$\Omega = E\{h(\tilde{y}, \tilde{x}, \theta^*) h(\tilde{y}, \tilde{x}, \theta^*)'\}.$$

Hint: Use Taylor series expansions to provide a formula for $\sqrt{N}(\hat{\theta}_N - \theta^*)$ from the first order condition for $\hat{\theta}_N$

$$\nabla H_N(\hat{\theta}_N) \hat{\Omega}_N^{-1} H_N(\hat{\theta}_N) = 0 \tag{9}$$

and a Taylor series expansion of $H_N(\hat{\theta}_N)$ about θ^*

$$H_N(\hat{\theta}_N) = H_N(\theta^*) + \nabla H_N(\tilde{\theta}_N)(\hat{\theta}_N - \theta^*) \tag{10}$$

where

$$\nabla H_N(\theta) \equiv \frac{1}{N} \sum_{i=1}^{N} \frac{\partial h}{\partial \theta}(y_i, x_i, \theta) \tag{11}$$

is the $(J \times K)$ matrix of partial derivatives of the moment conditions $H_N(\theta)$ with respect to θ and $\tilde{\theta}_N$ is a vector each of whose elements are on the line segment joining the corresponding components of $\hat{\theta}_N$ and θ^*. Use the above two equations to derive the following formula for $H_N(\hat{\theta}_N)$

$$H_N(\hat{\theta}_N) = M_N H_N(\theta^*) \tag{12}$$

where

$$M_N = \left[I - \nabla H_N(\hat{\theta}_N)[\nabla H_N(\hat{\theta}_N)' \hat{\Omega}_N^{-1} \nabla H_N(\tilde{\theta}_N)]^{-1} \nabla H_N(\hat{\theta}_N)' \hat{\Omega}_N^{-1} \right]. \tag{13}$$

Show that with probability 1 we have $M_N \to M$ where M is a $(J \times J)$ idempotent matrix. Then using this result, and using the Central Limit Theorem to show that

$$\sqrt{N} H_N(\theta^*) \underset{d}{\Longrightarrow} N(0, \Omega), \tag{14}$$

and using the probability result from Question 0 of Part II, show that the minimized value of the GMM criterion function does indeed converge in distribution to the $\chi^2(J-K)$ random variable as claimed in equation (8).

QUESTION 2 (Consistency of Bayesian posterior) Consider a Bayesian who has observes *IID* data $(X_1, \ldots, X_N)$, where $f(x|\theta)$ is the likelihood for a single observation, and $p(\theta)$ is the prior density over an unknown finite-dimensional parameter $\theta \in R^K$.

A. (10%) Use Bayes Rule to derive a formula for the posterior density of θ given $(X_1, \ldots, X_N)$.

B. (20%) Let $P(\theta \in A|X_1, \ldots, X_N\}$ be the posterior probability θ is in some set $A \subset \Theta$ given the first N observations. Show that this posterior probability satisfies the *Law of iterated expectations:*

$$E\{P(\theta \in A|X_1, \ldots, X_{N+1})|X_1, \ldots, X_N\} = P(\theta \in A|X_1, \ldots, X_N).$$

C. (20%) A *martingale* is a stochastic process $\{\tilde{Z}_t\}$ that satisfies $E\left\{\tilde{Z}_{t+1}|\mathcal{I}_t\right\} = \tilde{Z}_t$, where $\mathcal{I}_t$ denotes the information set at time t and includes knowledge of all past Z_t's up to time t, $\mathcal{I}_t \supset (\tilde{Z}_1, \ldots, \tilde{Z}_t)$. Use the result in part A to show that the process $\{\tilde{Z}_t\}$ where $\tilde{Z}_t = P(\theta \in A|\tilde{X}_1 \ldots, X_t)$ is a martingale. (We are interested in martingales because the *Martingale Convergence Theorem* can be used to show that if θ is finite-dimensional, then the posterior distribution converges with probability 1 to a point mass on the true value of θ generating the observations $\{X_i\}$. But you don't have to know anything about this to answer this question.)

D. (50%) Suppose that if θ is restricted to the K-dimensional simplex, $\theta = (\theta_1, \ldots, \theta_K)$ with $\theta_i \in (0,1)$, $i = 1, \ldots, K$, $1 = \sum_{i=1}^K \theta_i$, and the distribution of X_i given θ is multinomial with parameter θ, i.e.

$$Pr\{X_i = k\} = \theta_k, \quad k = 1, \ldots, K.$$

Suppose the prior distribution over θ, $p(\theta)$ is *Dirichlet* with parameter α:

$$p(\theta) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_K)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_K)} \theta_1^{\alpha_1 - 1} \cdots \theta_K^{\alpha_K - 1}$$

where both $\theta_i > 0$ and $\alpha_i > 0$, $i = 1, \ldots, K$. Compute the posterior distribution and show 1) the posterior is also Dirichlet (i.e. the Dirichlet is a conjugate family), and show directly that as $N \to \infty$ that the posterior distribution converges to a point mass on the true parameter θ generating the data.

QUESTION 3 Consider the *random utility model:*

$$\tilde{u}_d = v_d + \tilde{\epsilon}_d, \quad d = 1, \ldots, D \tag{15}$$

where $\tilde{u}_d$ is a decision-maker's payoff or utility for selecting alternative d from a set containing D possible alternatives (we assume that the individual only chooses one item). The term v_d is known as the deterministic or *strict utility* from alternative d and the error term $\tilde{\epsilon}_d$ is the random component of utility. In empirical applications v_d is often specified as

$$v_d = X_d \beta \tag{16}$$

where X_d is a vector of observed covariates and β is a vector of coefficients determining the agent's utility to be estimated. The interpretation is that X_d represents a vector of characteristics of the decision-maker and alternative d that are observable by the econometrician and ϵ_d represents characteristics of the agent and alternative d that affect the utility of choosing alternative d which are unobserved by the econometrician. Define the agent's *decision rule* $\delta(\epsilon_1, \ldots, \epsilon_D)$ by:

$$\delta(\epsilon) = argmax_{d=1,\ldots,D} \left[v_d + \tilde{\epsilon}_d \right] \tag{17}$$

i.e. $\delta(\epsilon)$ is the optimal choice for an agent whose unobserved utility components are $\epsilon = (\epsilon_1, \ldots, \epsilon_D)$. Then the agent's *choice probability* $P\{d|X\}$ is given by:

$$P\left\{d|X\right\} = \int I\{d = \delta(\epsilon)\} f(\epsilon|X) d\epsilon \tag{18}$$

where $X = (X_1, \ldots, X_D)$ is the vector of observed characteristics of the agent and the D alternatives and $f(\epsilon|X)$ is the conditional density function of the random components of utility given the values of observed components X, and $I\{\delta(\epsilon) = d\}$ is the *indicator function* given by $I\{\delta(\epsilon) = d\} = 1$ if $\delta(\epsilon) = d$ and 0 otherwise. Note that the integral above is actually a multivariate integral over the D components of $\epsilon = (\epsilon_1, \ldots, \epsilon_D)$, and simply represents the probability that the values of the vector of unobserved utilities ϵ lead the agent to choose alternative d.

Definition: The *Social Surplus Function* $U(v_1, \ldots, v_D, X)$ is given by:

$$U(v_1, \ldots, v_D, X) = E\left\{ \max_{d=1,\ldots,D} [v_d + \tilde{\epsilon}_d] \Big| X \right\} = \int_{\epsilon_1} \cdots \int_{\epsilon_D} \max_{d=1,\ldots,D} [v_d + \epsilon_d] f(\epsilon_1, \ldots, \epsilon_D | X) d\epsilon_1 \cdots d\epsilon_D \tag{19}$$

The Social Surplus function is the expected maximized utility of the agent.[1]

A. (50%) Prove the *Williams-Daly-Zachary Theorem:*

$$\frac{\partial U(v_1, \ldots, v_D, X)}{\partial v_d} = P\{d|X\} \tag{20}$$

and discuss its relationship to *Roy's Identity.*

Hint: Interchange the differentiation and expectation operations when computing $\partial U / \partial v_d$:

$$\begin{aligned} \frac{\partial U(v_1, \ldots, v_D, X)}{\partial v_d} &= \partial / \partial v_d \int_{\epsilon_1} \cdots \int_{\epsilon_D} \max_{d=1,\ldots,D} [v_d + \epsilon_d] f(\epsilon_1, \ldots, \epsilon_D | X) d\epsilon_1 \cdots d\epsilon_D \\ &= \int_{\epsilon_1} \cdots \int_{\epsilon_D} \partial / \partial v_d \max_{d=1,\ldots,D} [v_d + \epsilon_d] f(\epsilon_1, \ldots, \epsilon_D | X) d\epsilon_1 \cdots d\epsilon_D \end{aligned}$$

[1]If we think of an economy consisting of a population of agents each with their own observed vector of utilities ϵ and $f(\epsilon|X)$ is the density function representing the distribution of these "types" in the population, then $U(v_1, \ldots, v_D, X)$ represents the indirect or maximized utility of a typical person in the population. This is the reason U is referred to as a Social Surplus Function.

and show that

$$\partial/\partial v_d \max_{d=1,\ldots,D}[v_d + \epsilon_d] = I\{d = \delta(\epsilon)\}.$$

B. (50%) Consider the special case of the random utility model when $\epsilon = (\epsilon_1, \ldots, \epsilon_D)$ has a multivariate (Type I) *extreme value distribution:*

$$f(\epsilon|X) = \prod_{d=1}^{D} \exp\{-\epsilon_d\} \exp\{-\exp\{-\epsilon_d\}\}. \tag{21}$$

Show that the conditional choice probability $P\{d|X\}$ is given by the *multinomial logit formula:*

$$P\{d|X\} = \frac{\exp\{v_d\}}{\sum_{d'=1}^{D} \exp\{v_{d'}\}}. \tag{22}$$

Hint 1: Use the Williams-Daly-Zachary Theorem, showing that in the case of the extreme value distribution (21) the Social Surplus function is given by

$$U(v_1, \ldots, v_D, X) = \gamma + \log\left[\sum_{d=1}^{D} \exp\{v_d\}\right]. \tag{23}$$

where $\gamma = .577216\ldots$ is Euler's constant.

Hint 2: To derive equation (23) show that the extreme value family is *max-stable:* i.e. if $(\epsilon_1, \ldots, \epsilon_D)$ are *IID* extreme value random variables, then $\max_d\{\epsilon_d\}$ also has an extreme value distribution. Also use the fact that the expectation of a single extreme value random variable with location parameter α and scale parameter σ is given by:

$$E\{\tilde{\epsilon}\} = \int_{-\infty}^{+\infty} \epsilon \exp\{-\epsilon\} \exp\{-\exp\{-\epsilon\}\} d\epsilon = \alpha + \sigma\gamma, \tag{24}$$

and the CDF is given by

$$F(x|\alpha, \sigma) = P\{\tilde{\epsilon} \le x|\alpha, \sigma\} = \exp\left\{-\exp\left\{\frac{-(x-\alpha)}{\sigma}\right\}\right\}. \tag{25}$$

Hint 3: Let $(\epsilon_1, \ldots, \epsilon_D)$ be *INID* (independent, non-identically distributed) extreme value random variables with location parameters $(\alpha_1, \ldots, \alpha_D)$ and common scale parameter σ. Show that this family is max-stable by proving that $\max(\epsilon_1, \ldots, \epsilon_D)$ is an extreme value random variable with scale parameter σ and location parameter

$$\alpha = \sigma \log\left[\sum_{d=1}^{D} \exp\{\alpha_d/\sigma\}\right] \tag{26}$$

QUESTION 4 (Latent Variable Models) The *Binary Probit Model* can be viewed as a simple type of latent variable model. There is an underlying linear regression model

$$\tilde{y} = X\beta^* + \epsilon \tag{27}$$

but where the dependent variable $\tilde{y}$ is *latent,* i.e. it is not observed by the econometrician. Instead we observe the dependent variable y given by

$$y = \begin{cases} 1 & \text{if} \quad \tilde{y} > 0 \\ 0 & \text{if} \quad \tilde{y} \le 0 \end{cases} \tag{28}$$

1. (5%) Assume that the error term $\epsilon \sim N(0, \sigma^2)$. Show that the scale of β^* and the parameter σ^2 is not simultaneously identified and therefore without loss of generality we can normalize $\sigma^2 = 1$ and interpret the estimated β coefficients as being the true coefficients β^* divided by σ:
$$\beta = \frac{\beta^*}{\sigma}. \tag{29}$$

2. (10%) Derive the conditional probability $\Pr\{y = 1|X\}$ in terms of X, β and the standard normal CDF, Φ and use this probability to write down the likelihood function for N *IID* observations of pairs $\{(y_i, X_i)\}, i = 1, \ldots, N$.

3. (20%) Show that β can be consistently estimated by nonlinear least squares by writing down the least squares problem and sketching a proof for its consistency.

4. (20%) Derive the asymptotic distribution of the maximum likelihood estimator by providing an analytical formula for the asymptotic covariance matrix of the MLE estimator $\hat{\beta}_N$ (**Hint:** This is the inverse of the information matrix $\mathcal{I}$. Derive a formula for $\mathcal{I}$ in terms of Φ, X and β and possibly other terms.)

5. (20%) Derive the asymptotic distribution of the nonlinear least squares estimator and compare it to the maximum likelihood estimator. Is the nonlinear least squares estimator asymptotically inefficient?

6. (25%) Show that the nonlinear least squares estimator of β is subject to heteroscedasticity by deriving an explicit formula for the conditional variance of the error term in the nonlinear regression formulation of the estimation problem. Can you form a more efficient estimator by correcting for this heteroscedasticity in a two stage feasible GLS procedure (i.e. in stage 1 computing an initial consistent, but inefficient estimator of β by ordinary nonlinear least squares and in stage two using this initial consistent estimator to correct for the heteroscedasticity and using the stage two estimator of β as the feasible GLS estimator)? If so, is this feasible GLS procedure asymptotically efficient? If you believe so, provide a sketch of the derviation of the asymptotic distribution of the feasible GLS estimator. Otherwise provide a counterexample or a sketch of an argument why you believe the feasible GLS procedure is asymptotically inefficient relative to the maximum likelihood estimator.