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

PROBLEM SET 3

Classical Methods(I)

(Due: February 24, 1997)

QUESTION 1 Let $Y \sim N(0, \Sigma)$ where Y is $k \times 1$ random vector and Σ is $k \times k$ positive definite matrix. Prove $Y'\Sigma^{-1}Y \sim \chi^2(k)$.

QUESTION 2 Suppose regression model

$$y = X\beta^* + \varepsilon$$

(where y, ε are $N \times 1$, X is $N \times k$, and β^* is $k \times 1$) with normality assumption on disturbance, $\varepsilon \sim N(0, \Omega)$. Derive the maximum likelihood estimator (MLE) of β^* when Ω is known, and show that MLE is identical to the GLS estimator.

QUESTION 3 Show that the Cramer-Rao lower bound, I^{-1}, corresponds to the GLS covariance matrix, $(X'\Omega^{-1}X)^{-1}$, in Question 2.

QUESTION 4 Suppose following SUR model,

$$\begin{aligned} y_{i1} &= X_{i1}\beta_1^* + \varepsilon_{i1} \\ &\vdots \\ y_{ip} &= X_{ip}\beta_p^* + \varepsilon_{ip} \end{aligned}$$

where β_1^* is $k_1 \times 1$, ... , β_p^* is $k_p \times 1$, and $k = \sum_{i=1}^p k_i$. Also suppose covariance matrix of $\varepsilon_i = (\varepsilon_{i1}, \cdots, \varepsilon_{ip})'$ is Ω. Derive the covariance matix of $k \times 1$ OLS estimator $\widehat{\beta} = \left(\widehat{\beta}_1', \cdots, \widehat{\beta}_p'\right)'$, and show this covariance matrix is larger than that of GLS estimator, $\widehat{\beta}_{GLS}$.

QUESTION 5 Derive the maximum likelihood estimator for the model

$$y_i = X_i\beta + \epsilon_i, \quad i = 1, \ldots, N$$

where ϵ_i are *IID double exponential* random variables with mean 0 and scale parameter σ^2:

$$f(\epsilon) = K \exp\left\{ -\left| \frac{\epsilon}{\sqrt{\sigma^2}} \right| \right\}.$$

A. Derive a formula for K so that f is a valid probability density (i.e. so $\int_{-\infty}^{\infty} f(\epsilon)d\epsilon = 1$).

B. Derive the maximum likelihood estimator for (β, σ^2).

Hint: Show that the MLE for β is identical to the *Least Absolute Deviations* estimator $\hat{\beta}_{\text{lad}}$ defined by:

$$\hat{\beta}_{\text{lad}} = \underset{\beta}{argmin} \sum_{i=1}^N |y_i - X_i\beta|.$$

QUESTION 6 (Bayesian) Suppose $\varepsilon_i \sim N(0, \Omega)$ in SUR model of Question 4. Suppose the prior for the parameters $\theta = (\beta, \Omega^{-1})$ is given by the product of two independent priors for the $k \times 1$ vector β and the $p \times p$ inverse covariance matrix Ω^{-1}:

$$p(\theta) = p(\beta, \Omega^{-1}) = p(\beta|\beta_0, A_0^{-1})p(\Omega^{-1}|\rho_0, R_0).$$

Suppose $p(\beta|\beta_0, A_0^{-1}) = N(\beta_0, A_0^{-1})$ i.e. the prior for β is a multivariate normal distribution with hyperparameters β_0 (the mean of the prior) and A_0^{-1} (the covariance matrix of the prior). Suppose $p(\Omega^{-1}|\rho_0, R_0) = W(\rho, R_0)$, i.e. the prior for Ω^{-1} is a Wishart distribution with hyperparameters ρ (1×1) and R_0 $(p \times p)$. Recall that the density of a Wishart is given by:

$$p(\Omega^{-1}|\rho_0, R_0) = \frac{|R_0|^{-\rho_0/2}|\Omega^{-1}|^{(\rho_0-(p-1))/2}\exp\left\{-\frac{1}{2}\mathrm{tr}(R_0^{-1}\Omega^{-1})\right\}}{2^{\rho_0 p/2}\pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma\left(\frac{\rho_0+1-i}{2}\right)}.$$

Show that the posterior distribution of (β, Ω^{-1}) is *conditionally conjugate* i.e. it can be decomposed as a product of a Normal and Wishart as follows:

$$\begin{aligned}
p(\beta|\Omega^{-1}, y, X) &\sim N\left(\widehat{\beta}, A_N^{-1}\right)\\
p(\Omega^{-1}|\beta, y, X) &\sim W\left(N+\rho_0, R_N\right)
\end{aligned}$$

where

$$\begin{aligned}
\widehat{\beta} &= A_N^{-1}\left(A_0\beta_0 + \sum_{i=1}^{N} X_i'\Omega^{-1}y_i\right)\\
A_N &= \left(A_0 + \sum_{i=1}^{N} X_i'\Omega^{-1}X_i\right)\\
R_N &= \left(R_0^{-1} + \sum_{i=1}^{N}(y_i - X_i\beta)(y_i - X_i\beta)'\right)^{-1}.
\end{aligned}$$

Hint: Show that the likelihood function can be expressed as:

$$f(y_1, \ldots, y_N|\beta, \Omega^{-1}, X_1, \ldots, X_N) = \pi^{-N/2}|\Omega^{-1}|^{N/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{N}(y_i - X_i\beta)'\Omega^{-1}(y_i - X_i\beta)\right\}$$

where y_i is $p \times 1$ and X_i is $p \times k$. Also use the fact that $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ so we can write:

$$\begin{aligned}
\mathrm{tr}\left[\sum_{i=1}^{N}(y_i - X_i\beta)'\Omega^{-1}(y_i - X_i\beta)\right] &= \sum_{i=1}^{N}\mathrm{tr}\left[(y_i - X_i\beta)'\Omega^{-1}(y_i - X_i\beta)\right]\\
&= \mathrm{tr}\left[\Omega^{-1}\sum_{i=1}^{N}(y_i - X_i\beta)(y_i - X_i\beta)'\right].
\end{aligned}$$

QUESTION 7 Consider the *random utility model:*

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

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

$$v_d = X_d\beta \tag{2}$$

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:

$$\delta(\epsilon) = argmax_{d=1,\ldots,D}\left[v_d + \tilde{\epsilon}_d\right] \tag{3}$$

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:

$$P\left\{d|X\right\} = \int I\{d = \delta(\epsilon)\}f(\epsilon|X)d\epsilon \tag{4}$$

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:

$$U(v_1, \ldots, v_D, X) = E\left\{\max_{d=1,\ldots,D}[v_d + \tilde{\epsilon}_d]\Big|X\right\} = \int_{\epsilon_1}\cdots\int_{\epsilon_D}\max_{d=1,\ldots,D}[v_d+\epsilon_d]f(\epsilon_1, \ldots, \epsilon_D|X)d\epsilon_1\cdots d\epsilon_D \tag{5}$$

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

Problem: Prove the *Williams-Daly-Zachary Theorem:*

$$\frac{\partial U(v_1, \ldots, v_D, X)}{\partial v_d} = P\{d|X\} \tag{6}$$

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

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

Question 8 Consider the special case of the random utility model when $\epsilon=(\epsilon_1,\ldots,\epsilon_D)$ has a multivariate (Type I) *extreme value distribution:*

$$f(\epsilon|X) = \prod_{d=1}^{D} \exp\{-\epsilon_d\}\exp\{-\exp\{-\epsilon_d\}\}. \tag{8}$$

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

$$P\{d|X\} = \frac{\exp\{v_d\}}{\sum_{d'=1}^{D}\exp\{v_{d'}\}}. \tag{9}$$

Hint 1: Use the Williams-Daly-Zachary Theorem, showing that in the case of the extreme value distribution (8) the Social Surplus function is given by

$$U(v_1,\ldots,v_D,X) = \gamma + \log\left[\sum_{d=1}^{D}\exp\{v_d\}\right]. \tag{10}$$

where $\gamma = .577216\ldots$ is Euler's constant.
Hint 2: To derive equation (9) 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:

$$E\{\tilde{\epsilon}\} = \int_{-\infty}^{+\infty} \epsilon\exp\{-\epsilon\}\exp\{-\exp\{-\epsilon\}\}\,d\epsilon = \alpha+\sigma\gamma, \tag{11}$$

and the CDF is given by

$$F(x|\alpha,\sigma) = P\{\tilde{\epsilon}\le x|\alpha,\sigma\} = \exp\left\{-\exp\left\{\frac{-(x-\alpha)}{\sigma}\right\}\right\}. \tag{12}$$

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

$$\alpha = \sigma\log\left[\sum_{d=1}^{D}\exp\{\alpha_d/\sigma\}\right] \tag{13}$$