Practice Final

Question 1

A researcher is studying the effect of an endogenous explanatory variable X on a binary outcome Y. The data-generating process (DGP) is defined as follows:

$$
\begin{aligned}
Y_1 =& 1\{Y_1^\star > 0\} \\
Y_1^\star =& X\beta + Y_2\alpha + U \\
Y_2 =& X\gamma + Z\delta + V
\end{aligned}
$$

Where Z is the exogenous variable.

(a) Suppose $U, V|X, Z \sim \mathcal{N}$, write down the joint distribution under endogeneity.

(b) Define $U = \theta V + e$, derive the conditional and unconditional distribution of e under the assumption $Var(U) = 1$.

(c) Derive the conditional expectation of Y_1 by substituting $U = \theta V + e$ inside $Y_1^\star$.

(d) Define the ATE as the average treatment effect of increasing the endogenous variable Y_2 by Δ. State clearly for what distribution you are integrating, and derive an analytical expression for it.

(e) Derive the APE and explain what this object is.

(f) Suppose you ignore the endogeneity problem and decide to estimate the Probit model and the APE via MLE. What is the probability limit of $\hat{\alpha}$? Is $\widehat{APE}$ consistent for the true APE?

(g) Carefully describe how you would estimate the model using the control function approach (Standardized probit with unit variance error).

(h) Is your estimator consistent for the true α? And for the APE? Prove it.

(i) Another approach to estimate the model is via conditional MLE, based on the joint distribution $f(Y_1, Y_2|X, Z)$. Derive the log-likelihood and write down the maximization problem.

Question 2

In some empirical applications, economic agents choose one alternative from a set of alternatives to minimize a specific objective function, such as cost, regret, disutility, or loss, rather than maximizing it. To model this minimization behavior, we assume the existence of a latent function C_{ij} that encompasses an observable component $D_{ij}(\theta)$ and an unobservable component ε_{ij}, so that

$$C_{ij}(\theta) = D_{ij}(\theta) + \varepsilon_{ij} \text{ for } j = 1, \ldots, J,$$

where $C_{ij}(\theta)$ is the 'cost' that individual i incurs from choosing alternative j. Individuals then make a choice that results in the smallest C value. More specifically, the observed choice of individual i, denoted by Y_i, equals

$$Y_i = \arg \min_{j=1,\ldots,J} C_{ij}(\theta).$$

(a) Under what distributional assumptions on $\{\varepsilon_{ij}\}_{j=1}^{J}$ can we obtain

$$\Pr(Y_i = j \mid D_{i,1}(\theta), \ldots, D_{i,J}(\theta)) = \frac{\exp\{-D_{ij}(\theta)\}}{\sum_{h=1}^{J} \exp\{-D_{ih}(\theta)\}}?$$

(b) Let the assumptions in (a) hold. Assume further that

$$D_{ij}(\theta) = \alpha_j + X_{ij}\beta + W_{ij}\gamma_j + Z_i\delta_j,$$

where each of X_{ij} and W_{ij} can vary freely across both i and j. What normalizations (if any) should we impose on the parameters $\{\alpha_j\}, \beta, \{\gamma_j\}$ and $\{\delta_j\}$ so that the model becomes identified?

(c) The choice probability in (a) satisfies the IIA property. Explain what this means.

(d) Is there any way to test the assumptions in (a) against the same set of assumptions, except that ε_{iJ} is not independent of $\{\varepsilon_{ij}\}_{j=1}^{J-1}$. Explain.