Instructions:
Answer ALL questions in Parts I, II, and III
Use a separate answer book for each Part.
You have four hours to complete the exam.
Calculators and other electronic devices are not allowed.

1 Part I

Question 1:

Suppose that

\[ X \equiv \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ \rho \sigma_1 \sigma_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix} \right) \]

Let

\[ U = X_2 - \frac{\rho \sigma_2}{\sigma_1} X_1 \]

Show that \( U \) is independent of \( X_1 \).

Question 2:

Let \( X_1, \ldots, X_n \) be i.i.d. \( N(0, \theta) \). We would like to test \( H_0 : \theta = \theta' \) against \( H_1 : \theta < \theta' \), where \( \theta' \) is a fixed positive number. Prove that the set \( \{ (x_1, \ldots, x_n) : \sum_{i=1}^n x_i^2 \leq c \} \) for some \( c > 0 \) characterizes a uniformly most powerful critical region.

Question 3:

Let \( X_1, \ldots, X_n \) be i.i.d. with common PDF \( f(x) = \theta^x e^{-\theta} / x! \), \( x = 0, 1, 2, \ldots \). What is the MLE of \( \Pr[X_{10} = 1 \text{ or } 2] \)? Justify your answer.

Question 4:

Let \( X_1, \ldots, X_n \) be i.i.d. \( \sim \) Binomial \((1, \pi)\). Show that \( \bar{X} = n^{-1} \sum_{i=1}^n X_i \) has the minimal variance among all unbiased estimators. Justify your answer by explicitly invoking some theorem(s).
2 Part II

Question 1:

Suppose that you have \( n \) i.i.d. observations on \((Y_i, X_i)\) from the model:

\[
Y_i = X_i \beta_0 + \epsilon_i \\
\epsilon_i | X \sim N(0, \sigma^2 \exp(Z_i \gamma_0))
\]

where \( \beta_0 \) and \( \gamma_0 \) are unknown parameter vectors, \( \sigma^2 \) is an unknown positive constant and \( Z_i \) is a subvector of \( X_i \).

1. Define and justify the maximum likelihood estimator of \( \beta_0 \).

2. Define and justify the feasible generalized least squares estimator of \( \beta_0 \).

3. Evaluate the claim: \( \hat{\beta}_{ML} \) and \( \hat{\beta}_{FGLS} \) are numerically identical”.

4. Evaluate the claim: \( \hat{\beta}_{ML} \) and \( \hat{\beta}_{FGLS} \) are asymptotically equivalent”.

In parts (1) and (2) ‘justify’ means ‘explain why one would use this estimator’. In parts (3) and (4) ‘evaluate’ means ‘argue whether it is always true, or true under some conditions, or false’. Make sure to provide proofs for all your claims.

Question 2:

Suppose that we have an i.i.d. sample of \( n \) observations on \((Y_i, X_i)\) from the censored regression model

\[
Y_i = \max(Y_i^*, 0)
\]

where the latent variable \( Y_i^* \) is given by a linear regression model

\[
Y_i^* = X_i \beta_0 + \sigma_0 \epsilon_i
\]

where \( \sigma_0 > 0 \) is unknown and \( \epsilon_i \) is distributed according to a known continuous distribution \( F \) with density function \( f \) independent of \( X_i \).

1. Define the maximum likelihood estimator of \( \beta_0 \).

2. Derive the conditional mean of \( Y_i \) given \( X_i \), i.e. \( E(Y_i | X_i) \).

3. Derive the marginal effect of \( X_i \) on \( E(Y_i | X_i) \).

4. Evaluate the claim: “Censoring leads to an attenuation of the marginal effect of \( X_i \) relative to its effect in the latent regression.”

Note that \( F \) need NOT be the cdf of the standard normal distribution.
3 Part III

Question 1: True/Questionable/False? (No points are given for just stating true/questionable or false. The explanation is what counts.)

1. A large value of the $J$ statistic is evidence against the validity of the moment restrictions.

2. The process $X_t$ that satisfies $X_t + 1.9X_{t-1} + .88X_{t-2} = \varepsilon_t + .2\varepsilon_{t-1} + .7\varepsilon_{t-2}$ for white noise $\varepsilon_t$ is causal and invertible.

3. An MA$(q)$ process (for a finite positive integer $q$) is ergodic.

4. If $X_n = O_p(n^{-\delta})$ for some $\delta > 0$ then $X_n = o_p(1)$.

5. When instruments are weak, an applied researcher should not use a Wald statistic to test parameter hypotheses but instead use the statistic $LM_{CUE}$.

Question 2: Consider the linear simultaneous equations model

\[
\begin{align*}
Y &= X\beta + \varepsilon, \\
X &= Z\Pi + \nu,
\end{align*}
\]

where $Y$ is $n \times 1$, $X$ is $n \times d$. The observations are i.i.d. and you can assume conditional homoskedasticity. The sample size $n$ is larger than $d$. Interest focuses on estimation of $\beta$.

1. If $Z$ is an $n \times n$ invertible matrix, show that 2SLS and OLS are identical.

   Now consider 2SLS estimators $\hat{\beta}_i$ ($i = 1, 2$) based on instruments $Z_i$ of dimensions $n \times k_i$, where $k_1 < k_2$ and $Z_1$ consists of the first $k_1$ columns of $Z_2$.

2. Verify that the variance of $\hat{\beta}_2$ is smaller than the one of $\hat{\beta}_1$ (in the positive definite sense).

3. Discuss the statement “The more instruments we use for 2SLS, the better the estimator becomes”.

Question 3: Show that for a covariance stationary process $Y_t$, the linear projection of $Y_{t+1}$ on a constant and $Y_t$ is given by

\[
\hat{E}(Y_{t+1}|Y_t) = (1 - \rho_1)\mu + \rho_1 Y_t,
\]

where $\mu := E(Y_t)$, $\rho_1 := \gamma_1/\gamma_0$, and $\gamma_k$, as usually, denotes the covariance of $Y_t$ at lag $k$. Show that for an AR(2) process with AR parameters $\phi_i$ ($i = 1, 2$), the implied forecast is

\[
\mu + \phi_1/\phi_2(Y_t - \mu).
\]