July/August

Qualifying Exam: Econometrics, ~~June~~ 2024

Answer all four questions

QUESTION 1 (25 points)

(A)-(8 points) State and explain the Gauss-Markov theorem in the context of the Linear Regression model, under the assumptions (2)-(4) (table 1) and $\sum_{t=1}^{n}(x_t - \overline{x})^2 \neq 0$.

Table 1: Traditional Linear Regression model	
$Y_t = \beta_0 + \beta_1 x_t + \varepsilon_t,\ t=1,2,...,n$	
(1) $(\varepsilon_t \vert X_t=x_t) \backsim \mathsf{N}(.,.)$,	$t, s=1,2,...,n$
(2) $E(\varepsilon_t \vert X_t=x_t) = 0$,	$t, s=1,2,...,n$
(3) $Var(\varepsilon_t \vert X_t=x_t) = \sigma^2$,	$t, s=1,2,...,n$
(4) $E(\varepsilon_t \varepsilon_s \vert X_t=x_t) = 0,\ t>s$,	$t, s=1,2,...,n$

(B)-(5 points) Compare and contrast the specification in table 1 with that of table 2 in terms of:

(i) relating their assumptions (1)-(4) vs. [1]-[5] and

(ii) the possibility of assessing their validity using preliminary data analysis using graphical techniques.

Table 2: Normal, Linear Regression Model			
Statistical GM:		$Y_t = \beta_0 + \beta_1 x_t + u_t,\quad t \in \mathbb{N} := (1,2,...,n,...)$	
[1]	Normality:	$(Y_t \vert X_t=x_t) \backsim \mathsf{N}(.,.),\ f(y_t \vert x_t;\boldsymbol{\theta}) = \frac{\exp - \frac{(y_t - \beta_0 - \beta_1 x_t)^2}{2\sigma^2}}{\sigma\sqrt{2\pi}}$	$t \in \mathbb{N}$.
[2]	Linearity:	$E(Y_t \vert X_t=x_t)) = \beta_0 + \beta_1 x_t$,	$t \in \mathbb{N}$.
[3]	Homoskedasticity:	$Var(Y_t \vert X_t=x_t) = \sigma^2$,	$t \in \mathbb{N}$.
[4]	Independence:	$\{(Y_t \vert X_t=x_t),\ t \in \mathbb{N}\}$ independent process,	$t \in \mathbb{N}$.
[5]	t-invariance:	$\boldsymbol{\theta} := (\beta_0, \beta_1, \sigma^2)$ are *constant* over t,	$t \in \mathbb{N}$.
$\beta_0 = [E(Y_t) - \beta_1 E(X_t)] \in \mathbb{R},\ \beta_1 = \frac{Cov(Y_t,X_t)}{Var(X_t)} \in \mathbb{R},\ \sigma^2 = (Var(Y_t) - \frac{[Cov(Y_t,X_t)]2}{Var(X_t)}) \in \mathbb{R}_+$			

(C)-(7 points) (i) Explain why the formulae for the OLS estimators of (β_0, β_1) coincide with those of the Maximum Likelihood (ML) estimators.

(ii) Despite that, "the OLS [under (1)-(4)] and ML[under [1]–[5]] estimators of (β_0, β_1) have different finite sampling distributions and optimal *finite* sample properties". Explain.

(D)-(5 points) Discuss the limitations of the Gauss-Markov theorem for inference purposes and explain why its results are not informative enough to provide a proper frequentist test for the hypotheses:

$$H_0\colon\ \beta_1 = 0 \text{ vs. } H_1\colon\ \beta_1 \neq 0.$$

QUESTION 2 (25 points)

(A)-(10 points) Consider the simple Normal model (one parameter):

$$X_t \backsim \mathsf{NIID}(\mu, \sigma^2),\ t{=}1, 2, ..., n, ..., \tag{1}$$

with σ^2 is known, and 'NIID' stands for 'Normal, Independent and Identically Distributed'.

The sampling distribution of $\frac{\sqrt{n}(\overline{X}_n-\mu)}{\sigma}$, where $\overline{X}_n{=}\frac{1}{n}\sum_{t=1}^{n} X_t$, is often stated in traditional econometric textbooks by:

$$d(\mathbf{X}){=}\frac{\sqrt{n}(\overline{X}_n-\mu)}{\sigma} \backsim \mathsf{N}(0, 1) \tag{2}$$

(i) Explain why this is misleading by explicating what ensures $E[d(\mathbf{X})]{=}0$ and $Var[d(\mathbf{X})]{=}1$ for the sampling distribution in (2), and relate your answer to the reasoning underlying **estimation** and **hypothesis testing** for the hypotheses:

$$H_0\colon\ \mu \leq \mu_0,\ \text{vs.}\quad H_1\colon\ \mu > \mu_0. \tag{3}$$

(ii) Using your answer in (**A**)-(i) construct a $(1{-}\alpha)$ two-sided Confidence Interval (CI) for μ.

(iii) Explain why the $(1{-}\alpha)$ probability cannot be assigned to an observed CI.

(B)-(10 points)

(i) Using your answer in (**A**)-(i) construct an optimal α significance level Neyman-Pearson (N-P) test based on $d(\mathbf{X})$ and explain what 'optimal' means in terms of its relevant properties.

(ii) Using your answer in (**A**)-(i) define and explain the concepts of (a) type I error probability, (b) type II error probability, (c) power of the test and (d) the p-value, and (e) compare and contrast (a) and (d).

(C)-(5 points) State the fallacies of acceptance and rejection and explain why the accept/reject rules and the p-value are vulnerable to these fallacies when they are interpreted as providing evidence for or against a hypothesis irrespective of the significance level α and the sample size n.

Question 3

Consider a housing market analysis with some "treated" homes (e.g. located in a high-risk flood area) and "control homes."

Assume the true population models for a treated and control home i with sale prices of y_{1i} (if treated) and y_{0i} (if untreated), respectively, are given as follows:

$$\begin{aligned} y_{1i} &= \mu\left(\mathbf{x}_i, \boldsymbol{\beta}\right) + \gamma + \nu_i + \epsilon_i \\ y_{0i} &= \mu\left(\mathbf{x}_i, \boldsymbol{\beta}\right) + \nu_i + \epsilon_i, \end{aligned} \tag{1}$$

where $\mu(.)$ is some function of observed home and neighborhood characteristics $\mathbf{x}_i$ and corresponding parameters $\boldsymbol{\beta}$, γ is the treatment effect, ν_i is a spatial fixed effect (e.g. school zone), and ϵ_i is an error term capturing additional unobservables that affect home prices, with $\epsilon_i \sim n\left(0, \sigma^2\right)$.

As discussed in class, a generic Average Treatment Effect on the Treated (ATT) under 1:1 matching can be obtained by paring each treated observation with *a single* matched control to learn about the treatment effect. Individual differences are then averaged to yield a sample-level estimate for the sought treatment effect γ. Formally:

$$\hat{\gamma} = \frac{1}{n_1} \sum_{i=1}^{n_1} \left(y_{1i} - \hat{y}_{0i}\right) \tag{2}$$

where the summation is over all treated observations, and $\hat{y}_{i0}$ is the estimated counterfactual sale price for treated observation i.

Part (a) *4 points*
Letting control observations be indexed by j, what will be the estimate of $\hat{y}_{i0}$ in this case (1:1 matching)?

Part (b) *16 points*

1. Show the finite sample bias of $\hat{\gamma}_B$, i.e. $E\left(\frac{1}{n_1}\sum_{i=1}^{n_1}\left(y_{0i} - \hat{y}_{i0}\right)\right)$ by substituting the population model in (1) for y_{0i} and the estimate of $\hat{y}_{i0}$ you found in the previous question, respectively.
2. Describe the general / sufficient condition(s) that must hold for this bias to vanish.
3. How could you assure that $\nu_i = \nu_j$ for all treated observations?
4. Assume now that that $\mathbf{x}_i \neq \mathbf{x}_j$ and $\nu_i \neq \nu_j$ for a specific matched pair. Could the bias for this specific pair still vanish? If so, how? How likely is this condition going to hold for the entire sample of treated?

Question 4

Consider the Bayesian estimation of a CLRM without explanatory variables (i.e. just a constant term). At the observation level this model can be written as

$$\begin{aligned} y_i &= \mu + \epsilon_i \\ \epsilon_i &\sim n\left(0, \sigma^2\right) \end{aligned} \tag{1}$$

Thus, the only parameters in this model are the population mean μ and variance σ^2. Assume throughout that σ^2 is *known*.

You opt for a normal prior for μ, i.e.

$$\mu \sim n\left(\mu_0, V_0\right), \tag{2}$$

where μ_0 and V_0 are the prior mean and variance, respectively. (Note that both are scalars, of course).

Part (a) *5 points*

1. Write down the regression model for the full sample of n observations.
2. Write down the likelihood function for the full sample (call it $p\left(\mathbf{y}|\mu, \sigma^2\right)$).

Part (b) *5 points*
Since σ^2 is known, the conditional posterior of μ, $p\left(\mu|\sigma^2, \mathbf{y}\right)$ is the end-product for this analysis. Recall that for a CLRM with covariates, the moments for the conditional posterior can be expressed as

$$\begin{aligned} \mathbf{V}_1 &= \left(\mathbf{V}_0^{-1} + \tfrac{1}{\sigma^2}\mathbf{X}'\mathbf{X}\right)^{-1} \\ \boldsymbol{\mu}_1 &= \mathbf{V}_1\left(\mathbf{V}_0^{-1}\boldsymbol{\mu}_0 + \tfrac{1}{\sigma^2}\mathbf{X}'\mathbf{y}\right) \end{aligned} \tag{3}$$

where $\boldsymbol{\mu}_1$ and $\mathbf{V}_1$ are the conditional posterior mean and variance, respectively.

Working from these expressions, derive the conditional posterior variance of μ for your model (call it V_1). Show that it is always smaller than the prior variance V_0 for any $n, V_0 > 0$.

Part (c) *10 points*

1. Derive the conditional posterior mean (call it μ_1) and show that it can be written as a weighted average of the prior mean and the sample mean $\bar{y}$, with the weights summing to one.
2. State the condition under which the sample mean will receive a larger weight than the prior mean. Elaborate on the effect of the prior variance V_0 and the sample size n on the relative weight of the sample mean.