Qualifying Exam Questions based on ECON5125

Question 1 (25/100 points)

(**A**) [8 points]. Consider the simple Normal model:

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

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

(i) Explain why the following sampling distribution for the standardized $\overline{X}_n{=}\frac{1}{n}\sum_{t=1}^{n} X_t$:

$$\tfrac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \backsim \mathsf{N}(0, 1), \tag{1}$$

can be derived by employing two different forms of reasoning, factual (what if $\mu{=}\mu^*$) and hypothetical (what if $\mu{=}\mu_0$), explaining the difference. Note that μ^* denotes the 'true' value of μ,

(ii) Using your answer in (i) explain how (1) provides the basis for testing the hypotheses:

$$H_0\colon\ \mu \leq \mu_0, \text{ vs. } \quad H_1\colon\ \mu > \mu_0, \tag{2}$$

to give rise to an optimal Neyman-Pearson (N-P) test based on the test statistic $d(\mathbf{X}){=}\frac{\sqrt{n}(\overline{X}_n - \mu_0)}{\sigma}$, and a rejection region $C_1(\alpha)$. Define what 'optimal' means in this case.

(**B**) [7 points] (i) State the sampling distributions of $d(\mathbf{X})$ under both the null and alternative hypotheses in (2).

(ii) Using your answer in (b(i), employ probabilistic notation ($\mathbb{P}(....)$) to define:
[i] the type I error probability, [ii] the power of the test, [iii] the p-value, and
[iv] compare and contrast [i] and [iii].

(**C**) [5 points] Using your answer in (**A**) to explain how (1), evaluated under factual reasoning, can be used to construct a $(1{-}\alpha)$ Confidence interval for μ^*, the 'true' value of μ.

(**D**) [5 points] State the fallacies of acceptance and rejection, explaining why the accept and reject rules and the p-value are vulnerable to these fallacies when 'accept H_0' is interpreted as 'evidence for H_0', and 'reject H_0' is interpreted as 'evidence for H_1'.

Question 2 (25/100 points)

The traditional Linear Regression (LR) model is specified in terms of probabilistic assumptions assign to the error process $\{(\varepsilon_t|X_t{=}x_t),\ t{\in}\mathbb{N}{:=}(1,2,...,n,...)\}$:

$$Y_t{=}a_0 + a_1 x_t + \varepsilon_t,\ (\varepsilon_t|X_t{=}x_t) \backsim \mathsf{NIID}(0,\sigma_\varepsilon^2),\quad t{\in}\mathbb{N}, \tag{6}$$

where 'NIID' stands for 'Normal, Independent and Identically Distributed'.

Table 1 shows the statistical model, implicit in (6), specified in terms of probabilistic assumptions assigned to the observable process $\{\mathbf{Z}_t{:=}(Y_t, X_t),\ t{\in}\mathbb{N}\}$.

Table 1: Statistical 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:	$(Y_t\|X_t{=}x_t) \backsim \mathsf{N}(.,.)$,	$t{\in}\mathbb{N}$.
[2]	Linearity:	$E(Y_t\|X_t{=}x_t) {=} \beta_0 + \beta_1 x_t$,	
[3]	Homoskedasticity:	$Var(Y_t\|X_t{=}x_t) {=} \sigma^2$,	
[4]	Independence:	$\{(Y_t\|X_t{=}x_t),\ t{\in}\mathbb{N}\}$ independent process,	
[5]	t-invariance:	$\boldsymbol{\theta}{:=}(\beta_0, \beta_1, \sigma^2)$ are *constant* over t,	

$\beta_0{=}\ [E(Y_t){-}\beta_1 E(X_t)]{\in}\mathbb{R},\ \beta_1{=}\frac{Cov(Y_t,X_t)}{Var(X_t)}{\in}\mathbb{R},\ \sigma^2{=}(Var(Y_t) - \frac{[Cov(Y_t,X_t)]2}{Var(X_t)}){\in}\mathbb{R}_+,$

(**A**) [10 points]

(i) Explain how the Normal/Linear Regression model in Table 2 can be derived from the observable vector process $\{\mathbf{Z}_t{:=}(Y_t, X_t),\ t{\in}\mathbb{N}\}$, when one imposes the Normal, Independent and Identically Distributed assumptions, with the bivariate Normal:

$$\begin{pmatrix} Y_t \\ X_t \end{pmatrix} \backsim \mathsf{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}\right). \tag{7}$$

$$(Y_t|X_t{=}x_t) \backsim \mathsf{N}(\beta_0{+}\beta_1 x_t, \sigma^2),\ x{\in}\mathbb{R},\quad X_t \backsim \mathsf{N}(\mu_2, \sigma_{22}). \tag{8}$$

Show how the NIID assumptions reduce the joint distribution of $\{\mathbf{Z}_t{:=}(Y_t, X_t),\ t{\in}\mathbb{N}\}$.

(ii) Compare and contrast the two specifications in terms of their probabilistic assumptions, explaining why [1]-[5] in Table 1 are the only ones that matter for modeling and inference purposes.

(**B**) [7 points] The traditional aPP Autoregressive of order one model [AR(1)] is specified in terms of probabilistic assumptions assign to the error term process $\{(\varepsilon_t|X_t{=}x_t),\ t{\in}\mathbb{N}\}$:

$$Y_t{=}a_0 + a_1 Y_{t-1} + \varepsilon_t,\ (\varepsilon_t|\sigma(Y_{t-1})){\backsim}\mathsf{NIID}(0,\sigma_\varepsilon^2),\quad t{\in}\mathbb{N}{:=}(1,2,...,n,...) \tag{9}$$

Explain why assuming that the stochastic process $\{Y_t,\ t{\in}\mathbb{N}\}$ is Normal, Markov (M) and Stationary (S) gives rise to the statistical Normal AR(1) model implicit in () by specifying it using the specification in Table 1 as a template.

(**C**) [8 points] Using your answers in (**A**) and (**B**)**,** compare the statistical LR (Table 1) and the statistical AR(1) models in terms of their (i) model assumptions, (ii) statistical parameterizations as well as (iii) explain whether weak exogeneity holds for their respective parametrizations, relating to:

$$\text{LR: } f(x_t, y_t; \boldsymbol{\theta}_{LR}){=}f(y_t|x_t; \boldsymbol{\phi}_1)\cdot f_x(x_t; \boldsymbol{\phi}_2),\ \text{for all } (x_t, y_t){\in}\mathbb{R}{\times}\mathbb{R}, \tag{10}$$

$$\text{AR(1): } f(y_t, y_{t-1}; \boldsymbol{\theta}_{AR}){=}f(y_t|y_{t-1}; \boldsymbol{\phi}_1)\cdot f_x(y_{t-1}; \boldsymbol{\phi}_2),\ \text{for all } (y_t, y_{t-1}){\in}\mathbb{R}{\times}\mathbb{R}. \tag{11}$$

Question 3: (est. time = 20 minutes)

Consider the TRUE linear regression model, expressed for the full sample as:

$$\mathbf{y}^* = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{1}$$

where $\boldsymbol{\epsilon}$ has the usual CLRM properties, i.e follows a normal density with mean $\mathbf{0}$ and variance matrix $\sigma_\epsilon^2\mathbf{I}$.

Assume, however, that the *dependent variable* is measured with error, with the relationship between the observed $\mathbf{y}$ and the true $\mathbf{y}^*$ given as:

$$\begin{aligned} &\mathbf{y} = \mathbf{y}^* + \boldsymbol{\mu}, \\ &\boldsymbol{\mu} \sim n\left(\mathbf{0}, \sigma_\mu^2\mathbf{I}\right) \end{aligned} \tag{2}$$

where you can assume that $\boldsymbol{\mu}$ is uncorrelated with $\boldsymbol{\epsilon}$.

Part (a), 6 points

1. Express the model in (1) in terms of $\mathbf{y}$ and $\mathbf{X}$, and show the form of the resulting error term. Derive the error term's expectation and variance matrix.
2. Show that the OLS estimator of $\boldsymbol{\beta}$ (call it $\mathbf{b}$) remains *unbiased* in this case.
3. Overall, what is therefore the main (and only) undesirable effect introduced by the measurement error?

Part (b), 10 points

Now assume instead the measurement error on the dependent variable takes the following form:

$$\mathbf{y} = \alpha\mathbf{y}^*, \quad \alpha \neq 0 \tag{3}$$

1. Express the model in (3) in terms of $\mathbf{y}$ and $\mathbf{X}$, and show the form of the resulting error term. Derive the error term's expectation and variance matrix.
   (To be specific, your equation should start with $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + ...$)
2. Show that the OLS estimator of $\boldsymbol{\beta}$ (call it $\mathbf{b}$) becomes *biased* in this case.
3. Does this bias make intuitive sense? Explain.
4. Show the (true) expectation of $\mathbf{y}^*|\mathbf{X}$. Then show that this model also produces (not surprisingly) *biased* predictions $\hat{\mathbf{y}}$.

Question 4: (est. time = 20 minutes)

The VT student recreation office sampled at random graduating senior students who held at least one campus club membership (e.g. sports, dancing, arts & crafts, debate, etc.) during their time at Tech. Let y_i be the observed number of club memberships for sampled student i.

For the sample described, note that $\sum_{i=1}^{n} y_i = 240$, $\sum_{i=1}^{n} ln\left(y_i\right) = 214.4542$, and $n = 200$.

You first assume that the distribution of y_i can be characterized by the (continuous) Pareto distribution, given as:

$$\begin{aligned} f\left(y_i\right) &= \theta y_i^{-\theta-1} \\ F\left(y_i\right) &= 1 - y_i^{-\theta} \quad \text{with} \quad y_i \geq 1;\ \theta > 0 \end{aligned} \tag{1}$$

where $f(.)$ and $F(.)$ denote, respectively the *pdf* and *cdf* for this distribution.

Part (a), 10 points

1. Write the log-likelihood for the i^{th} observation and the full sample.
2. Derive the gradient for the sample and solve (analytically and numerically) for the MLE estimator of θ (call it $\hat{\theta}$). Report the numerical result to a precision of *four decimals*.
3. Does the second order condition confirm that your estimator of θ does indeed maximize the likelihood? Explain.
4. Given your results, what is the estimated probability that a student in this population participates held more than two club memberships, i.e. that $y_i > 2$? (Hint: Recall that the *cdf* for the Pareto density is given above). Report this numerical result to a precision of *four decimals*.

Part (b), 10 points

Now consider the *truncated-at-one Poisson* as an alternative sample distribution for y_i. Recall that the *un*-truncated Poisson density is given as:

$$f\left(y_i\right) = \frac{exp\left(-\lambda\right)\lambda^{y_i}}{y_i!}, \quad y_i = 0, 1, 2\ldots;\ \lambda > 0 \tag{2}$$

The truncated-at-one Poisson must satisfy $y_i = 1, 2, 3\ldots$, so it excludes the value of 0. Note that, given that the Poisson is a *discrete* density, the truncated Poisson can be generically expressed as:

$$f\left(y_i | y_i > 0\right) = \frac{f\left(y_i\right)}{Pr\left(y_i > 0\right)} = \frac{f\left(y_i\right)}{1 - f\left(0\right)} \tag{3}$$

1. Derive the explicit form of the density for the truncated-at-one Poisson. Simplify as much as possible, such that only λ^{y_i} remains in the numerator of the resulting fraction.
2. Write the log-likelihood for the *ith* observation and the full sample.
3. Suppose that the maximum likelihood estimator of λ for this sample is 2.56. What is the estimated probability that a student in this population participated in more than two club organizations?
4. *In words*, how would you decide on which model (Pareto or truncated Poisson) better fits the data in a classical estimation framework, using predictive measures of fit (list at least 3 examples)?