Qualifying Exam: Econometrics, June 2024

**Answer all four questions**

**Question 1**-(25 points)

(**a**)-(5 points) Explain how the Normal/Linear Regression model in table 1 can be viewed as a parameterization of the vector process $\{\mathbf{Z}_t:=(Y_t, X_t),\ t\in\mathbb{N}:=(1,2,...,n,...)\}$, assumed to be Normal (N), Independent (I) and Identically Distributed (ID), $\mathbf{Z}_t \backsim \mathsf{NIID}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),\ t\in\mathbb{N}$, stemming from the distribution of the sample $f(\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_n; \boldsymbol{\phi})$, where the relevant bivariate Normal distribution takes the form:

$$\begin{pmatrix} Y_t \\ X_t \end{pmatrix} \backsim \mathsf{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}\right). \tag{1}$$

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

(**b**)-(10 points) Show how the parametrizations of $(\beta_0, \beta_1, \sigma^2)$ in Table 1 are determined by (1).

Define and explain the concept of weak exogeneity in the context of the reduction:

$$f(x, y; \boldsymbol{\theta}) = f(y|x; \boldsymbol{\phi}_1) \cdot f_x(x; \boldsymbol{\phi}_2),\ \text{for all } (x, y)\in\mathbb{R}_X \times \mathbb{R}_Y, \tag{3}$$

and apply it to the case of the Normal/Linear Regression specified in **(a)-(b)**.

(**c**)-(5 points) Explain why assuming that the stochastic process $\{Y_t,\ t\in\mathbb{N}\}$ is Normal (N), Markov (M) and Stationary (S) simplifies the joint distribution of the sample $\mathbf{Y}:=(Y_1, Y_2, ..., Y_n)$, and gives rise the Normal AR(1) model; Explain each step.

(**d**)-(5 points) Using your answer in **(c),** specify the Normal AR(1) model using the specification in table 1 as a template, and compare the two models in terms of their (i) model assumptions, (ii) statistical parameterizations as well as (iii) how they relate to weak exogeneity.

**Table 1: Normal, Linear Regression Model**

Statistical GM: $Y_t = \beta_0 + \beta_1 x_t + u_t,\quad t\in\mathbb{N}$

| | | | |
|---|---|---|---|
| [1] | Normality: | $(Y_t\|X_t{=}x_t) \backsim \mathsf{N}(.,.)$, | $t\in\mathbb{N}$. |
| [2] | Linearity: | $E\left(Y_t\|X_t{=}x_t\right) = \beta_0 + \beta_1 x_t$, | |
| [3] | Homoskedasticity: | $Var\left(Y_t\|X_t{=}x_t\right) = \sigma^2$, | |
| [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 $t$, | |

Statistical parametrizations of $(\beta_0, \beta_1, \sigma^2)$?

**Question 2**-(25 points)

**(a)**-(5 points) State the Neyman-Pearson (N-P) lemma, and explain why invoking this lemma in the context of the simple Bernoulli model, $X_t \backsim \mathsf{BerIID}(\theta, (1-\theta)),\ t\in\mathbb{N}$, to derive an optimal test for the simple vs. hypotheses:

$$H_0\colon\ \theta=\theta_0,\ \text{vs.}\quad H_1\colon\ \theta=\theta_1,\ \ \theta_1 > \theta_0, \tag{4}$$

constitutes a blatant abuse of the original lemma. Explain how the result of the N-P lemma can be extend to more realistic cases including the simple Bernoulli model with $\theta\in(0,1)$.

**(b)**-(3 points) Explain how you would reformulate the null ($H_0$) and alternative ($H_1$) hypotheses in (4) to accord with the archetypal N-P formulation to properly frame the N-P hypotheses to avoid fallacious results.

**(c)**-(8 points) Define and explain the concepts of the type I ($\alpha$) and type II ($\beta$) error probabilities and the power of a test. Explain why the traditional choice of $\alpha$=.05 could easily give rise to the small $n$ (sample size) or large $n$ problems because of the in-built trade-off between $\alpha$ and $\beta$. Are there any ways to address the small and large $n$ problems in practice?

**(d)**-(5 points) Define and explain the following properties of N-P tests: (i) Uniformly Most Powerful test, (ii) Unbiased test, and (iii) Consistent test.

**(e)**-(4 points) Discuss the notion of statistical adequacy and explain what is likely to go wrong when any of the model assumptions are invalid for the data in question.

## 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:

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

where $\mu(.)$ is some function of observed home and neighborhood characteristics $\mathbf{x}_i$ and corresponding parameters $\boldsymbol{\beta}$, $\gamma$ is the treatment effect, and $\epsilon_i$ is an error term capturing additional unobservables that affect home prices, with $E\left(\epsilon_i|\mathbf{x}_i\right) = 0$.

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

$$\begin{aligned} \hat{\gamma} &= \frac{1}{n_1} \sum_{i=1}^{n_1} \left(y_{1i} - \hat{y}_{0i}\right), \quad \text{with} \\ \hat{y}_{0i} &= \tfrac{1}{M} \sum_{j \in L_{i,M}} y_{0j}, \end{aligned} \tag{2}$$

where $n_1$ denotes the total number of treated observations, $\hat{y}_{i0}$ is the estimated counterfactual sale price for treated observation $i$, $L_{i,M}$ is the set of $M$ controls that were matched to treated observation $i$, and $y_{0j}$ is the observed sale price of control observation $j$.

**Part (a)**
First assume 1:1 matching, i.e. $M = 1$, such that $\hat{y}_{0i} = y_{0j}$, the price of the single matched control. Let the finite sample bias of the basic matching estimator be given as $B\left(\hat{\gamma}_B\right) = E\left(\frac{1}{n_1} \sum_{i=1}^{n_1} \left(y_{0i} - \hat{y}_{i0}\right)\right)$.

1. Express this bias as a function of the population model in (1), and show formally the condition under which it goes to zero.

2. Using intuition, how likely is this condition to hold as (i) the dimension of $\mathbf{x}$ increases, and / or (ii) the number of continuous variables in $\mathbf{x}$ increases?

**Part (b)**
Continuing with 1:1 matching, now consider an adjusted matching estimator where the estimated counterfactual takes the following form:

$$\hat{y}_{0i} = y_{0j} + \hat{\mu}\left(\mathbf{x}_i, \boldsymbol{\beta}\right) - \hat{\mu}\left(\mathbf{x}_j, \boldsymbol{\beta}\right), \tag{3}$$

where $\hat{\mu}(.)$ is an estimate of $\mu(.)$, using for example a linear or nonparametric regression model. Show that in this case there are two alternative conditions for the bias of $\hat{\gamma}$ to vanish (i.e. the bias goes to zero if *either* of them holds).

**Part (c)**

Now consider the case of matching with multiple controls, i.e. $M > 1$, with counterfactual $\hat{y}_{0i} = \frac{1}{M} \sum_{j \in L_{i,M}} y_{0j}$ for the basic case, and $\hat{y}_{0i} = \hat{\mu}\left(\mathbf{x}_i, \boldsymbol{\beta}\right) + \frac{1}{M} \sum_{j \in L_{i,M}} \left(y_{0j} - \hat{\mu}\left(\mathbf{x}_j, \boldsymbol{\beta}\right)\right)$ for the adjusted case.

Show formally that the results from parts (a) and (b) hold for this situation as well.

## Question 4

Consider the Poisson model for a random integer variate $y$ with parameter $\lambda$, given as

$$\begin{aligned} p(y|\lambda) &= \frac{\lambda^y exp(-\lambda)}{y!}, \quad \text{with} \\ E(y|\lambda) &= V(y|\lambda) = \lambda, \quad \lambda > 0, y \in \{0, 1, 2, 3, ....\} \end{aligned} \tag{1}$$

**Part (a)**
Now consider a sample of $n$ observations from this distribution, with each observation generically labeled $y_i, i = 1 \dots n$.

1. Write down the log-likelihood function for the full sample (call it $L(\lambda)$).
2. Derive the sample gradient $g(\lambda)$ and solve for the Maximum Likelihood Estimator $\hat{\lambda}$.
3. Derive the sample Hessian $H(\lambda)$ and show that the MLE solution is indeed a maximum.
4. Derive the information matrix $I(\lambda)$.
5. Show that the score identity holds $(E_y(g(.)) = 0)$
6. Show that the information matrix identity holds $(V_y(g(\lambda)) = I(\lambda))$

**Part (b)**
Now consider the Bayesian approach to learning about $\lambda$.

Suppose you stipulate a *gamma* prior density for $\lambda$ with shape parameter $a$ and inverse scale ("rate") parameter $b$, given as

$$\begin{aligned} p(\lambda) &= g(a, b) = \frac{b^a}{\Gamma(a)} \lambda^{(a-1)} exp(-b\lambda), \quad \text{with} \\ E(\lambda) &= \frac{a}{b}, \; V(\lambda) = \frac{a}{b^2}, \quad \lambda, a, b > 0, \end{aligned} \tag{2}$$

1. Write down the joint distribution for the sample data (in *un*-logged form). Call it $p(\mathbf{y}|\lambda)$.
2. Show that the posterior distribution of $\lambda$, given your collected data from the Poisson, is also a gamma. Show the form of the posterior shape and rate parameters (you can call them $a^*$ and $b^*$).

**Part (c)**
Show that the posterior expectation can be written as a weighted average of the prior expectation and the MLE solution. What happens to this posterior expectation as $n \to \infty$?