Departments of Economics and of Agricultural and Applied Economics

Ph.D. Qualifying Exam August 2023

PART III:

Econometrics

Aug. 3, 2023

Question 1 (20 minutes)

1. (a) State and explain the Gauss-Markov theorem in the context of the Linear Regression (LR) model, assuming that $\sum_{t=1}^{n}(x_t-\overline{x})^2\neq 0$ and some of the probabilistic assumptions in table 1 hold.

Table 1: Traditional Linear Regression model

$$Y_t=\beta_0+\beta_1 x_t+\varepsilon_t,\;\; t{=}1,2,...,n$$

$$\left.\begin{array}{l}(1)\;(\varepsilon_t|X_t{=}x_t)\backsim \mathsf{N}(.,.),\\ (2)\;E\left(\varepsilon_t|X_t{=}x_t\right)=0,\\ (3)\;Var\left(\varepsilon_t|X_t{=}x_t\right)=\sigma^2,\\ (4)\;E\left(\varepsilon_t\varepsilon_s|X_t{=}x_t\right){=}0,\;\; t\neq s,\end{array}\right\}\; t,s{=}1,2,...,n$$

(b) Compare and contrast the specification in table 1 with that of table 2 in terms of

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

(ii) the difficulty in assessing their validity using a preliminary data analysis.

(iii) Explain why the additional assumption in traditional textbook econometrics:

(5) the LR model in Table 1 includes all the relevant explanatory variables (X_t),

cannot be a statistical assumption.

Table 2: Normal, Linear Regression Model

$$\begin{array}{lll}\text{Statistical GM:} & Y_t=\beta_0+\beta_1 x_t+u_t, & t\in\mathbb{N}{:=}(1,2,...,n,...)\end{array}$$

$$\left.\begin{array}{lll}[1] & \text{Normality:} & (Y_t|X_t{=}x_t)\backsim\mathsf{N}(.,.),\;\; f(y_t|x_t;\boldsymbol{\theta}){=}(\frac{1}{\sigma\sqrt{2\pi}})\exp\{-(y_t{-}\beta_0{-}\beta_1 x_t)^2/2\sigma^2\}\\ [2] & \text{Linearity:} & E\left(Y_t|X_t{=}x_t\right)){=}\beta_0+\beta_1 x_t,\\ [3] & \text{Homoskedasticity:} & Var\left(Y_t|X_t{=}x_t\right){=}\sigma^2,\\ [4] & \text{Independence:} & \{(Y_t|X_t{=}x_t)\,,\; t\in\mathbb{N}\}\text{ independent process,}\\ [5] & \text{t-invariance:} & \boldsymbol{\theta}{:=}\left(\beta_0,\beta_1,\sigma^2\right)\text{ are } \textit{constant} \text{ over } t,\end{array}\right\}\; t\in\mathbb{N}.$$

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

(c) (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 (2)-(4)] and ML[under [1]–[5]] estimators of (β_0,β_1) have different finite sampling distributions and optimal *finite* sample properties". Explain.

(d) Discuss the limitations of the Gauss-Markov theorem for inference purposes and explain why its results are not informative enough to test the hypotheses: H_0: $\beta_1{=}0$ vs. H_1: $\beta_1\neq 0$.

Question 2: Weighted Regression (20 points)

Consider a linear regression model where every observation $i = 1 \dots n$ receives a known weight $\sqrt{w_i}$, leading to the following model at the individual level:

$$\begin{aligned} &\sqrt{w_i} y_i = \sqrt{w_i}\mathbf{x}_i'\boldsymbol{\beta} + \sqrt{w_i}\epsilon_i, \quad \text{with} \\ &\epsilon_i \sim n\left(0, \sigma^2\right) \end{aligned} \tag{1}$$

As usual, $\mathbf{x}_i$ comprises a set of k explanatory variables, $\boldsymbol{\beta}$ is the corresponding vector of coefficients, and the *unweighted* error term follows the typical i.i.d. normal density with zero mean and common variance σ^2.

Let $\mathbf{W}^{1/2}$ be an n by n diagonal matrix with the individual weights on the diagonal, and $\mathbf{W} = \mathbf{W}^{1/2} * \mathbf{W}^{1/2}$ the corresponding diagonal matrix with the squared weights on the diagonal. To be perfectly clear:

$$\mathbf{W}^{1/2} = \begin{bmatrix} \sqrt{w_1} & 0 & \dots & 0 \\ 0 & \sqrt{w_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \sqrt{w_n} \end{bmatrix} \quad \mathbf{W} = \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & w_n \end{bmatrix} \tag{2}$$

The full-sample model in *weighted form* can then be written as:

$$\begin{aligned} &\mathbf{W}^{1/2}\mathbf{y} = \mathbf{W}^{1/2}\mathbf{X}\boldsymbol{\beta} + \mathbf{W}^{1/2}\boldsymbol{\epsilon}, \quad \text{or} \\ &\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{\epsilon}} \end{aligned} \tag{3}$$

where $\mathbf{y}$ is the n by 1 vector of outcomes, $\mathbf{X}$ is the full n by k matrix of explanatory variables, and $\boldsymbol{\epsilon}$ is the n by 1 vector of error terms.

It naturally follows that

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{4}$$

Part (a), 9 points

1. Derive the variance-covariance matrix for the weighted error vector $\tilde{\boldsymbol{\epsilon}}$. How, if at all, does it deviate from the CLRM assumptions?
2. Using the OLS formula on the transformed data (second line in (3)), derive the least squares estimator $\tilde{\mathbf{b}}$, and show it in terms of $\mathbf{y}$, $\mathbf{X}$, and $\mathbf{W}$.
3. Derive the expectation and variance-covariance matrix for this estimator, given $\mathbf{X}$.
4. Show that the variance of the estimator reduces to the standard OLS variance if every observation receives an equal weight of 1.
   What would be the consequences for inference if the weighted model is the true model, but the analyst uses the standard expression for the variance of the OLS estimator?

Part (b), 6 points

1. Consider the vector of predicted *weighted* outcomes, $\hat{\tilde{\mathbf{y}}} = \mathbf{W}^{1/2}\hat{\mathbf{y}} = \tilde{\mathbf{X}}\tilde{\mathbf{b}}$. Solve for the unweighted predictions $\hat{\mathbf{y}}$ in terms of $\mathbf{y}$, $\mathbf{X}$, and $\mathbf{W}$.
2. Consider the unweighted residuals $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \tilde{\mathbf{M}}\mathbf{y}$. Solve for $\tilde{\mathbf{M}}$ in terms of $\mathbf{y}$, $\mathbf{X}$, and $\mathbf{W}$.
3. Show that $\mathbf{e} = \tilde{\mathbf{M}}\boldsymbol{\epsilon}$.

Part (c), 5 points

Assume you are interested in the agricultural output of farm i in county $c = 1 \dots C$, given by

$$\begin{aligned} &y_{ic} = \mathbf{x}_{ic}'\boldsymbol{\beta} + \epsilon_{ic}, \quad \text{with} \\ &\epsilon_{ic} \sim n\left(0, \sigma^2\right) \end{aligned} \tag{5}$$

where y_{ic} is output, $\mathbf{x}_{ic}$ is a vector of farm characteristics, and the farm-specific regression error ϵ_{ic} has the typical CLRM properties.

Since you were unable to obtain information from some farms, you decide to instead estimate a model of *county-level average output*, i.e.:

$$\begin{aligned} &\bar{y}_c = \bar{\mathbf{x}}_c'\boldsymbol{\gamma} + \bar{\epsilon}_c, \quad \text{with} \\ &\bar{y}_c = \tfrac{1}{n}\sum_{i=1}^{n_c} y_{ic},\ \bar{\mathbf{x}}_c = \tfrac{1}{n}\sum_{i=1}^{n_c} \mathbf{x}_{ic},\ \bar{\epsilon}_c = \tfrac{1}{n}\sum_{i=1}^{n_c} \epsilon_{ic}, \end{aligned} \tag{6}$$

where n denotes the total number of counties, and n_c captures the number of farms in county c.

1. Derive the variance of the averaged error $\bar{\epsilon}_c$. How, if at all, does it deviate from the CLRM assumptions?
2. Let $\boldsymbol{\epsilon}$ be the n by 1 vector of averaged errors for the entire sample of counties. Derive the variance-covariance matrix of $\boldsymbol{\epsilon}$ (you can call it $\boldsymbol{\Omega}$).
3. Knowing $\boldsymbol{\Omega}$, what would be the most efficient estimator for $\boldsymbol{\gamma}$ (call it $\hat{\boldsymbol{\gamma}}$)? Show its form (you do not need to derive it mathematically).

Question 3: Bayesian problem (20 points)

Consider the Exponential model for a random variate y with *scale* parameter λ, given as

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

Part (a), 4 points
Now consider a sample of n observations from this distribution, with each observation generically labeled $y_i, i = 1 \ldots n$.

Suppose you stipulate a *inverse gamma* prior density for λ with shape parameter ν_0 and scale parameter τ_0, given as

$$\begin{aligned} p\left(\lambda\right) &= ig\left(\nu_0, \tau_0\right) = \frac{\tau_0^{\nu_0}}{\Gamma\left(\nu_0\right)} \lambda^{-(\nu_0+1)} exp\left(-\tau_0 \lambda^{-1}\right), \quad \text{with} \\ E\left(\lambda\right) &= \frac{\tau_0}{\nu_0 - 1},\ V\left(\lambda\right) = \frac{\tau_0^2}{\left(\nu_0-1\right)^2\left(\nu_0-2\right)}, \quad \lambda, \nu_0, \tau_0 > 0, \end{aligned} \tag{2}$$

1. Write down the joint distribution for the sample data (in *un*-logged form). Call it $p\left(\mathbf{y}|\lambda\right)$.
2. Show that the posterior distribution of λ, given your collected data from the Exponential, is also an inverse-gamma. Show the form of the posterior shape and scale parameters (you can call them ν_1 and τ_1).

Part (b), 10 points

1. Derive the posterior expectation of λ.
2. Derive the MLE estimate for λ, call it $\hat{\lambda}$. Also derive the standard error of $\hat{\lambda}$.
3. Show that the posterior expectation can be written as a weighted average of the prior expectation and the MLE estimate. What happens to this posterior expectation as $n \to \infty$?

Part (c), 6 points
Assume you study the waiting time for patients at the Emergency Room (ER) of the local hospital. Let y_i be the waiting time by patient i, in minutes. In terms of prior information, assume you know from other nearby hospitals that the average waiting time over the last 12 months was 10 minutes for the typical patient.

Your own data shows that of 10 patients sampled over the course of one day, the sum of all wait times was 160 minutes.

1. Setting $\nu_0 = 3$, find the prior scale parameter τ_0 to set the prior expectation such that it corresponds to the average waiting time at the other hospitals (given the relationship between $E\left(y|\lambda\right)$ and λ in (1)).
2. Solve for the the posterior expectation $E\left(\lambda|\mathbf{y}\right)$ as well as the posterior variance $V\left(\lambda|\mathbf{y}\right)$ and posterior standard deviation, with precision to the third decimal.
3. Assume the hospital is prepared to add more staff to the ER room if the posterior expectation, plus three standard deviations exceeds 25 minutes. What will you recommend to the hospital's administration?