

Definition of  $\sigma$ -algebra

kernel estimator:

$$\omega(u, v) \triangleq \frac{1}{h} k\left(\frac{u - v}{h}\right)$$

with  $h > 0$ .

**Problem 1.**

The local polynomial estimator can be derived as

$$(\tilde{\beta}_0, \dots, \tilde{\beta}_k) = \arg \min_{\beta_0, \dots, \beta_k} \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{m_i} (Y_{ij} - \sum_{j=0}^{k_i} \frac{g^{(j)}(t)}{j!} (t_{ij} - t)^j)^2 k_h(t - t_{ij})$$

where

$$k_h(v) = \frac{1}{h} k\left(\frac{v}{h}\right).$$

It implies that  $\tilde{g}_h(t) = \tilde{\beta}_0$ .

Show that, when  $k = 1$ : (local linear fit)

$$\tilde{g}_h(t) = \frac{\sum_{i=1}^n \sum_{j=1}^{m_i} \omega(t, t_{ij}) Y_{ij}}{\sum_{i=1}^n \sum_{j=1}^{m_i} \omega(t, t_{ij})}$$

where

$$\omega(t, t_{ij}) = k_h(t - t_{ij}) \{S_{n,2} + (t - t_{ij}) S_{n,2}\},$$

with

$$S_{i,j} = \sum_{i=1}^n \sum_{j=1}^{m_i} k_h(t - t_{ij}) (t - t_{ij})^j.$$

*Solution.*

■

**Problem 2.**

Explain

1. Total regression coefficient estimator
2. Partial regression coefficient estimator
3. Sequential regression coefficient estimator

and when can they be the same

*Solution.* Total Regression Coefficient Estimator:

The total regression coefficient estimator measures the relationship between a dependent variable and an independent variable while considering the combined influence of all other independent variables in the model. It quantifies the change in the dependent variable associated with a one-unit change in

the specific independent variable, assuming all other variables are held constant. Partial Regression Coefficient Estimator:

The partial regression coefficient estimator quantifies the relationship between a dependent variable and a specific independent variable while controlling for the effects of all other independent variables in the model. It isolates the unique impact of that particular independent variable on the dependent variable. Sequential Regression Coefficient Estimator:

The sequential regression coefficient estimator involves estimating regression coefficients step by step, typically by adding or removing variables from the model based on selection criteria. It iteratively adjusts the coefficients as variables are added or removed, aiming to find the best subset of predictors. When Can They Be the Same:

The total regression coefficient estimator, partial regression coefficient estimator, and sequential regression coefficient estimator can be the same under certain condition:

Simple Linear Regression:

When you are dealing with a simple linear regression model (i.e., there is only one independent variable), all three estimators will be the same. This is because there are no other independent variables to control for or select from, and the total, partial, and sequential estimators will all measure the relationship between the single independent variable and the dependent variable in the same way.

■

### Problem 3. Time Series Correlation Structure

The first-order autoregressive model with equally spaced time points:

$$\varepsilon_{i,j} = \rho \varepsilon_{i,j-1} + Z_{i,j} \text{ with } Z_{i,j} \stackrel{\text{iid}}{\sim} (0, (1 - \rho^2)\sigma^2), |\rho| < 1.$$

Show that

$$\text{Cov}[\varepsilon_{i,j}, \varepsilon_{i,j-k}] = \rho^k \sigma^2$$

### Problem 4. Cross-Sectional Data

For Cross-Sectional Data

$$Y_i = X_i^T \beta + \varepsilon_i, \varepsilon \stackrel{\text{iid}}{\sim} (0, \sigma^2),$$

prove that the least square estimator  $\hat{\beta} = (X^T X)^{-1} X^T Y$  is the best linear unbiased estimator (BLUE) of  $\beta$ .

### Problem 5.

GLM:  $Y = X\beta + \varepsilon$ , where  $\varepsilon \sim (0, \sigma^2 V)$ . Let  $A = I - X(X^T X)^{-1} X^T$  and  $B$  be the  $N \times (N - (p - 1))$  matrix such that  $BB^T = A$  and  $B^T B = I_{N - (p - 1)}$ .

For any fixed  $V$ , the MLE of  $\beta$  is  $\hat{\beta}(V) = GY$ , with  $G = (X^T V^{-1} X)^{-1} X^T V^{-1}$ . Show that

$$|GG^T - GBB^T G^T|^{\frac{1}{2}} = |X^T X|^{-\frac{1}{2}}$$

**Problem 6. Serial correlation plus measurement error**

Measurement error:  $\varepsilon \sim (0, \sigma^2 H + \tau^2 I_N)$ .

Show that the corresponding variogram is  $r(u) = \tau^2 + \sigma^2(1 - \rho(u))$  for  $u \geq 0$ , with  $r(0) = \tau^2 > 0$ .

**Problem 7. Repeated Measures**

$$y_{hij} = \beta_h + r_{hj} + U_{hi} + z_{hij}, h = 1, \dots, g, i = 1, \dots, n_h, j = 1, \dots, m,$$

where  $\beta_h$  is the main effect for the  $h$ -th treatment, and  $r_{hj}$  is the interaction effect between the  $h$ -th treatment and the  $j$ -th time with  $\sum_{j=1}^m r_{hj} = 0, \forall h$ .

For the hypothesis  $H_0 : \beta_1 = \dots = \beta_g$  find the F statistics  $F = \frac{MBTSS_1}{MRSS_1}$  of the ANOVA table.

*Solution.*

■