Correlated Data Analysis: Modeling, Analytics and Applications

PETER X.-K. SONG

Problem Set 1

Problem 1.1 To study the impact of ignoring autocorrelation in statistical inference, consider a longitudinal data of equal length $Y_i = (Y_{i1}, \ldots, Y_{in}), i = 1, \ldots, K$, with the first two moments given by

$$\mathrm{E}(Y_{ij}) = \beta_0 + \beta_1 t_{ij}, \quad \mathrm{Var}(Y_{ij}) = \sigma^2,$$

and

$$\mathrm{cov}(Y_{ij}, Y_{ij'}) = \sigma^2 \rho, \rho \in (0,1), \ j \neq j' = 1, \ldots, n.$$

$V(\rho) = \mathrm{Var}_\rho(\widehat{\beta}_1)$, the variance of the weighted least squares estimator $\widehat{\beta}_1$.
(a) Prove that

$$\frac{V(0)}{V(\rho)} = \frac{(1-\rho) + n\rho(1-\phi)}{(1-\rho)(1+(n-1)\rho)}, \ \rho > 0.$$

where $\phi = n\sum_{i=1}^K (\bar{t}_i - \bar{t})^2 / \sum_i \sum_j (t_{ij} - \bar{t})^2$.
(b) Use R software to draw the plot of the logarithm of $V(0)/V(\rho)$ versus ρ for selected length of time series, $n = (2, 5, 10)$ and $\phi = (0.0, 0.2, 0.5, 0.8, 1.0)$. Comment on these plots. Note that the extreme cases $\phi = 0.0$ and $\phi = 1.0$ correspond to the within-subject comparison and the between-subject comparison.