Problem Set 2

Buan 6340

Due: Wed, 2018-06-20-11:59pm

Linear Algebra Questions

Question 1 We sometimes want to think about a matrix a partition of submatrices. Accordingly, we might define:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

Using this definition, it can be proven that:

$$A^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right)^{-1} & -\left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right)^{-1} A_{12}A_{22}^{-1} \\ -\left(A_{22} - A_{21}A_{11}^{-1}A_{12}\right)^{-1} A_{21}A_{11}^{-1} & \left(A_{22} - A_{21}A_{11}^{-1}A_{12}\right)^{-1} \end{bmatrix}$$

Consider a linear model

$$y = X\beta_x + Z\beta_z + e$$

Using the partition matrix inverse, find formulas for $\hat{\beta}_x$ and $\hat{\beta}_z$.

Question 2 Suppose I propose the following two formulas

$$\widehat{Cov}\left(x, y\right) = \frac{x' M_0 y}{n-1}, \quad \widehat{Var}\left(x\right) = \frac{x' M_0 x}{n-1}$$

What M_0 will make the above formulas true? How does M_0 relate to the formulas from Question 1?

Programming Questions

Question 3 Consider the following data generating process (DGP). $\{y_t\}$ is defined by

$$y_t = y_{t-1} + e_t, \quad e_t \sim iid\mathcal{N}\left(0, \sigma^2\right)$$

for $t = 1, \ldots, T$ and $\{x_{ti}\}$ is defined by

$$x_{ti} = x_{t-1,i} + u_{ti}, \; u_{ti} \sim iid\mathcal{N}\left(0, \sigma^2\right)$$

where $t = 1, \ldots, T$ and $i = 1, \ldots, n$. Clearly, x_{ti} and y_t are completely independent. Construct monte carlo simulations where you generate $\{y_t\}$ and $\{x_t\}$ using this DGP. In the simulation, run the following model using ordinary least squares for different input values and compute the R^2 statistic:

$$y_t = \beta_0 + \beta_1 x_{t1} + \cdots + \beta_r x_{tr} + \varepsilon_t$$

Find the limiting values of R^2 as $T \to \infty$ for the cases $n = 1, 2, 3, 5, 10$. Once you have completed that, repeat this simulation with the following DGP and compare your findings:

$$\begin{aligned} y_t &= e_t, \quad e_t \sim iid\mathcal{N}\left(0, \sigma^2\right) \\ x_{ti} &= u_{ti}, \; u_{ti} \sim iid\mathcal{N}\left(0, \sigma^2\right). \end{aligned}$$

Question 4 Consider the following DGP for $\{y_t\}$:

$$y_t = \rho y_{t-1} + e_t,\ e_t \sim iid\mathcal{N}\left(0, \sigma^2\right)$$

for $t = 1, \ldots, T$. When $|\rho| < 1$, $\{y_t\}$ is stationary. We are only concerned here with the stationary case. Run the following two OLS regressions:

$$\begin{aligned} y_t &= \beta_1 y_{t-1} + \varepsilon_t & (1) \\ y_t &= \beta_0 + \beta_1 y_{t-1} + \varepsilon_t & (2) \end{aligned}$$

Using simulation and by changing parameter values, find an approximate equation for the bias of $\hat{\beta}_1$ ($E\left[\hat{\beta}_1 - \rho\right]$) in the above regressions. [Hint: there is no bias in (1) but there is in (2).] Then by using simulation, show that $\hat{\beta}_1$ is consistent even though it may be biased ($E\left[\hat{\beta}_1\right] \rightarrow \rho$ and $Var\left[\hat{\beta}_1\right] \rightarrow 0$ as $T \rightarrow \infty$) in both regressions. In some panel data models, analysts often to include lags of the dependent variable as controls. Based on your findings from this question, when do you think this might cause a problem?