1. Consider the classic least squares model, $y = X\beta + u$ with $\mathbb{E}(u \mid X) = 0$, $\mathbb{V}\mathrm{ar}(u \mid X) = \sigma^2 \times I$. Here I is $n \times n$ identity matrix, X is $n \times k$ matrix with linearly independent columns.

   The estimator $\hat{\beta}$ is obtained by ordinary least squares, $\hat{y}$ is the vector of in-sample forecasts, the residual vector is $\hat{u} = \hat{y} - y$.

   (a) Find $\mathbb{E}(\hat{y} \mid X)$, $\mathbb{E}(\hat{u} \mid X)$.

   (b) Find $\mathbb{V}\mathrm{ar}(\hat{u} \mid X)$, $\mathbb{V}\mathrm{ar}(\hat{y} \mid X)$.

   (c) Find $\mathbb{C}\mathrm{ov}(\hat{u}, \hat{\beta} \mid X)$, $\mathbb{C}\mathrm{ov}(\hat{u}, \hat{y} \mid X)$.

2. Consider a simple linear regression $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$ with L^2 penalty,

$$\sum_{i=1}^{n} (\hat{y}_i - y_i)^2 + \lambda \cdot \hat{\beta}_2^2 \to \min_{\hat{\beta}_1, \hat{\beta}_2}.$$

   (a) Find $\hat{\beta}_1$ and $\hat{\beta}_2$ for the penalized problem.

   (b) Prove that you can obtain exactly the same $\hat{\beta}_1$ and $\hat{\beta}_2$ just by adding some special additional observations to the initial dataset and using vanilla least squares without penalty. Describe these special additional observations explicitely.

   Comment: this idea of equivalence between penalty and adding observations is widely used in bayesian vector autoregressive models.

3. Consider the matrix A

$$A = \begin{pmatrix} 2 & 5 \\ 2 & 5 \\ 4 & 6 \\ 4 & 7 \\ 3 & 7 \end{pmatrix}$$

   (a) Standardize the columns of A. Let's denote the new matrix by X.

   (b) Calculate $X^T X$, its eigenvalues and eigenvectors.

   (c) Calculate $X X^T$, its eigenvalues and eigenvectors.

   (d) Calculate singular value decomposition of X.

   (e) Calculate the first principal component of X.