Problem Set 1
Due Thursday, January 17

1.) Let $X$ be a $(T \times k)$ matrix whose columns are linearly independent, and let $M = I_T - X(X'X)^{-1}X'$. Show that $M$ is symmetric and idempotent. Calculate the eigenvalues and rank of $M$, and show that it is positive semidefinite.

2.) Consider a regression of $y_t$ on $x_t$ where the first element of $x_t$ is a constant term. The $R^2$ or coefficient of determination is defined as

\[
R^2 = 1 - \frac{\sum_{t=1}^{T}(y_t - x'_t b)^2}{\sum_{t=1}^{T}(y_t - \bar{y})^2}
\]

where $b$ is the OLS regression coefficient and $\bar{y}$ is the sample mean. Show that $0 \leq R^2 \leq 1$.

3.) Let $P$ be a nonsingular symmetric $(k_1 \times k_1)$ matrix, $Q$ a nonsingular symmetric $(k_2 \times k_2)$ matrix, and $R$ an arbitrary $(k_1 \times k_2)$ matrix. Verify the following formula for the inverse of a partitioned matrix:

\[
\begin{bmatrix}
P & R \\
R' & Q
\end{bmatrix}^{-1} =
\begin{bmatrix}
W & -WRQ^{-1} \\
-Q^{-1}R'W & (Q^{-1} + Q^{-1}RWRQ^{-1})
\end{bmatrix}
\]

for $W = (P - RQ^{-1}R')^{-1}$.

4.) Consider a regression of $y_t$ on $x_t$, where we partition the regressors into two groups: $x_t = (x'_{1t}, x'_{2t})'$ where $k_1$ of the variables are included in the subvector $x_{1t}$ and the remaining $k_2$ variables are in $x_{2t}$:

\[
y_t = x'_{1t} \beta_1 + x'_{2t} \beta_2 + \varepsilon_t.
\]

The usual OLS regression coefficients are given by

\[
\begin{bmatrix}
b_1 \\
b_2
\end{bmatrix} = (X'X)^{-1}X'y
\]

for $X = [X_1 \ X_2]$ a $(T \times (k_1 + k_2))$ matrix and $X_i$ the $(T \times k_i)$ matrix whose $t$th row is $x'_{it}$. Use the results from question (3) to show that the OLS estimate $b_1$ could equivalently be calculated as follows: (a) regress $y_t$ on $x_{2t}$ alone and calculate the residuals $e_{2t}$, for $e_{2t}$ the $t$th element of $e_2 = M_2y$ with $M_2 = I_T - X_2(X_2'X_2)^{-1}X_2'$; regress each element of $x_{1t}$ on $x_{2t}$ and calculate the residuals $\tilde{x}_{1t}$, where $\tilde{x}'_{1t}$ is the $t$th row of $\tilde{X}_2 X_{1t}$; (c) regress $e_{2t}$ on $\tilde{x}_{1t}$, to obtain a $(k_1 \times 1)$ vector $\hat{\beta}_1$ that is numerically identical to $b_1$ given above.
5.) Consider a special case of the previous result when the second explanatory variable is the constant term in the regression, so that \( x_{2t} = 1 \) for \( t = 1, ..., T \) and \( k_2 = 1 \). Describe in words the interpretation of \( e_{2t} \) and \( \tilde{x}_{1t} \) for this case.

6.) Suppose you have performed an initial OLS regression of a scalar \( y_t \) on a \((k \times 1)\) vector of explanatory variables \( x_t \),
\[
y_t = x_t' \beta + \varepsilon_t
\]
and obtained the OLS estimates \( \hat{b} \), \( s^2 \), OLS residuals \( e = y - X\hat{b} \), and \( R^2 \). You are asked to predict the consequences for OLS estimation where you regress \( y_t \) not on the original \( x_t \) but instead on a linear transformation of the original regressors,
\[
x_t^* = Qx_t,
\]
where \( Q \) is a nonsingular \((k \times k)\) matrix, and you now perform the OLS regression
\[
y_t = x_t'^* \beta^* + \varepsilon_t^*.
\]

a.) Write the simplest possible formulas for the OLS estimates on the transformed data \( \hat{b}^* \), \( s^2 \), \( e^* \), and \( R^2 \) as functions of the original \( \hat{b} \), \( s^2 \), \( e \), and \( R^2 \).

b.) As a special case, consider the regression on a constant and a scalar \( x_t \):
\[
y_t = \beta_1 + \beta_2 x_t + \varepsilon_t.
\]
What happens to the estimated values for \( \beta_1 \) and \( \beta_2 \) if you multiply \( x_t \) by 10 and regress
\[
y_t = \beta_1^* + \beta_2^* x_t^* + \varepsilon_t^*
\]
for \( x_t^* = 10x_t \)? What is the relation between the \( t \)-statistic for testing the null hypothesis \( \beta_2 = 0 \) using the original regression and the \( t \)-statistic for testing the null hypothesis \( \beta_2^* = 0 \) on the transformed regression?