An application of result 2, part iv

We will use our result $E(\mathbf{y}'\mathbf{A}\mathbf{y}) = trace(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$ to derive the expected value of the residual sum of squares for a least-squares regression model. For the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, with $E(\boldsymbol{\varepsilon}) = 0$, $V(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2\mathbf{I}_n$, and $rank(\mathbf{X}) = k+1$, we have

$$\begin{aligned}
E(RSS) &= E[(\mathbf{y}-\widehat{\mathbf{y}})'(\mathbf{y}-\widehat{\mathbf{y}})] \\
&= E[(\mathbf{y}-\mathbf{X}\mathbf{b})'(\mathbf{y}-\mathbf{X}\mathbf{b})] \\
&= E[(\mathbf{y}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y})'(\mathbf{y}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y})] \\
&= E[(\,(\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\,\mathbf{y})'][(\,(\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\,\mathbf{y})] \\
&= E[\mathbf{y}'(\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')'(\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}] \\
&= E[\mathbf{y}'(\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')(\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}] \\
&= E[\mathbf{y}'(\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}]
\end{aligned}$$

Thus,

$$\begin{aligned}
E(RSS) &= E[\mathbf{y}'(\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}] \\
&= trace((\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{I}_n\sigma_\varepsilon^2) + (E(\mathbf{y}))'(\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')(E(\mathbf{y})) \\
&= \sigma_\varepsilon^2\; trace((\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')) + (\mathbf{X}\boldsymbol{\beta})'(\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')(\mathbf{X}\boldsymbol{\beta}) \\
&= \sigma_\varepsilon^2\; [trace(\mathbf{I}_n)-trace(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')] + (\boldsymbol{\beta}'\mathbf{X}')(\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')(\mathbf{X}\boldsymbol{\beta}) \\
&= \sigma_\varepsilon^2\; [n-trace(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')] + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\
&= \sigma_\varepsilon^2\; [n-trace(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')] \\
&= \sigma_\varepsilon^2\; [n-trace(\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1})] \\
&= \sigma_\varepsilon^2\; [n-trace(\mathbf{I}_{k+1})] \\
&= \sigma_\varepsilon^2\; [n-(k+1)] \\
&= \sigma_\varepsilon^2\; [n-k-1]
\end{aligned}$$

A matrix $\mathbf{A}$ is called idempotent if $\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \mathbf{A}$. We saw in the first part of the derivation above that the matrix $(\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$ is idempotent. We will see more idempotent matrices in Chapters 9 and 10 of the text, and we will see that they play a key role in the method of least-squares. In Chapter 10 we will give a geometric explanation of the role of idempotent matrices.