Variance Estimation

Lemma 4.1:

Suppose z is a random vector with

$$E(z) = \boldsymbol{\mu} \quad \text{and} \quad \text{Var}(z) = \boldsymbol{\Sigma}.$$

Then $\boldsymbol{E}(z'\boldsymbol{A}z) = \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu} + tr(\boldsymbol{A}\boldsymbol{\Sigma})$.

Proof of Lemma 4.1:

$$
\begin{aligned}
E(\boldsymbol{z}'\boldsymbol{A}\boldsymbol{z}) &= E(tr(\boldsymbol{z}'\boldsymbol{A}\boldsymbol{z})) \\
&= E(tr(\boldsymbol{A}\boldsymbol{z}\boldsymbol{z}')) \\
&= tr(E(\boldsymbol{A}\boldsymbol{z}\boldsymbol{z}')) \\
&= tr(\boldsymbol{A}E(\boldsymbol{z}\boldsymbol{z}')) \\
&= tr(\boldsymbol{A}(\mathrm{Var}(\boldsymbol{z}) + E(\boldsymbol{z})E(\boldsymbol{z})')) \\
&= tr(\boldsymbol{A}(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}')) \\
&= tr(\boldsymbol{A}\boldsymbol{\Sigma}) + tr(\boldsymbol{A}\boldsymbol{\mu}\boldsymbol{\mu}') \\
&= tr(\boldsymbol{A}\boldsymbol{\Sigma}) + tr(\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}) = \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu} + tr(\boldsymbol{A}\boldsymbol{\Sigma}).
\end{aligned}
$$

□

Result 4.2:

Under the GMM, an unbiased estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{\text{SSE}}{n-r},$$

where $r = rank(\boldsymbol{X})$ and

$$\begin{aligned}
\text{SSE} &= \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = (\boldsymbol{y} - \hat{\boldsymbol{y}})'(\boldsymbol{y} - \hat{\boldsymbol{y}}) = [(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y}]'(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y} \\
&= \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P_X})'(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y} = \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P_X})(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y} \\
&= \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y} \\
&= \text{“Sum of Squared Errors.”}
\end{aligned}$$

Proof of Result 4.2:

$$
\begin{aligned}
E(\text{SSE}) &= E(\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{P}_X)\boldsymbol{y}) \\
&= (E(\boldsymbol{y}))'(\boldsymbol{I}-\boldsymbol{P}_X)E(\boldsymbol{y}) + tr((\boldsymbol{I}-\boldsymbol{P}_X)\text{Var}(\boldsymbol{y})) \\
&= (\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{I}-\boldsymbol{P}_X)\boldsymbol{X}\boldsymbol{\beta} + tr((\boldsymbol{I}-\boldsymbol{P}_X)(\sigma^2\boldsymbol{I})) \\
&= (\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{X}-\boldsymbol{P}_X\boldsymbol{X})\boldsymbol{\beta} + tr(\sigma^2(\boldsymbol{I}-\boldsymbol{P}_X)) \\
&= (\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{X}-\boldsymbol{X})\boldsymbol{\beta} + \sigma^2 tr(\boldsymbol{I}-\boldsymbol{P}_X) \\
&= \sigma^2(n - rank(\boldsymbol{X})) = \sigma^2(n-r).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
E(\hat{\sigma}^2) &= E\left(\frac{\text{SSE}}{n-r}\right) \\
&= \frac{1}{n-r}E(\text{SSE}) = \frac{1}{n-r}\sigma^2(n-r) \\
&= \sigma^2.
\end{aligned}
$$

$\square$

The “Regression Sum of Squares” or “Sum of Squares for Regression” is

$$\begin{aligned} \text{SSR} = \hat{\boldsymbol{y}}'\hat{\boldsymbol{y}} &= (\boldsymbol{P_X y})'\boldsymbol{P_X y} \\ &= \boldsymbol{y}'\boldsymbol{P}_X'\boldsymbol{P_X y} \\ &= \boldsymbol{y}'\boldsymbol{P_X y}. \end{aligned}$$

Find $E(\text{SSR})$.

$$
\begin{aligned}
E(\boldsymbol{y}'\boldsymbol{P_X}\boldsymbol{y}) &= (\boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{P_X}\boldsymbol{X}\boldsymbol{\beta} + tr(\boldsymbol{P_X}(\sigma^2\boldsymbol{I})) \\
&= \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{P_X}\boldsymbol{X}\boldsymbol{\beta} + \sigma^2 tr(\boldsymbol{P_X}) \\
&= \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} + \sigma^2 rank(\boldsymbol{P_X}) \\
&= \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} + r\sigma^2.
\end{aligned}
$$

Show that the Total Sum of Squares $\boldsymbol{y}'\boldsymbol{y} = \text{SSR} + \text{SSE}$.

We know $\boldsymbol{y} = \hat{\boldsymbol{y}} + \hat{\boldsymbol{\varepsilon}}$, where $\hat{\boldsymbol{y}} = \boldsymbol{P_X y} \in \mathcal{C}(\boldsymbol{X})$ and $\hat{\boldsymbol{\varepsilon}} = (\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y} \in \mathcal{C}(\boldsymbol{X})^{\perp}$.
It follows that

$$\boldsymbol{y}'\boldsymbol{y} = (\hat{\boldsymbol{y}} + \hat{\boldsymbol{\varepsilon}})'(\hat{\boldsymbol{y}} + \hat{\boldsymbol{\varepsilon}}) = \hat{\boldsymbol{y}}'\hat{\boldsymbol{y}} + \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = \text{SSR} + \text{SSE}.$$

Written our more extensively, the argument is

$$\begin{aligned}
\boldsymbol{y}'\boldsymbol{y} &= (\boldsymbol{P_X y} + (\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y})'(\boldsymbol{P_X y} + (\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y}) \\
&= \boldsymbol{y}'\boldsymbol{P}_{\boldsymbol{X}}'\boldsymbol{P_X y} + \boldsymbol{y}'\boldsymbol{P}_{\boldsymbol{X}}'(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y} + \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P_X})'\boldsymbol{P_X y} + \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P_X})'(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y} \\
&= \boldsymbol{y}'\boldsymbol{P_X P_X y} + \boldsymbol{y}'\boldsymbol{P_X}(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y} + \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{P_X y} + \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P_X})(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y} \\
&= \boldsymbol{y}'\boldsymbol{P_X y} + \boldsymbol{y}'\boldsymbol{P_X}(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y} + \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{P_X y} + \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y} \\
&= \boldsymbol{y}'\boldsymbol{P_X y} + \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y} = \text{SSR} + \text{SSE}.
\end{aligned}$$

Even more simply,

$$\text{SSR} + \text{SSE} = \boldsymbol{y}'\boldsymbol{P_X}\boldsymbol{y} + \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y} = \boldsymbol{y}'(\boldsymbol{P_X} + (\boldsymbol{I} - \boldsymbol{P_X}))\boldsymbol{y} = \boldsymbol{y}'\boldsymbol{I}\boldsymbol{y} = \boldsymbol{y}'\boldsymbol{y}.$$