Underfitting and Overfitting

Underfitting

Suppose the true model is

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\eta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\eta}$ is an unknown fixed vector and $\boldsymbol{\varepsilon}$ satisfies the GMM.

Suppose we incorrectly assume that

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

This example of misspecifying the model is known as underfitting.

Note that $\boldsymbol{\eta}$ may equal $\boldsymbol{W}\boldsymbol{\alpha}$ for some design matrix $\boldsymbol{W}$ whose columns could contain explanatory variables excluded from $\boldsymbol{X}$.

What are the implications of underfitting?

Find $E(\boldsymbol{c}'\hat{\boldsymbol{\beta}})$.

Find $E(\hat{\sigma}^2)$.

Derivation of $E(\boldsymbol{c}'\hat{\boldsymbol{\beta}})$

$$
\begin{aligned}
E(\boldsymbol{c}'\hat{\boldsymbol{\beta}}) &= E(\boldsymbol{a}'\boldsymbol{X}\hat{\boldsymbol{\beta}}) \\
&= E(\boldsymbol{a}'\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y}) \\
&= \boldsymbol{a}'\boldsymbol{P}_{\boldsymbol{X}}E(\boldsymbol{y}) \\
&= \boldsymbol{a}'\boldsymbol{P}_{\boldsymbol{X}}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\eta}) \\
&= \boldsymbol{a}'\boldsymbol{X}\boldsymbol{\beta} + +\boldsymbol{a}'\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{\eta} \\
&= \boldsymbol{c}'\boldsymbol{\beta} + \boldsymbol{a}'\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{\eta}.
\end{aligned}
$$

$\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ is biased for $\boldsymbol{c}'\boldsymbol{\beta}$ unless $\boldsymbol{a}'\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{\eta} = 0$.

Note that if $\boldsymbol{\eta}$ is close to $\mathbf{0}$, the bias $\boldsymbol{a}'\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{\eta}$ may be small.

If $\boldsymbol{\eta} \in \mathcal{C}(\boldsymbol{X})^{\perp} = \mathcal{N}(\boldsymbol{X}')$, then

$$\boldsymbol{X}'\boldsymbol{\eta} = \mathbf{0} \Rightarrow \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{\eta} = \mathbf{0}$$

$$\Rightarrow \boldsymbol{a}'\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{\eta} = 0.$$

As an example of this last point, suppose we fit a multiple regression but omit one explanatory variable.

Suppose for our sample of n observations, the vector $\boldsymbol{x}^* = \boldsymbol{x} - \bar{x}\mathbf{1}$ contains the values of the missing variable centered so that the sample mean is zero.

If the sample covariance of the missing variable $\boldsymbol{x}$ with each of the included variables $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ is 0, then the LSE of the multiple regression coefficients $\hat{\boldsymbol{\beta}}$ will still be unbiased for $\boldsymbol{\beta}$ even though $\boldsymbol{x}$ is excluded $\because \boldsymbol{X}'\boldsymbol{x}^* = \mathbf{0}$.

Derivation of $E(\hat{\sigma}^2)$

$$
\begin{aligned}
(n-r)E(\hat{\sigma}^2) &= E(\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{P_X})\boldsymbol{y}) \\
&= (\boldsymbol{X\beta}+\boldsymbol{\eta})'(\boldsymbol{I}-\boldsymbol{P_X})(\boldsymbol{X\beta}+\boldsymbol{\eta}) + tr((\boldsymbol{I}-\boldsymbol{P_X})\sigma^2\boldsymbol{I}) \\
&= \boldsymbol{\eta}'(\boldsymbol{I}-\boldsymbol{P_X})\boldsymbol{\eta} + \sigma^2(n-r) \\
&\because \boldsymbol{X}'(\boldsymbol{I}-\boldsymbol{P_X}) = \boldsymbol{0} \quad \text{and} \quad (\boldsymbol{I}-\boldsymbol{P_X})\boldsymbol{X} = \boldsymbol{0}. \\
\therefore E(\hat{\sigma}^2) &= \frac{\boldsymbol{\eta}'(\boldsymbol{I}-\boldsymbol{P_X})\boldsymbol{\eta}}{n-r} + \sigma^2.
\end{aligned}
$$

Note that

$$\begin{aligned}\boldsymbol{\eta}'(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{\eta} &= \boldsymbol{\eta}'(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{X}})'(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{\eta} \\ &= \|(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{\eta}\|^2 \\ &= \|\boldsymbol{\eta}-\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{\eta}\|^2.\end{aligned}$$

Thus, $E(\hat{\sigma}^2) = \sigma^2$ iff

$$
\begin{aligned}
(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{\eta} = \boldsymbol{0} &\iff \boldsymbol{\eta} \in \mathcal{N}(\boldsymbol{I} - \boldsymbol{P_X}) \\
&\iff \boldsymbol{\eta} \in \mathcal{C}(\boldsymbol{P_X}) = \mathcal{C}(\boldsymbol{X}) \\
&\iff \exists\ \boldsymbol{\alpha} \ni \boldsymbol{X\alpha} = \boldsymbol{\eta} \\
&\iff \exists\ \boldsymbol{\alpha} \ni E(\boldsymbol{y}) = \boldsymbol{X\beta} + \boldsymbol{\eta} \\
&\qquad\qquad = \boldsymbol{X\beta} + \boldsymbol{X\alpha} \\
&\qquad\qquad = \boldsymbol{X}(\boldsymbol{\beta} + \boldsymbol{\alpha}) \\
&\iff E(\boldsymbol{y}) \in \mathcal{C}(\boldsymbol{X}).
\end{aligned}
$$

Example 1

Consider an experiment with two experimental units (mice in this case) for each of two treatments.

We might assume the GMM holds with

$$E(\boldsymbol{y}) = E\left(\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{bmatrix}\right) = \boldsymbol{X}\boldsymbol{\beta} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \end{bmatrix} = \begin{bmatrix} \mu + \tau_1 \\ \mu + \tau_1 \\ \mu + \tau_2 \\ \mu + \tau_2 \end{bmatrix}.$$

Example 1 (continued)

Suppose the person who conducted the experiment neglected to mention that, in each treatment group, one of the experimental units was male and the other was female.

Example 1 (continued)

Then the true model may require

$$E(\boldsymbol{y}) = \begin{bmatrix} \mu+\tau_1+\alpha/2 \\ \mu+\tau_1-\alpha/2 \\ \mu+\tau_2+\alpha/2 \\ \mu+\tau_2-\alpha/2 \end{bmatrix} = \begin{bmatrix} \mu+\tau_1 \\ \mu+\tau_1 \\ \mu+\tau_2 \\ \mu+\tau_2 \end{bmatrix} + \begin{bmatrix} \alpha/2 \\ -\alpha/2 \\ \alpha/2 \\ -\alpha/2 \end{bmatrix}$$

$$= \boldsymbol{X\beta} + \begin{bmatrix} 1/2 \\ -1/2 \\ 1/2 \\ -1/2 \end{bmatrix} \begin{bmatrix} \alpha \end{bmatrix} = \boldsymbol{X\beta} + \boldsymbol{W\alpha} = \boldsymbol{X\beta} + \boldsymbol{\eta}.$$

Example 1 (continued)

If we analyze the data assuming the GMM with $E(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta}$, determine

1. $E(\widehat{\tau_1 - \tau_2})$, and
2. $E(\hat{\sigma}^2)$.

Example 1 (continued)

From slide 6,

$$
\begin{aligned}
\widehat{E(\tau_1 - \tau_2)} &= \tau_1 - \tau_2 + \boldsymbol{a}'\boldsymbol{P}_X\boldsymbol{\eta} \\
&= \tau_1 - \tau_2 + \boldsymbol{a}'\boldsymbol{P}_X \begin{bmatrix} 1/2 \\ -1/2 \\ 1/2 \\ -1/2 \end{bmatrix} \begin{bmatrix} \alpha \end{bmatrix} \\
&= \tau_1 - \tau_2 + \boldsymbol{a}'\boldsymbol{0} \begin{bmatrix} \alpha \end{bmatrix} \\
&= \tau_1 - \tau_2.
\end{aligned}
$$

Thus, the LSE of $\tau_1 - \tau_2$ is unbiased in this case.

Example 1 (continued)

From slide 10,

$$\begin{aligned} E(\hat{\sigma}^2) &= \frac{\boldsymbol{\eta}'(\boldsymbol{I}-\boldsymbol{P_X})\boldsymbol{\eta}}{n-r} + \sigma^2 \\ &= \frac{\boldsymbol{\eta}'(\boldsymbol{I}\boldsymbol{\eta}-\boldsymbol{P_X}\boldsymbol{\eta})}{n-r} + \sigma^2 \\ &= \frac{\boldsymbol{\eta}'(\boldsymbol{\eta}-\boldsymbol{0})}{n-r} + \sigma^2 \\ &= \frac{\boldsymbol{\eta}'\boldsymbol{\eta}}{n-r} + \sigma^2 \\ &= \frac{\alpha^2}{4-2} + \sigma^2. \end{aligned}$$

Thus, $\hat{\sigma}^2$ is biased for σ^2 in this case.

Example 2

Once again consider an experiment with two experimental units (mice) for each of two treatments.

Suppose we assume the GMM holds with

$$E(\boldsymbol{y}) = E\left(\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{bmatrix}\right) = \boldsymbol{X}\boldsymbol{\beta} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \end{bmatrix} = \begin{bmatrix} \mu + \tau_1 \\ \mu + \tau_1 \\ \mu + \tau_2 \\ \mu + \tau_2 \end{bmatrix}.$$

Example 2 (continued)

Suppose the person who conducted the experiment neglected to mention that both experimental units in treatment group 1 were female and that both experimental units in treatment group 2 were male.

Example 2 (continued)

Then the true model may require

$$E(\boldsymbol{y}) = \begin{bmatrix} \mu+\tau_1+\alpha/2 \\ \mu+\tau_1+\alpha/2 \\ \mu+\tau_2-\alpha/2 \\ \mu+\tau_2-\alpha/2 \end{bmatrix} = \begin{bmatrix} \mu+\tau_1 \\ \mu+\tau_1 \\ \mu+\tau_2 \\ \mu+\tau_2 \end{bmatrix} + \begin{bmatrix} \alpha/2 \\ \alpha/2 \\ -\alpha/2 \\ -\alpha/2 \end{bmatrix}$$

$$= \boldsymbol{X\beta} + \begin{bmatrix} 1/2 \\ 1/2 \\ -1/2 \\ -1/2 \end{bmatrix} \begin{bmatrix} \alpha \end{bmatrix} = \boldsymbol{X\beta} + \boldsymbol{W\alpha} = \boldsymbol{X\beta} + \boldsymbol{\eta}.$$

Example 2 (continued)

If we analyze the data assuming the GMM with $E(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta}$, determine

1. $E(\widehat{\tau_1 - \tau_2})$, and
2. $E(\hat{\sigma}^2)$.

Example 2 (continued)

From slide 6,

$$
\begin{aligned}
E(\widehat{\tau_1 - \tau_2}) &= \tau_1 - \tau_2 + \boldsymbol{a}'\boldsymbol{P}_X\boldsymbol{\eta} \\
&= \tau_1 - \tau_2 + \boldsymbol{a}'\boldsymbol{P}_X \begin{bmatrix} 1/2 \\ 1/2 \\ -1/2 \\ -1/2 \end{bmatrix} \begin{bmatrix} \alpha \end{bmatrix} \\
&= \tau_1 - \tau_2 + \boldsymbol{a}' \begin{bmatrix} 1/2 \\ 1/2 \\ -1/2 \\ -1/2 \end{bmatrix} \begin{bmatrix} \alpha \end{bmatrix} = \tau_1 - \tau_2 + \alpha.
\end{aligned}
$$

Example 2 (continued)

Note that $\widehat{\tau_1 - \tau_2} = \bar{y}_{1\cdot} - \bar{y}_{2\cdot}$.

The previous slide shows that $\bar{y}_{1\cdot} - \bar{y}_{2\cdot}$ is not an unbiased estimator of the difference between treatment effects.

However, $\bar{y}_{1\cdot} - \bar{y}_{2\cdot}$ is an unbiased estimator of the difference between the means of the two treatment groups; i.e.,

$$E(\bar{y}_{1\cdot} - \bar{y}_{2\cdot}) = (\mu + \tau_1 + \alpha/2) - (\mu + \tau_2 - \alpha/2) = \tau_1 - \tau_2 + \alpha.$$

Part of the difference may be due to treatment, but part may be due to sex of the mice.

Example 2 (continued)

From slide 10,

$$\begin{aligned} E(\hat{\sigma}^2) &= \frac{\boldsymbol{\eta}'(\boldsymbol{I} - \boldsymbol{P}_X)\boldsymbol{\eta}}{n-r} + \sigma^2 \\ &= \frac{\boldsymbol{\eta}'(\boldsymbol{I}\boldsymbol{\eta} - \boldsymbol{P}_X\boldsymbol{\eta})}{n-r} + \sigma^2 \\ &= \frac{\boldsymbol{\eta}'(\boldsymbol{\eta} - \boldsymbol{\eta})}{n-r} + \sigma^2 \\ &= \sigma^2. \end{aligned}$$

Thus, $\hat{\sigma}^2$ is unbiased for σ^2 in this case.

Example 2 (continued)

Because $\boldsymbol{\eta} \in \mathcal{C}(\boldsymbol{X})$, both assumptions

$$E(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta} \quad \text{and} \quad E(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\eta}$$

are equivalent to $E(\boldsymbol{y}) \in \mathcal{C}(\boldsymbol{X})$.

Thus, even though we ignore sex of the mice, our model for the mean is correct.

The only mistake we would make is to assume that the difference in means for the treatment groups is due only to treatment rather than to a combination of treatment and sex.

Overfitting

Now suppose we consider the model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where

$$\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2] \quad \text{and} \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} \quad \ni \quad \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2.$$

Furthermore, suppose that (unknown to us) $\boldsymbol{X}_2\boldsymbol{\beta}_2 = \mathbf{0}$.

In this case, we say that we are overfitting.

Note that we are fitting a model that is more complicated than it needs to be.

To examine the impact of the overfitting, consider the case where $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2]$ is of full-column rank.

If we were to fit the simpler and correct model $\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}$, the LSE of $\boldsymbol{\beta}_1$ is $\tilde{\boldsymbol{\beta}}_1 = (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1\boldsymbol{y}$. Then

$$\begin{aligned} E(\tilde{\boldsymbol{\beta}}_1) &= (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'E(\boldsymbol{y}) \\ &= (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_1\boldsymbol{\beta}_1 \\ &= \boldsymbol{\beta}_1. \end{aligned}$$

$$
\begin{aligned}
\text{Var}(\tilde{\boldsymbol{\beta}}_1) &= (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\text{Var}(\boldsymbol{y})\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1} \\
&= \sigma^2(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1} \\
&= \sigma^2(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}.
\end{aligned}
$$

If we were to fit the full model

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$$

that is correct but more complicated than it needs to be, then the LSE of $\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix}$ is

$$\begin{bmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{bmatrix} = \left([\boldsymbol{X}_1, \boldsymbol{X}_2]'[\boldsymbol{X}_1, \boldsymbol{X}_2]\right)^{-1} [\boldsymbol{X}_1, \boldsymbol{X}_2]'\boldsymbol{y}$$

$$= \begin{bmatrix} \boldsymbol{X}_1'\boldsymbol{X}_1 & \boldsymbol{X}_1'\boldsymbol{X}_2 \\ \boldsymbol{X}_2'\boldsymbol{X}_1 & \boldsymbol{X}_2'\boldsymbol{X}_2 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{X}_1'\boldsymbol{y} \\ \boldsymbol{X}_2'\boldsymbol{y} \end{bmatrix}.$$

If $\boldsymbol{X}_1'\boldsymbol{X}_2 = \boldsymbol{0}$, then

$$\begin{bmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1'\boldsymbol{X}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{X}_2'\boldsymbol{X}_2 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{X}_1'\boldsymbol{y} \\ \boldsymbol{X}_2'\boldsymbol{y} \end{bmatrix}$$

$$= \begin{bmatrix} (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1\boldsymbol{y} \\ (\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2\boldsymbol{y} \end{bmatrix} = \begin{bmatrix} \tilde{\boldsymbol{\beta}}_1 \\ (\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2\boldsymbol{y} \end{bmatrix}.$$

Now suppose $\boldsymbol{X}_1'\boldsymbol{X}_2 \neq \mathbf{0}$.

$$
\begin{aligned}
E\left(\begin{bmatrix}\hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2\end{bmatrix}\right) &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E(\boldsymbol{y}) \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} \\
&= \boldsymbol{\beta} = \begin{bmatrix}\boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2\end{bmatrix}.
\end{aligned}
$$

Thus, $E(\hat{\boldsymbol{\beta}}_1) = \boldsymbol{\beta}_1$.

$$
\begin{aligned}
\text{Var}(\hat{\boldsymbol{\beta}}) &= \text{Var}\left(\begin{bmatrix}\hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2\end{bmatrix}\right) \\
&= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= \sigma^2 \begin{bmatrix}\boldsymbol{X}_1'\boldsymbol{X}_1 & \boldsymbol{X}_1'\boldsymbol{X}_2 \\ \boldsymbol{X}_2'\boldsymbol{X}_1 & \boldsymbol{X}_2'\boldsymbol{X}_2\end{bmatrix}^{-1}.
\end{aligned}
$$

By Exercise A.72,

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{E}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{E}^{-1} \\ -\boldsymbol{E}\boldsymbol{C}\boldsymbol{A}^{-1} & \boldsymbol{E}^{-1} \end{bmatrix},$$

where $\boldsymbol{E} = \boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}$.

Thus, $\mathrm{Var}(\hat{\boldsymbol{\beta}}_1)$ is σ^2 times

$$(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1} + (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2(\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}$$
$$= (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1} + (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2(\boldsymbol{X}_2'(\boldsymbol{I} - \boldsymbol{P}_{X_1})\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}.$$

Thus,

$$\text{Var}(\hat{\boldsymbol{\beta}}_1) - \text{Var}(\tilde{\boldsymbol{\beta}}_1) = \sigma^2(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2(\boldsymbol{X}_2'(\boldsymbol{I} - \boldsymbol{P}_{X_1})\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}.$$

In a homework problem, you will show that

$$\text{Var}(\hat{\boldsymbol{\beta}}_1) - \text{Var}(\tilde{\boldsymbol{\beta}}_1) \quad \text{is NND.}$$

Thus, one cost of overfitting is increased variability of estimators of regression coefficients.

How is estimation of σ^2 affected?

Let $r_1 = rank(\boldsymbol{X}_1)$ and $r_2 = rank(\boldsymbol{X}_2)$.

If we fit a simpler model $\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}$, then

$$\tilde{\sigma}^2 = \frac{\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{y}}{n - r_1} \quad \text{and}$$

$$E(\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{y}) = (n - r_1)\sigma^2$$

$$\Rightarrow E(\tilde{\sigma}^2) = \sigma^2.$$

If we overfit with the model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, then

$$\hat{\sigma}^2 = \frac{\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{y}}{n - r} \qquad \text{and}$$

$$E(\hat{\sigma}^2) = \sigma^2.$$

Thus, overfitting does not lead to biased estimation of σ^2.

However, as we will see later in the course, overfitting leads to a loss of degrees of freedom $(n - r < n - r_1)$, which can lead to a loss of power for testing hypotheses about $\boldsymbol{\beta}$.