Why $s^2 = \frac{(\hat{\varepsilon}'\hat{\varepsilon})}{(N-k)}$ is a good candidate for estimating σ^2

Guillem Riambau (Yale-NUS)

February 12, 2016

- Goal: find an unbiased for σ^2, where $\sigma^2 = Var(\varepsilon_i) \forall i$
- It is shown here that the best unbiased candidate is $s^2 = \frac{(\hat{\varepsilon}' \hat{\varepsilon})}{(N-k)}$
- Why do we care about σ^2?

- Goal: find an unbiased for σ^2, where $\sigma^2 = Var(\varepsilon_i)\forall i$
- It is shown here that the best unbiased candidate is $s^2 = \frac{(\hat{\varepsilon}'\hat{\varepsilon})}{(N-k)}$
- Why do we care about σ^2?
- $\mathbb{VAR}\hat{\beta} = \sigma^2(X'X)^{-1}$
- Need unbiased estimator to make good inference about $\hat{\beta}$

- Goal: find an unbiased for σ^2, where $\sigma^2 = Var(\varepsilon_i) \forall i$
- Since $\mathbb{E}(\varepsilon_i|X) = 0$, $\Rightarrow$ $Var(\varepsilon_i) = \mathbb{E}(\varepsilon_i^2|X)$
- Since $\hat{\varepsilon}$ is a good estimator of ε, the natural candidate: $\frac{1}{N}\sum_{i=1}^{N}\hat{\varepsilon}_i^2$
- In vector notation, $\frac{1}{N} \times \hat{\varepsilon}'\hat{\varepsilon}$

Goal: find $\frac{1}{N}\mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon})$

$$\hat{\varepsilon} = y - X\hat{\beta}$$

$$\begin{aligned}\hat{\varepsilon} &= y - X\hat{\beta} \\ &= y - X(X'X)^{-1}X'y\end{aligned}$$

$$
\begin{aligned}
\hat{\varepsilon} &= y - X\hat{\beta} \\
&= y - X(X'X)^{-1}X'y \\
&= I_n y - X(X'X)^{-1}X'y
\end{aligned}
$$

$$\begin{aligned} \hat{\varepsilon} &= y - X\hat{\beta} \\ &= y - X(X'X)^{-1}X'y \\ &= I_n y - X(X'X)^{-1}X'y \\ &= (I_n - X(X'X)^{-1}X')y \end{aligned}$$

$$
\begin{aligned}
\hat{\varepsilon} &= y - X\hat{\beta} \\
&= y - X(X'X)^{-1}X'y \\
&= I_n y - X(X'X)^{-1}X'y \\
&= (I_n - X(X'X)^{-1}X')y \\
&= M_x y
\end{aligned}
$$

$$
\begin{aligned}
\hat{\varepsilon} &= y - X\hat{\beta} \\
&= y - X(X'X)^{-1}X'y \\
&= I_n y - X(X'X)^{-1}X'y \\
&= (I_n - X(X'X)^{-1}X')y \\
&= M_x y
\end{aligned}
$$

since we define $M_x = (I_n - X(X'X)^{-1}X')$

$$\hat{\varepsilon} = M_x y$$

$$\hat{\varepsilon} = M_x y$$
$$= M_x(X\beta + \varepsilon)$$

$$
\begin{aligned}
\hat{\varepsilon} &= M_x y \\
&= M_x(X\beta + \varepsilon) \\
&= M_x X\beta + M_x\varepsilon
\end{aligned}
$$

$$
\begin{aligned}
\hat{\varepsilon} &= M_x y \\
&= M_x(X\beta + \varepsilon) \\
&= M_x X\beta + M_x\varepsilon \\
&= (I_n - X(X'X)^{-1}X')X\beta + M_x\varepsilon
\end{aligned}
$$

$$
\begin{aligned}
\hat{\varepsilon} &= M_x y \\
&= M_x(X\beta + \varepsilon) \\
&= M_x X\beta + M_x\varepsilon \\
&= (I_n - X(X'X)^{-1}X')X\beta + M_x\varepsilon \\
&= X\beta - X(X'X)^{-1}X'X\beta + M_x\varepsilon
\end{aligned}
$$

$$
\begin{aligned}
\hat{\varepsilon} &= M_x y \\
&= M_x(X\beta + \varepsilon) \\
&= M_x X\beta + M_x\varepsilon \\
&= (I_n - X(X'X)^{-1}X')X\beta + M_x\varepsilon \\
&= X\beta - X(X'X)^{-1}X'X\beta + M_x\varepsilon \\
&= X\beta - X\beta + M_x\varepsilon
\end{aligned}
$$

$$
\begin{aligned}
\hat{\varepsilon} &= M_x y \\
&= M_x(X\beta + \varepsilon) \\
&= M_x X\beta + M_x\varepsilon \\
&= (I_n - X(X'X)^{-1}X')X\beta + M_x\varepsilon \\
&= X\beta - X(X'X)^{-1}X'X\beta + M_x\varepsilon \\
&= X\beta - X\beta + M_x\varepsilon \\
&= M_x\varepsilon
\end{aligned}
$$

$$
\begin{aligned}
\hat{\varepsilon} &= M_x y \\
&= M_x(X\beta + \varepsilon) \\
&= M_x X\beta + M_x\varepsilon \\
&= (I_n - X(X'X)^{-1}X')X\beta + M_x\varepsilon \\
&= X\beta - X(X'X)^{-1}X'X\beta + M_x\varepsilon \\
&= X\beta - X\beta + M_x\varepsilon \\
&= M_x\varepsilon
\end{aligned}
$$

Recall: goal is to find $\mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon})$

$$\Rightarrow \hat{\varepsilon}'\hat{\varepsilon} = (M_x\varepsilon)' M_x\varepsilon$$

Recall: goal is to find $\mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon})$

$$\begin{aligned}\Rightarrow \hat{\varepsilon}'\hat{\varepsilon} &= (M_x\varepsilon)' M_x \varepsilon \\ &= \varepsilon' M_x' M_x \varepsilon\end{aligned}$$

Recall: goal is to find $\mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon})$

$$\Rightarrow \hat{\varepsilon}'\hat{\varepsilon} = (M_x\varepsilon)'M_x\varepsilon$$
$$= \varepsilon' M_x' M_x \varepsilon$$

(Problem set: M_x symmetric and idempotent)

Recall: goal is to find $\mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon})$

$$\Rightarrow \hat{\varepsilon}'\hat{\varepsilon} = (M_x\varepsilon)'M_x\varepsilon$$
$$= \varepsilon' M_x' M_x \varepsilon$$
(Problem set: M_x symmetric and idempotent)
$$= \varepsilon' M_x M_x \varepsilon$$

Recall: goal is to find $\mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon})$

$$
\begin{aligned}
\Rightarrow \hat{\varepsilon}'\hat{\varepsilon} &= (M_x\varepsilon)'M_x\varepsilon \\
&= \varepsilon' M_x' M_x \varepsilon \\
&\text{(Problem set: } M_x \text{ symmetric and idempotent)} \\
&= \varepsilon' M_x M_x \varepsilon \\
&= \varepsilon' M_x \varepsilon
\end{aligned}
$$

Recall: goal is to find $\mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon})$

$$\Rightarrow \hat{\varepsilon}'\hat{\varepsilon} = (M_x\varepsilon)'M_x\varepsilon$$
$$= \varepsilon' M_x' M_x \varepsilon$$
(Problem set: M_x symmetric and idempotent)
$$= \varepsilon' M_x M_x \varepsilon$$
$$= \varepsilon' M_x \varepsilon$$
(and since this is a scalar:$(1 \times N)(N \times N)(N \times 1)$)

Recall: goal is to find $\mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon})$

$$\Rightarrow \hat{\varepsilon}'\hat{\varepsilon} = (M_x\varepsilon)'M_x\varepsilon$$
$$= \varepsilon' M_x' M_x \varepsilon$$
(Problem set: M_x symmetric and idempotent)
$$= \varepsilon' M_x M_x \varepsilon$$
$$= \varepsilon' M_x \varepsilon$$
(and since this is a scalar:$(1 \times N)(N \times N)(N \times 1)$)
$$= tr(\varepsilon' M_x \varepsilon)$$

Recall: goal is to find $\mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon})$

$$
\begin{aligned}
\Rightarrow \hat{\varepsilon}'\hat{\varepsilon} &= (M_x\varepsilon)'M_x\varepsilon \\
&= \varepsilon' M_x' M_x \varepsilon \\
&\text{(Problem set: } \mathrm{M}_x \text{ symmetric and idempotent)} \\
&= \varepsilon' M_x M_x \varepsilon \\
&= \varepsilon' M_x \varepsilon \\
&\text{(and since this is a scalar:} (1 \times N)(N \times N)(N \times 1)) \\
&= tr(\varepsilon' M_x \varepsilon)
\end{aligned}
$$

$$\Rightarrow \hat{\varepsilon}'\hat{\varepsilon} = tr(\varepsilon' M_x \varepsilon)$$

$\Rightarrow \hat{\varepsilon}' \hat{\varepsilon} = tr(\varepsilon' M_x \varepsilon)$

(since, as long as size consistent, tr(ABC) = tr(BCA)=tr(CAB) etc.)

$\Rightarrow \hat{\varepsilon}'\hat{\varepsilon} = tr(\varepsilon' M_x \varepsilon)$

(since, as long as size consistent, tr(ABC) = tr(BCA)=tr(CAB) etc.)

$= tr(M_x \varepsilon\varepsilon')$

$\Rightarrow \hat{\varepsilon}'\hat{\varepsilon} = tr(\varepsilon' M_x \varepsilon)$

(since, as long as size consistent, tr(ABC) = tr(BCA)=tr(CAB) etc.)

$= tr(M_x \varepsilon \varepsilon')$

Recall: goal is to find $\mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon})$

$$\Rightarrow \mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon}) = \mathbb{E}(tr(M_x \varepsilon \varepsilon'))$$

$$\Rightarrow \hat{\varepsilon}'\hat{\varepsilon} = tr(\varepsilon' M_x \varepsilon)$$

(since, as long as size consistent, tr(ABC) = tr(BCA)=tr(CAB) etc.)

$$= tr(M_x \varepsilon\varepsilon')$$

Recall: goal is to find $\mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon})$

$$\Rightarrow \mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon}) = \mathbb{E}(tr(M_x \varepsilon\varepsilon'))$$
$$= tr(M_x \mathbb{E}(\varepsilon\varepsilon'))$$

$\Rightarrow \hat{\varepsilon}'\hat{\varepsilon} = tr(\varepsilon' M_x \varepsilon)$

(since, as long as size consistent, tr(ABC) = tr(BCA)=tr(CAB) etc.)

$= tr(M_x \varepsilon\varepsilon')$

Recall: goal is to find $\mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon})$

$$\Rightarrow \mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon}) = \mathbb{E}(tr(M_x \varepsilon\varepsilon'))$$

$$= tr(M_x \mathbb{E}(\varepsilon\varepsilon'))$$

(seen in class) $\mathbb{E}(\varepsilon\varepsilon') = \mathbb{VAR}(\varepsilon) = \sigma^2 I_n$

$$\Rightarrow \hat{\varepsilon}'\hat{\varepsilon} = tr(\varepsilon' M_x \varepsilon)$$

(since, as long as size consistent, tr(ABC) = tr(BCA)=tr(CAB) etc.)

$$= tr(M_x \varepsilon\varepsilon')$$

Recall: goal is to find $\mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon})$

$$\Rightarrow \mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon}) = \mathbb{E}(tr(M_x \varepsilon\varepsilon'))$$

$$= tr(M_x \mathbb{E}(\varepsilon\varepsilon'))$$

(seen in class)$\mathbb{E}(\varepsilon\varepsilon') = \mathbb{VAR}(\varepsilon) = \sigma^2 I_n$

$$= tr(M_x \sigma^2 I_n)$$

$\Rightarrow \hat{\varepsilon}'\hat{\varepsilon} = tr(\varepsilon' M_x \varepsilon)$

(since, as long as size consistent, tr(ABC) = tr(BCA)=tr(CAB) etc.)

$= tr(M_x \varepsilon\varepsilon')$

Recall: goal is to find $\mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon})$

$$\Rightarrow \mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon}) = \mathbb{E}(tr(M_x \varepsilon\varepsilon'))$$
$$= tr(M_x \mathbb{E}(\varepsilon\varepsilon'))$$
(seen in class) $\mathbb{E}(\varepsilon\varepsilon') = \mathbb{VAR}(\varepsilon) = \sigma^2 I_n$
$$= tr(M_x \sigma^2 I_n)$$
$$= tr(M_x I_n)\sigma^2$$

$$\Rightarrow \hat{\varepsilon}'\hat{\varepsilon} = tr(\varepsilon' M_x \varepsilon)$$

(since, as long as size consistent, tr(ABC) = tr(BCA)=tr(CAB) etc.)

$$= tr(M_x \varepsilon\varepsilon')$$

Recall: goal is to find $\mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon})$

$$\Rightarrow \mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon}) = \mathbb{E}(tr(M_x \varepsilon\varepsilon'))$$

$$= tr(M_x \mathbb{E}(\varepsilon\varepsilon'))$$

(seen in class) $\mathbb{E}(\varepsilon\varepsilon') = \mathbb{VAR}(\varepsilon) = \sigma^2 I_n$

$$= tr(M_x \sigma^2 I_n)$$

$$= tr(M_x I_n)\sigma^2$$

$$= tr(M_x)\sigma^2$$

$$\Rightarrow \hat{\varepsilon}'\hat{\varepsilon} = tr(\varepsilon' M_x \varepsilon)$$

(since, as long as size consistent, tr(ABC) = tr(BCA)=tr(CAB) etc.)

$$= tr(M_x \varepsilon\varepsilon')$$

Recall: goal is to find $\mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon})$

$$\Rightarrow \mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon}) = \mathbb{E}(tr(M_x \varepsilon\varepsilon'))$$

$$= tr(M_x \mathbb{E}(\varepsilon\varepsilon'))$$

(seen in class) $\mathbb{E}(\varepsilon\varepsilon') = \mathbb{VAR}(\varepsilon) = \sigma^2 I_n$

$$= tr(M_x \sigma^2 I_n)$$

$$= tr(M_x I_n)\sigma^2$$

$$= tr(M_x)\sigma^2$$

Now, what is $tr(M_x)$?

$$tr(M_x) = tr(I_n - X(X'X)^{-1}X')$$

Now, what is $tr(M_x)$?

$$tr(M_x) = tr(I_n - X(X'X)^{-1}X')$$
$$= tr(I_n) - tr(X(X'X)^{-1}X')$$

Now, what is $tr(M_x)$?

$tr(M_x) = tr(I_n - X(X'X)^{-1}X')$
$= tr(I_n) - tr(X(X'X)^{-1}X')$
(since, as long as size consistent, tr(ABC) = tr(BCA)=tr(CAB) etc.)

Now, what is $tr(M_x)$?

$$tr(M_x) = tr(I_n - X(X'X)^{-1}X')$$
$$= tr(I_n) - tr(X(X'X)^{-1}X')$$
(since, as long as size consistent, tr(ABC) = tr(BCA)=tr(CAB) etc.)
$$= tr(I_n) - tr(X'X(X'X)^{-1})$$

Now, what is $tr(M_x)$?

$$tr(M_x) = tr(I_n - X(X'X)^{-1}X')$$
$$= tr(I_n) - tr(X(X'X)^{-1}X')$$
(since, as long as size consistent, tr(ABC) = tr(BCA)=tr(CAB) etc.)
$$= tr(I_n) - tr(X'X(X'X)^{-1})$$
$(N \times N) \qquad (k \times k)$

Now, what is $tr(M_x)$?

$tr(M_x) = tr(I_n - X(X'X)^{-1}X')$
$= tr(I_n) - tr(X(X'X)^{-1}X')$
(since, as long as size consistent, tr(ABC) = tr(BCA)=tr(CAB) etc.)
$= tr(I_n) - tr(X'X(X'X)^{-1})$
$(N \times N) \qquad (k \times k)$
$= tr(I_n) - tr(I_k)$

Now, what is $tr(M_x)$?

$tr(M_x) = tr(I_n - X(X'X)^{-1}X')$
$= tr(I_n) - tr(X(X'X)^{-1}X')$
(since, as long as size consistent, tr(ABC) = tr(BCA)=tr(CAB) etc.)
$= tr(I_n) - tr(X'X(X'X)^{-1})$
$(N \times N) \qquad (k \times k)$
$= tr(I_n) - tr(I_k)$
$= N - k$

Now, what is $tr(M_x)$?

$$tr(M_x) = tr(I_n - X(X'X)^{-1}X')$$
$$= tr(I_n) - tr(X(X'X)^{-1}X')$$
(since, as long as size consistent, tr(ABC) = tr(BCA)=tr(CAB) etc.)
$$= tr(I_n) - tr(X'X(X'X)^{-1})$$
$$(N \times N) \qquad (k \times k)$$
$$= tr(I_n) - tr(I_k)$$
$$= N - k$$

Since $\mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon}) = tr(M_x)\sigma^2$
$\Rightarrow \mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon}) = (N-k)\sigma^2$

$$\mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon}) = (N-k)\sigma^2$$
$$\Rightarrow \frac{1}{N}\mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon}) = \frac{(N-k)}{N}\sigma^2$$

Our candidate is biased. It yields a variance that is too small. We'd be making wrong inference. What could be unbiased?

$$\mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon}) = (N-k)\sigma^2$$
$$\Rightarrow \frac{1}{N}\mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon}) = \frac{(N-k)}{N}\sigma^2$$

Our candidate is biased. It yields a variance that is too small. We'd be making wrong inference. What could be unbiased?

$$\text{Let } s^2 = \frac{(\hat{\varepsilon}'\hat{\varepsilon})}{(N-k)}$$
$$\Rightarrow \mathbb{E}(s^2) = \frac{1}{(N-k)}\mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon}) = \frac{(N-k)\sigma^2}{(N-k)} = \sigma^2$$