An example where OLS is biased and the conventional variance estimator is wrong

Let (X_i, Y_i) be IID for $i = 1, \ldots, n$ with $\mathrm{E}\{X_i\} = \mathrm{E}\{Y_i\} = 0$. Let b minimize $\mathrm{E}\{(Y_i - X_i)^2\}$, so

$$b = \mathrm{E}\{X_i Y_i\} \big/ \mathrm{E}\{X_i^2\} \tag{1}$$

This is a parameter to be estimated, and the OLS estimator is

$$\hat{b}_n = \sum_{i=1}^{n} X_i Y_i \Big/ \sum_{i=1}^{n} X_i^2 \tag{2}$$

This problem does not fit into the usual OLS framework. To see why, let $\epsilon_i = Y_i - bX_i$. The ϵ_i are IID with mean 0 and $\epsilon_i \perp X_i$, which is all to the good. But ϵ_i and X_i will generally be dependent, so $\mathrm{E}\{\epsilon_i | X_i\}$ will generally not be 0, and $\mathrm{var}\{\epsilon_i | X_i\}$ will generally not be constant. Hence, $\hat{b}_n$ may be biased, and the conventional formula for $\widehat{\mathrm{var}}\,\hat{b}_n$ may be seriously in error.

In the following special case, calculations can be done quite explicitly. Suppose $X_i \sim \mathrm{N}(0, 1)$ and $Y_i = X_i^3$. Then

$$\hat{b}_n = \frac{1}{n} \sum_{i=1}^{n} X_i^4 \Big/ \frac{1}{n} \sum_{i=1}^{n} X_i^2 \tag{3}$$

$$\mathrm{E}\{\hat{b}_n | X_1, \ldots, X_n\} = \hat{b}_n \tag{4}$$

$$\mathrm{var}\{\hat{b}_n | X_1, \ldots, X_n\} = 0 \tag{5}$$

For the variance estimator, define the residuals as

$$e_i = Y_i - \hat{b}_n X_i \tag{6}$$

Now $\sum_{i=1}^{n} e_i X_i = 0$, but $\sum_{i=1}^{n} e_i \neq 0$ because there is no intercept in the regression. (None is needed, since the random variables are known to have expectation 0.) The usual estimate for the variance of $\hat{b}_n$ given $X_1, \ldots, X_n$ is

$$\widehat{\mathrm{var}}\,\hat{b}_n = \frac{1}{n-1} \sum_{i=1}^{n} e_i^2 \Big/ \sum_{i=1}^{n} X_i^2 \tag{7}$$

To determine the asymptotic behavior, we need some moments. The odd moments of X_i vanish by symmetry. For the even moments, integration by parts shows that

$$\mathrm{E}\{X_i^2\} = 1, \quad \mathrm{E}\{X_i^4\} = 3, \quad \mathrm{E}\{X_i^6\} = 5 \times 3 = 15, \quad \mathrm{E}\{X_i^8\} = 7 \times 15 = 105 \tag{8}$$

The regression parameter b in (1) is now easily computed, as

$$b = \mathrm{E}\{X_i^4\} \big/ \mathrm{E}\{X_i^2\} = 3 \tag{9}$$

By (4) and the strong law of large numbers, $\hat{b}_n$ is consistent:

$$\hat{b}_n \to 3 \text{ a.e. as } n \to \infty \tag{10}$$

We turn now to the variance estimator in (7). The denominator is asymptotic to n, since $\frac{1}{n}\sum_{i=1}^n X_i^2 \to 1$. For the numerator,

$$\frac{1}{n}\sum_{i=1}^n e_i^2 = \frac{1}{n}\sum_{i=1}^n Y_i^2 - \hat{b}_n^2 \frac{1}{n}\sum_{i=1}^n X_i^2 \to 15 - 3^2 = 6 \tag{11}$$

Since $n/(n-1) \to 1$,

$$\widehat{\text{var}}\, \hat{b}_n \approx 6/n \ \text{ a.e.}, \tag{12}$$

in the sense that the ratio approaches 1.

How far is $\hat{b}_n$ from b? The answer is given by the following proposition.

Proposition. Suppose X_i are IID N(0,1) and $Y_i = X_i^3$. The OLS estimator $\hat{b}_n$ is defined by (3). Then $\hat{b}_n = 3 + U_n + V_n + W_n$, where $\text{E}\{U_n\} = 0$, U_n is asymptotically $\text{N}(0, 42/n)$, $\text{E}\{V_n\} = -6/n$, and W_n is of order $1/n^{3/2}$.

Here, U_n, V_n, and W_n will be computed as explicit functions of $X_1, \ldots, X_n$. Given $X_1, \ldots, X_n$, the conditional bias in the OLS estimator is essentially U_n, which varies from one set of X's to another, (almost) following the normal distribution. The mean is 0, so the unconditional bias is about $-6/n$. The asymptotic distribution of U_n can be stated more rigorously as follows: $\sqrt{42/n}U_n \to$ N(0, 1) in law. The assertion about W_n is usually stated as follows: $W_n = O_P(1/n^{3/2})$, meaning that for any small positive δ, there is a large finite L with $P\{|W_n| < L/n^{3/2}\} > 1-\delta$. The argument will only be sketched, i.e., remainder terms will be dropped rather than estimated carefully; $\doteq$ means "approximately equal."

Define ξ_n and ζ_n as follows:

$$\xi_n = \frac{1}{n}\sum_{i=1}^n (X_i^2 - 1), \quad \zeta_n = \frac{1}{n}\sum_{i=1}^n (X_i^4 - 3). \tag{13}$$

Now

$$\hat{b}_n = \frac{3 + \zeta_n}{1 + \xi_n} \doteq (3 + \zeta_n)(1 - \xi_n + \xi_n^2) \doteq 3 + (\zeta_n - 3\xi_n) + 3\xi_n^2 - \xi_n\zeta_n. \tag{14}$$

The approximation in (14) is the "delta method"; cubic terms are ignored. In the proposition,

$$U_n = \zeta_n - 3\xi_n = \frac{1}{n}\sum_{i=1}^n (X_i^4 - 3X_i^2), \quad V_n = 3\xi_n^2 - \xi_n\zeta_n \tag{15}$$

The remainder term W_n is

$$\hat{b}_n - 3 - (\zeta_n - 3\xi_n) - 3\xi_n^2 + \xi_n\zeta_n.$$

The asymptotic normality of U_n follows from the central limit theorem; and $\mathrm{E}(U_n) = 0$ by (8). Moreover,

$$n \operatorname{var} U_n = \operatorname{var}\{X_i^4 - 3X_i^2\} = \mathrm{E}\{(X_i^4 - 3X_i^2)^2\} = E\{X_i^8 - 6X_i^6 + 9X_i^4\} = 105 - 90 + 27 = 42.$$

This completes the discussion of U_n. For V_n, cross-product terms in

$$\xi_n^2 = \frac{1}{n^2} \sum_{j,k=1}^{n} (X_j^2 - 1)(X_k^2 - 1)$$

or

$$\xi_n \zeta_n = \frac{1}{n^2} \sum_{j,k=1}^{n} (X_j^2 - 1)(X_k^4 - 3)$$

have mean 0, so

$$3\mathrm{E}\{\xi_n^2\} = \frac{3}{n}\mathrm{E}\{(X_i^2 - 1)^2\} = \frac{6}{n}$$

$$\mathrm{E}\{\xi_n \zeta_n\} = \frac{1}{n}\mathrm{E}\{(X_i^2 - 1)(X_i^4 - 3)\} = \frac{1}{n}\mathrm{E}\{X_i^6 - 3\} = \frac{12}{n}$$

and $\mathrm{E}\{V_n\} = -6/n$. This completes the discussion of V_n and the outline of the proof.

Remark. $\operatorname{var} V_n \approx 120/n^2$. Thus, V_n is around $-6/n$, give or take $\sqrt{120}/n$ or so. On the other hand, U_n is around 0, give or take $\sqrt{42}/\sqrt{n}$ or so. In telegraphic form, the argument for $\operatorname{var} V_n$ is as follows. We have shown that $\mathrm{E}\{V_n\} = -6/n$, and will show that $\mathrm{E}\{V_n^2\} = (156/n^2) + O(1/n^3)$. Indeed,

$$V_n = -\xi_n(\zeta_n - 3\xi_n) = \frac{1}{n^2} \sum_{j=1}^{n} (X_j^2 - 1) \sum_{k=1}^{n} (X_k^4 - 3X_k^2)$$

so that

$$\begin{aligned} V_n^2 &= \frac{1}{n^4} \sum_{j_1=1}^{n} (X_{j_1}^2 - 1) \sum_{j_2=1}^{n} (X_{j_2}^2 - 1) \sum_{k_1=1}^{n} (X_{k_1}^4 - 3X_{k_1}^2) \sum_{k_2=1}^{n} (X_{k_2}^4 - 3X_{k_2}^2) \\ &= \frac{1}{n^4} \sum_{j_1, j_2, k_1, k_2=1}^{n} (X_{j_1}^2 - 1)(X_{j_2}^2 - 1)(X_{k_1}^4 - 3X_{k_1}^2)(X_{k_2}^4 - 3X_{k_2}^2) \end{aligned}$$

Most of the terms have expectation zero, i.e., terms with 4 different indices, terms with 2 indices the same and 2 different, terms with 3 indices the same and 1 different. There are n terms with $j_1 = j_2 = k_1 = k_2$, which contribute $O(1/n^3)$ to $\mathrm{E}\{V_n^2\}$. There are $n(n-1)$ terms with $j_1 = j_2 \neq k_1 = k_2$; each is $\mathrm{E}\{(X_i^2 - 1)^2\}\mathrm{E}\{(X_i^4 - 3X_i^2)^2\} = 84$. Likewise, there are $n(n-1)$ terms with $j_1 = k_1 \neq j_2 = k_2$; each is $\left[\mathrm{E}\{X_i^6 - 3X_i^4\}\right]^2 = 36$. Finally, there are $n(n-1)$ terms with $j_1 = k_2 \neq j_2 = k_1$; each is 36, as before. Thus,

$$\mathrm{E}\{V_n^2\} = \frac{84 + 36 + 36}{n^2} + O\Big(\frac{1}{n^3}\Big),$$

as required.