EXERCISE ON WAYS TO LOOK AT REGRESSION

Suppose we have data drawn from an equal-probability weighted mixture of two bivariate normal distributions,

$$N\left(\begin{bmatrix}1\\1\end{bmatrix}, I\right) \text{ and } N\left(\begin{bmatrix}2\\4\end{bmatrix}, I\right),$$

where the variables are X and Y (so $E[X] = 1.5$ and $E[Y] = 2.5$). Let $\beta^* = (a, b)'$ be the squared-error-minimizing coefficients in the linear regression equation $Y = a + bX + u$.

(1) Calculate β^*.
(2) Calculate the function $f(X) = E[Y \mid X]$.
(3) Calculate $\sigma^2(X) = \text{Var}(u \mid X)$.
(4) Calculate $\text{Var}(Y - f(X) \mid X)$ at x values -1, 1, 1.5, 2, and 4. Compare to $\sigma^2(X)$ at those values of x and to the unconditional $\text{Var}(u)$
(5) Plot a scatter of 100 random data points drawn from this distribution, and on the same plot show the $a + bX$ line and the $f(x)$ line.
(6) (optional, as it's lots of algebra) Calculate the asymptotic covariance matrix of the least squares estimate of β^*.
(7) The GLS formula might suggest using weighted least squares here, with the weights on the observations one over the square root of the $\text{Var}(u \mid X)$ function you calculated in 3. Here this estimator does not improve on OLS, though. Why?

Date: February 8, 2018.