# Generalized Least Squares

*Practice Problems*

*These questions will be discussed during office hours on Tuesday.*

1. **Machine Speed**
   The number of defective items produced by a machine $Y$ is known to be linearly related to the speed setting of the machine $X$. The data can be found in the file `machine.txt` ($Y$ is in column 1, $X$ is in column 2).

   (a) Fit a linear regression function by OLS, obtain the residuals, and plot the residuals against $X$. What does the residual plot suggest?

   (b) Plot the squared residuals against $X$. What does this plot suggest about the relation between the standard deviation of the error term and $X$?

   (c) Estimate the variance function by regressing the squared residuals against $X$, and then calculate the estimated weight for each case.

   (d) Using the estimated weights, obtain the weighted least squares estimates of $\beta_0$ and $\beta_1$. Are these estimates similar to the ones obtained by the OLS in part (a)?

   (e) Compare the estimated standard deviations of the weighted least squares estimates in part (d) with those for the OLS in part (a). What do you find?

   (f) Iterate the steps in parts (c) and (d) one more time. Is there a substantial change in the estimated regression coefficients? If so, what should you do?

2. Derive the weighted least squares normal equations for fitting a simple linear regression function when $\sigma_i = kx_i$, where $k$ is a proportionality constant.

3. Observations on $Y$ are to be taken when $X$=10, 20, 30, 40, and 50, respectively. The true regression function is $\mathbb{E}(Y) = 20 + 10X$. The error terms are independent and normally distributed, with $\mathbb{E}(\epsilon_i) = 0$ and $Var(\epsilon_i) = 0.8X_i$.

   (a) Generate a random $Y$ observation for each $X$ level and calculate both the ordinary and weighted least squares estimates of the regression coefficient $\beta_1$ in the simple linear regression function.

   (b) Repeat part (a) 200 times generating new random numbers each time.

   (c) Calculate the mean and variance of the 200 OLS estimates $\beta_1$ and do the same for the 200 weighted least squares estimates.

   (d) Do both the OLS and WLS estimators appear to be unbiased? Explain. Which estimator appears to be more precise here? Comment.