**Question 10.2**

1. I expect the following signs for the coefficients:
   1. Positive: WAGE, EDUC
   2. Negative: AGE, KIDSL6, KIDS618, NWIFEINC
2. This equation cannot be consistently estimated because of the endogenous relationship between HOURS and WAGE. It’s almost the same as the demand and supply curve. If there is a need for more workers(hours), the employers will tend to increase the hourly wages. Once that happens, more women will begin to join the workforce. However, at some point, the supply of workers will be more than the demand and hence the employers will most likely reduce the hourly wages. In this way, WAGE and HOURS are endogenously related and this equation cannot be consistently estimated using the OLS estimators.
3. Although I cannot assess using the data, intuitively, I Conclude the following about EXPER and EXPER2:
   1. EXPER itself should not have any effect on HOURS. If I were to plot this, I imagine seeing a random plot.
   2. I cannot say whether EXPER is related to the residuals or not.
   3. EXPER should be highly correlated with WAGE.

Barring the second one, EXPER satisfies, at least intuitively, the requirements for a strong IV.

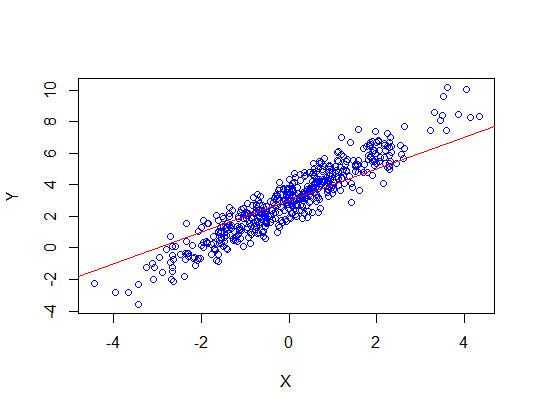
1. Using the IVs, the supply equation can be identified as follows:

Where is the fitted value based from the 2SLS method.

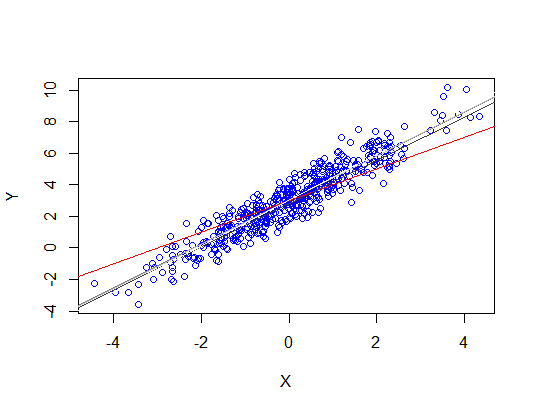
1. I will take the following steps to obtain the 2SLS estimates:
   1. Check if there are sufficient number of IVs – 2 in this case for 1 endogenous variable. I will go ahead to the next step.
   2. Regress WAGE on the remaining independent variables in the model (EDUC….NWIFEINC) and the IVs (EXPER and EXPER2).
   3. Remove the IVs and perform an F-Test to check if the F-value is greater than 10. If so, I will go to the next step. If not, I will stop here.
   4. Obtain fitted values
   5. Use the fitted values from the step above in place of the original values. The estimate thus obtained will be the 2SLS estimates.

**Question 10.6**

1. The calculated correlation is **0.6363961**.
2. The correlation is **0.65136**. This is pretty close to what I estimated in part a.
3. Plotted as follows:

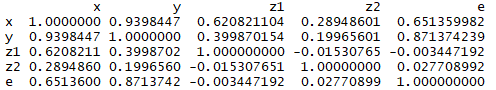


1. Having plotted each of the lines on the graph, I can say that they are not even close to approaching the true parameters.



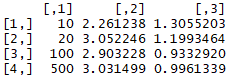
This tells me that for whatever issue is causing this, increasing the sample size is not a solution. It would appear that the error term is positively correlated with y, causing endogeneity in the model.

1. The sample correlations are as follows:

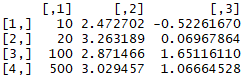


z1 and z2 would make good indicator variables, because both of them have a significant correlation with x and a minor correlation with e. Although their correlation with y is a little concerning, but not that much. z1 would make a better IV than z2 because it has a higher correlation and is likely more strongly related to x.

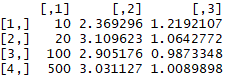
1. This time, they are definitely getting close to the true estimated.



1. Using z1 alone offers better results than using z2 alone. I would say this is so because z1 is a strongly related to x than z2 is – something I had mentioned in part e.



1. Looking purely at the coefficient of x, I would say that using z1 made more sense. 1.008 is farther from 1 than 0.996 is.



Question