1.0 GHJ 13.6

After running a bunch of regressions of one independent variable against the others, we notice some high adjusted R squared terms from D, Lnt1, and Lnt2. These variables all measure time. When we drop D, we note that Lnt1 and Lns both become significant. This is interesting, suggesting that the date the permit was issued tells us a lot about how large the plant is and how long it takes to get a permit. But when you drop D, the coefficient on pt becomes insignificant and drops markedly, indicating that partial turnkey plants also tended to be built at certain times. We observe quite a bit of instability here, and it’s not really easy to pull out what is happening - why should removing date of permit issue from the model alter the estimated effect of turnkey plants?

2.0 GHJ 13.8

We run the regression with all the variables, and then we drop either year of GNP deflator. Whichever we drop, the other becomes significant. But when they are both in there together, neither is significant. Looking at the simple correlations, we notice correlations of 0.99! This is a sure sign we are going to have problems. When we drop the last two observations, and run the complete model, we notice that two of our variables suddenly become significant, and even worse, several signs flip! All of these are signs of collinearity.

3.0 GHJ 15.1

This one is rather straightforward. There are two groups in the sample, and the groups have different variances, but the variance of the two groups differs. GLS is essentially about equalizing the variance of the errors in this instance. The variance of the errors is described by $W$ and is already embodied in the variables, so to equalize them we use $W^{-1}$. Using the inverse of the error covariance matrix basically undoes the problem - it weights the observations with the smaller error variance more, because these observations effectively carry more information.

\[
W^{-1} = \begin{bmatrix}
\frac{1}{\sigma_1^2}I & 0 \\
0 & \frac{1}{\sigma_2^2}I
\end{bmatrix}
\]

which makes it easy to find the matrix $P = \begin{bmatrix}
\frac{1}{\sigma_1}I & 0 \\
0 & \frac{1}{\sigma_2}I
\end{bmatrix}$.
This is because to take the ‘square root’ of a matrix with numbers on the diagonal and zeroes off, we just take the square root of each piece of the diagonal. When the matrix is not diagonal, it’s much harder.

\[ \hat{B} = \left[ \begin{bmatrix} \frac{1}{\sigma_1^2}X_1' \frac{1}{\sigma_2^2}X_2' \end{bmatrix} X_1 \right]^{-1} = \left[ \frac{X_1'X_1}{\sigma_1^2} + \frac{X_2'X_2}{\sigma_2^2} \right]^{-1} \]

b)

Working through the variance is pretty much identical.

### 4.0 GHJ 15.2

We are going to test the hypothesis that the variance for the first part of the sample is less than the variance from the second part of the sample. Notice that this is implicitly a one sided test. We are assuming that the variance from the first half is not larger than that from the second half, and so if we reject our null hypothesis that it is different, that’s because the second half variance is larger. Anyway, onwards...

From STATA, we get the estimates:

\[ \sigma_1^2 = 3.8134 \quad \text{and} \quad \sigma_2^2 = 9.8788 \]

We create our Goldfield-Quandt statistic:

\[ GQ = \frac{\sigma_2^2}{\sigma_1^2} = \frac{9.8788}{3.8134} = 2.59 \]

Checking on the F table, we find that this is just a hair larger than the F(13,13) 5% critical value, which is about 2.56, so we reject the hypothesis of equality.

I should point out that it is really bad form to do a GQ test with one subset of the observations, then redo it with another subset if the first one fails. It might have been worthwhile to even cut out another 10 variables in the middle of the sequence, and look only at the top or bottom quarter, but the more you cut out, the lower your degrees of freedom on the F table. What’s optimal? Hard to say. But you should pick one way, then do it, rather than keep trying different combinations of subsets till you get the answer you want.