**MODEL 1**

Table

Description automatically generated

Table

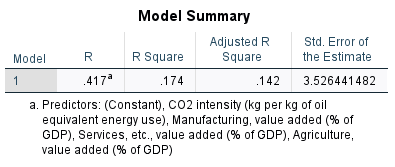
Description automatically generatedTable

Description automatically generated

**Are VIF’S greater than 3? (This is an indicator of multicollinearity). Which variables? How many variables are statistically significant at the 10% level?**

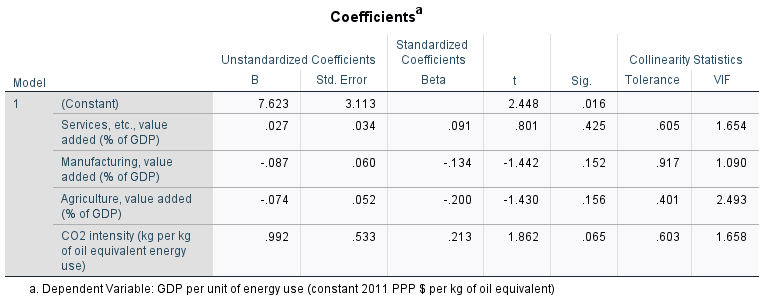
Two of the five explanatory variables have a VIF that is greater than 3. The two explanatory variables are CO2 Intensity and Fossil fuel energy consumption. None of the explanatory variables are statistically significant at the 10% level.

**MODEL 2**



Table

Description automatically generated



**Has the coefficient on CO2 intensity become more significant(significant at the 10% level)?**

Yes, it has become more statistically significant. Removing multicollinearity has repaired the standard errors of the estimates and the significant value is more appropriate.

**MODEL 3**

Table

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated

**Compare “Model 2” above and “Model 3”. How has adding Energy Use affected the value of the coefficients on Services, Manufacturing, Agriculture, and CO2 Intensity? Do you suspect Energy Use is an omitted variable?**

Some variables have become more statistically significant, so it leads me to believe that Energy Use is an omitted variable.

**MODEL 4**

Table

Description automatically generatedTable

Description automatically generatedTable

Description automatically generated

**Do any of the figures appear to have a vague fan shape? If so, which?** Yes, here are the scatterplots with heteroscedasticity.Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

**Briefly describe the problem that heteroskedasticity causes to a regression and how you can correct it.**

Y observations tend to be closer to the population regression line in one part of the graph over another. In other words, the model best fits **some** x observations since their y observations are closer to the regression line than others. To fix this, you take the log of the explanatory variable.