# Multicollinearity / Multikollinearität

Multikollinearität bezeichnet den linearen Zusammenhang, den zwei oder mehrere unabhängige Variablen eines Linearen Modells zueinander haben. Die Stärke dieses Zusammenhangs lässt sich mittels der Toleranz bzw. dessen Kehrwert: dem VIF-Wert ausdrücken. Hohe Multikollinearität ist problematisch, da sie zu ungenauen Schätzungen der Regressoren führt. Ferner ist der eigentliche Einfluss einer unabhängigen Variablen nur schwer oder gar nicht zu bestimmen.

In der Literatur finden sich unterschiedliche Grenzwerte für die Multikollinearität. Für meine Auswertung gehe ich von einem VIF-Grenzwert von 10 aus.

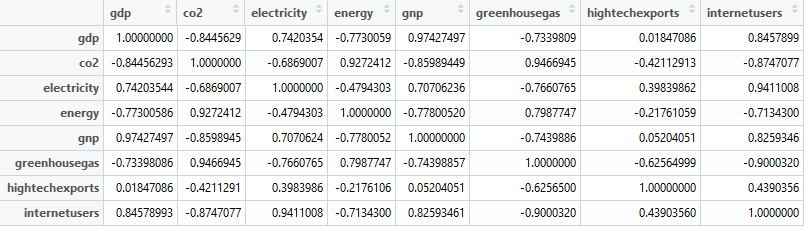


Abbildung Korrelationskoeffizienten der Variablen des Datensets zueinander

## Linear Model 1

|  |  |  |  |
| --- | --- | --- | --- |
|  | **co2** | **energy** | **greenhousegas** |
| **VIF** | 35.40566 | 10.09841 | 14.32909 |

Die Auswertung der VIF-Werte zeigt deutlich, dass alle unabhängigen Variablen den Grenzwert von 10 überschritten haben. Zwischen den Variablen bestehen starke lineare Abhängigkeiten, die die Aussagekraft des Modells stark beeinträchtigen. Abbildung 1 bestätigt die Multikollinearität. CO2 hat beispielsweise zu energy und greenhousegas jeweils eine Korrelationskoeffizienten von 0.92 und 0.94.

## Linear Model 2

|  |  |  |
| --- | --- | --- |
|  | **gnp** | **electricity** |
| **VIF** | 1.538878 | 1.538878 |

Trotz eines bestehenden linearen Zusammenhangs zwischen gnp und electricity von immerhin 0.7, reicht dieser nicht aus, um unseren VIF Grenzwert zu verletzen.

## Linear Model 3

|  |  |  |
| --- | --- | --- |
|  | **hightechexports** | **internetusers** |
| **VIF** | 1.32067 | 1.32067 |

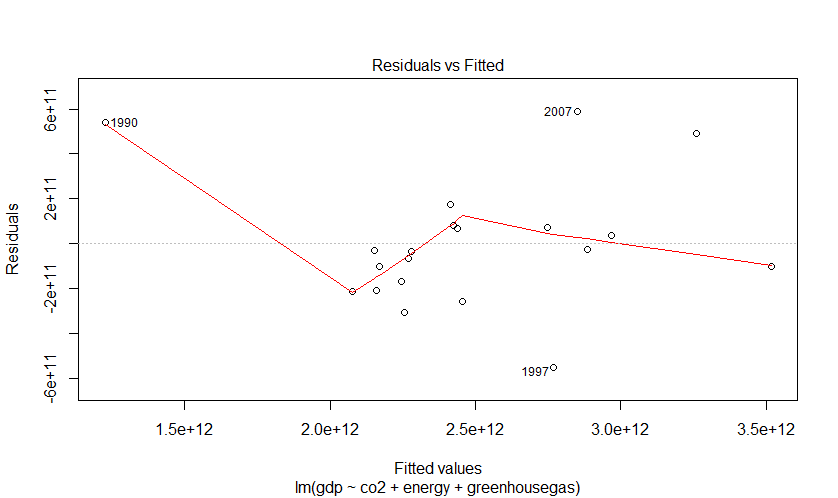
Zwischen hightechexports und internetusers gibt es keinen nennenswerten linearen Zusammenhang weswegen auch die VIF-Werte entsprechend gering ausfallen.

# Heteroscedasticity / Heteroskedastizität

Heteroskedastizität untersucht die Streuung der Residuen für den gesamten Vorhersagebereich. Wenn die Varianz der Residuen innerhalb des Vorhersagebereichs stark unterschiedlich ausfällt, liegt Homoskedastizität vor.

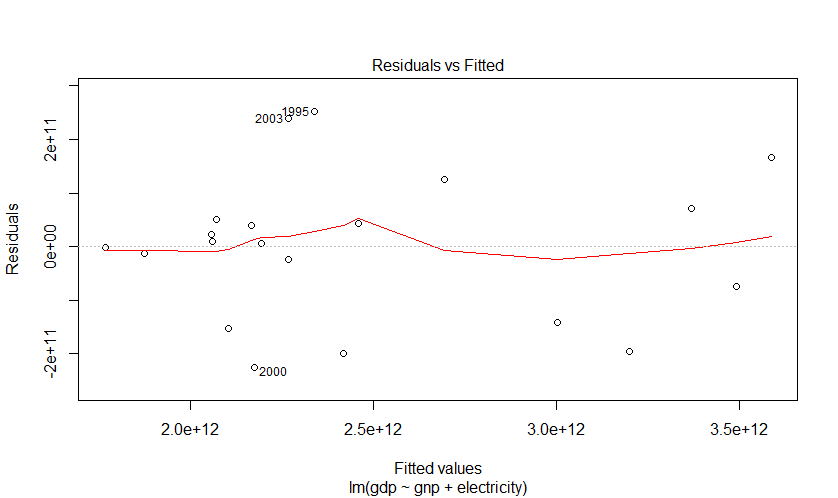
Homoskedastizität lässt sich sehr schön anhand eines Residuals-FittedValues-Graphen überprüfen.

## Linear Model 1



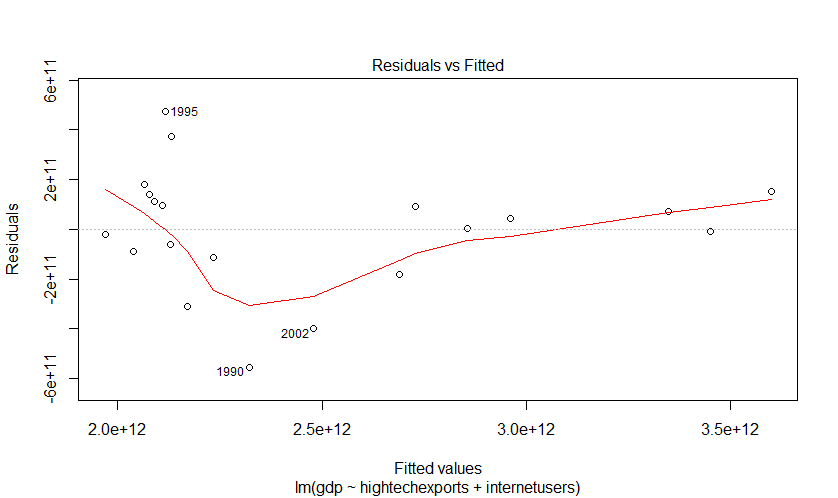
Die Streuung der Residuals ist relativ konstant über den gesamten Wertebereich hinweg. 🡪 Geringe Homoskedastizität

## Linear Model 2



Die Streuung der Residuals nimmt mit zunehmender Größe der Fitted Values zu 🡪 Hohe Homoskedastizität

## Linear Model 3



Die Streuung der Residuals ist relativ konstant über den gesamten Wertebereich hinweg. 🡪 Geringe Homoskedastizität