

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.

	gdp	co2	electricity	energy	gnp	greenhousegas	hightechexports	internetusers
gdp	1.00000000	-0.8445629	0.7420354	-0.7730059	0.97427497	-0.7339809	0.01847086	0.8457899
co2	-0.84456293	1.0000000	-0.6869007	0.9272412	-0.85989449	0.9466945	-0.42112913	-0.8747077
electricity	0.74203544	-0.6869007	1.0000000	-0.4794303	0.70706236	-0.7660765	0.39839862	0.9411008
energy	-0.77300586	0.9272412	-0.4794303	1.0000000	-0.77800520	0.7987747	-0.21761059	-0.7134300
gnp	0.97427497	-0.8598945	0.7070624	-0.7780052	1.0000000	-0.7439886	0.05204051	0.8259346
greenhousegas	-0.73398086	0.9466945	-0.7660765	0.7987747	-0.74398857	1.0000000	-0.62564999	-0.9000320
hightechexports	0.01847086	-0.4211291	0.3983986	-0.2176106	0.05204051	-0.6256500	1.0000000	0.4390356
internetusers	0.84578993	-0.8747077	0.9411008	-0.7134300	0.82593461	-0.9000320	0.43903560	1.0000000

Abbildung 1 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.

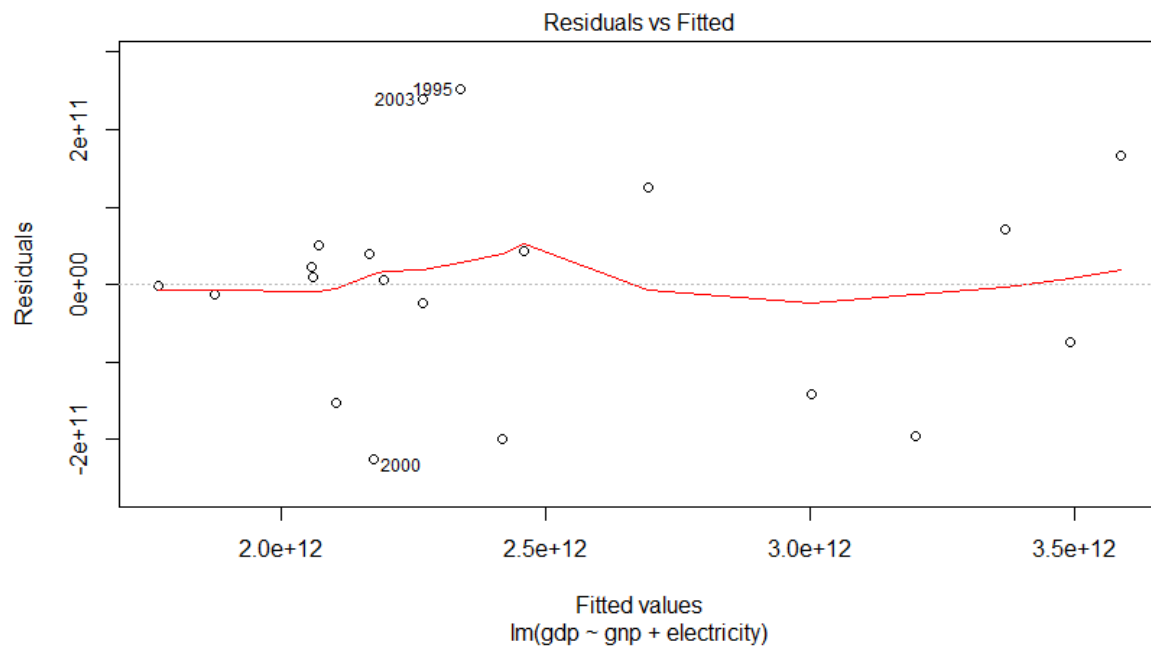
Homoskedastizität lässt sich sehr schön anhand eines Residuals-FittedValues-Graphen überprüfen.

Linear Model 1



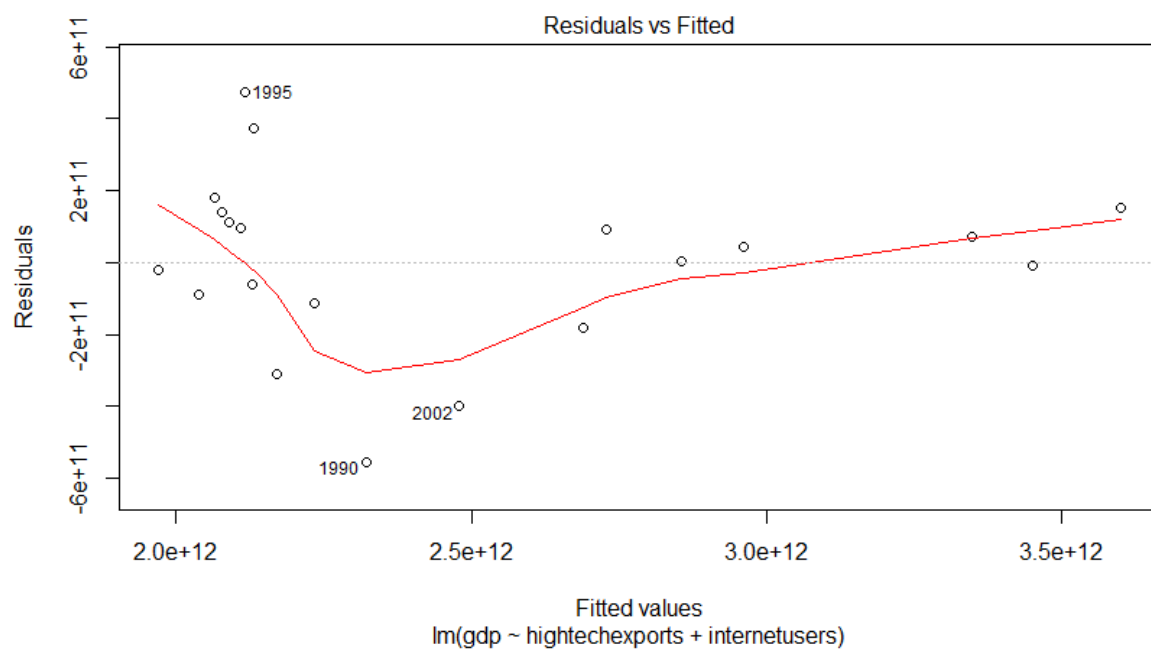
Die Streuung der Residuals ist relativ konstant über den gesamten Wertebereich hinweg. → Geringe Homoskedastizität

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