

4. Multikollinearität

4.1 Begriff der Multikollinearität

- Naturwissenschaften:

Experimentator kann die Werte der unabhängigen Variablen festlegen
(Unabhängigkeit, keine Korrelation)

- Wirtschaftswissenschaften:

Werte der unabhängigen Variablen werden beobachtet (Abhängigkeit, Korrelation)

Problematik in Wirtschaftswissenschaften:

Dadurch, dass die unabhängigen Variablen korreliert sind, wird die Messung ihres isolierten Einflusses auf die abhängige Variable erschwert. Eine Variable, die mit einer anderen Variablen korreliert ist, misst zum Teil den Einfluss der anderen Variablen mit. Wenn die unabhängigen Variablen in einem Regressionsmodell miteinander korreliert sind, spricht man von **Multikollinearität**.

Bei Multikollinearität ist der Gesamteinfluss der unabhängigen Variablen auf die abhängige Variable exakt erfassbar. Die Zurechenbarkeit des Gesamteinflusses der unabhängigen Variablen ist dagegen aufgrund ihrer wechselseitigen Abhängigkeit nicht eindeutig möglich.

Grenzfall: Exakte Multikollinearität

Eine exakte Multikollinearität, d.h. eine lineare Abhängigkeit zwischen den Regressoren würde dazu führen, dass der Rang der Beobachtungsmatrix \mathbf{X} kleiner als k wird, so dass die Produktmatrix $\mathbf{X}'\mathbf{X}$ nicht invertierbar ist. In diesem Extremfall könnte der OLS-Schätzer $\hat{\beta}$ nicht berechnet werden.

Betrachten wir hierzu das ökonometrische Modell

$$(4.1) \quad y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$$

mit der Scheinvariablen $x_{1t}=1$ und den beiden Einflussgrößen x_2 und x_3 . Wenn nun die beiden Regressoren x_2 und x_3 exakt linear miteinander verknüpft sind,

$$(4.2) \quad x_{3t} = \gamma x_{2t}$$

dann sind die Koeffizienten β_1 , β_2 und β_3 nicht mehr eindeutig bestimmbar, da die Inverse der Produktionsmatrix $\mathbf{X}'\mathbf{X}$ nicht existiert. Da jedoch ein perfekter Zusammenhang zwischen x_2 und x_3 vorliegt, d.h. der Korrelationskoeffizient r_{23} gleich 1 ist, reicht es aus, eine der beiden exogenen Variablen zur Erklärung von y heranzuziehen. Substituiert man (4.2) in (4.1), so erhält man die neue Regressionsgleichung

$$(4.3) \quad y_t = \beta_1 x_{1t} + (\beta_2 + \gamma\beta_3)x_{2t} + u_t$$

die nun ohne weiteres mit der OLS-Methode geschätzt werden kann. Dabei kann jedoch nur der Gesamteinfluss $\beta_2 + \gamma\beta_3$ der beiden Einflussgrößen x_2 und x_3 ermittelt werden. Er kann nicht in die Einzeleinflüsse β_2 und β_3 separiert werden, so dass man diese beiden Parameter als **nicht identifizierbar** bezeichnet.

Schätztechnisch bedeutet dies, dass an Stelle einer Regressionsebene nur eine Regressionsgerade im dreidimensionalen Raum bestimmt werden kann. Wenn zwischen den beiden Regressoren x_2 und x_3 anstelle von (4.2) eine lineare Abhängigkeit der Form

$$(4.4) \quad x_{3t} = \delta + \gamma \cdot x_{2t}$$

bestehen würde, dann würde die schätzbare Regressionsgleichung

$$(4.5) \quad y_t = (\beta_1 + \delta\beta_3)x_{1t} + (\beta_2 + \gamma\beta_3)x_{2t} + u_t$$

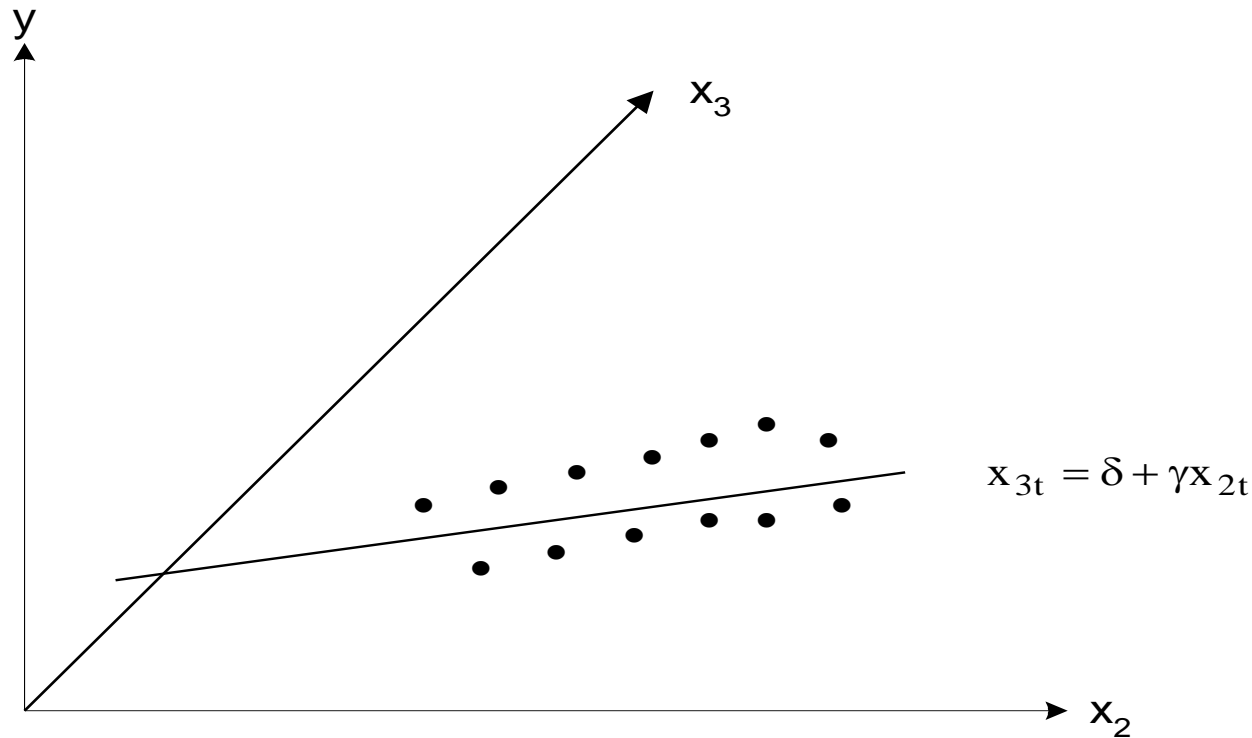
lauten. In diesem Fall wäre auch der Achsenabschnitt β_1 nicht mehr identifizierbar.

Auftreten exakter Multikollinearität:

- aufgrund definitorischer Beziehungen
- bei fehlerhafter Definition von Dummy-Variablen

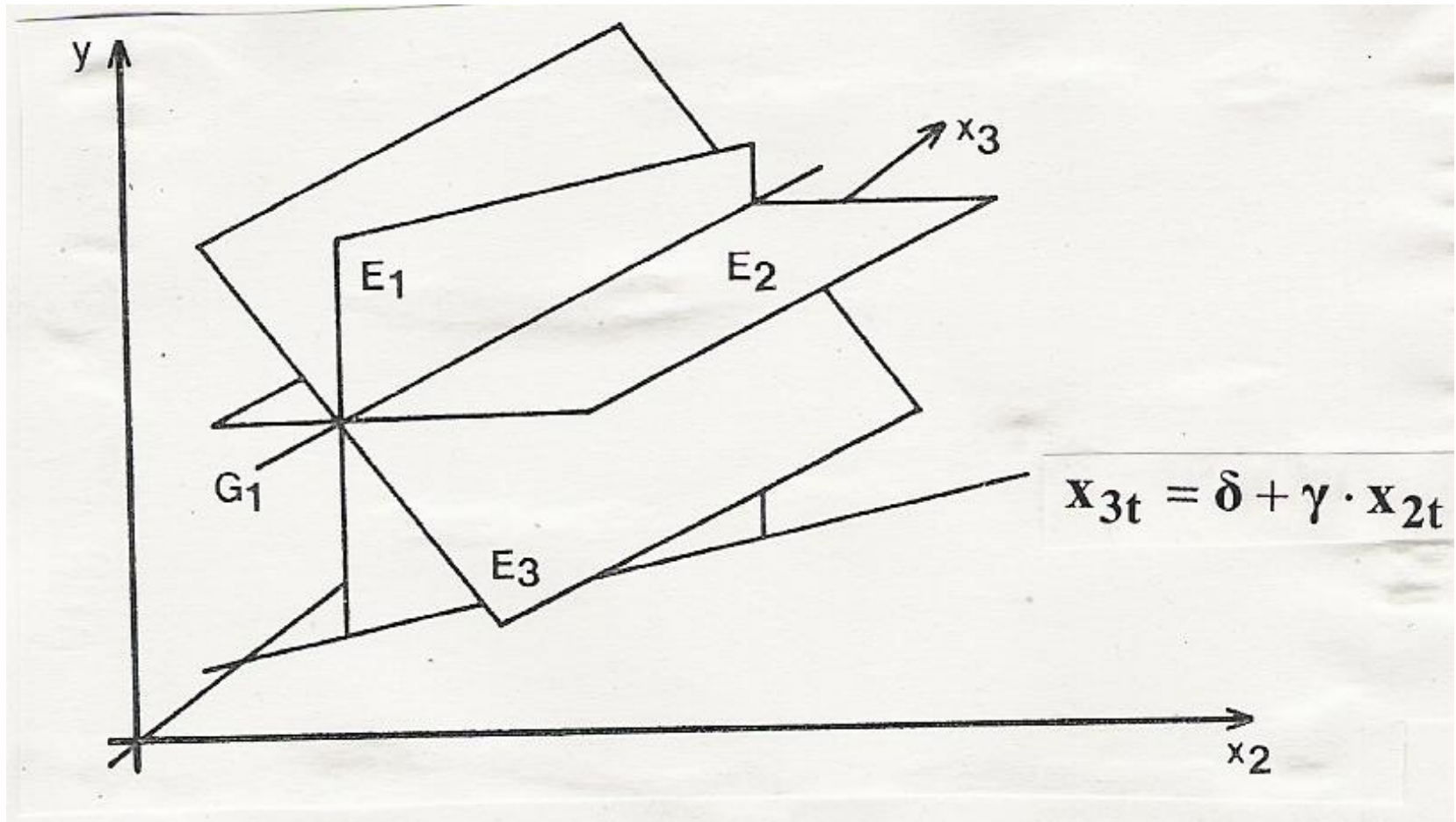
In der **ökonometrischen Praxis** tritt bei korrekter Modellierung in der Regel keine exakte Multikollinearität auf, sondern eine **nicht perfekte Multikollinearität**, die durch hohe, aber von +1 bzw. -1 verschiedene Korrelationen zwischen den unabhängigen Variablen zustande kommt.

Abbildung: Nicht-perfekte Multikollinearität



Die Schätzung der Regressionsebene in dem von den Variablen y , x_2 und x_3 aufgespannten Raum ist unsicher. Denn mit zunehmender Nähe der Projektionen der Beobachtungswerte zu der Geraden geht die Regressionsebene, die sich der dreidimensionalen Punktwolke optimal anpasst, immer mehr in eine Regressionsgerade im dreidimensionalen Raum über. Wenn alle projizierten Beobachtungen auf der Geraden (4.4) liegen, ist die Lage der Regressionsebene vollkommen unbestimmt. Je mehr die projizierten Beobachtungen dagegen in der x_2 , x_3 -Ebene streuen (=stärkere Tendenz zur Unabhängigkeit), umso breiter ist die Basis zur Bestimmung einer Regressionsebene im dreidimensionalen Raum.

Abb.: Perfekte Multikollinearität



4.2 Auswirkungen der Multikollinearität

Bei Multikollinearität, die aus statistischen Abhängigkeiten zwischen den exogenen Variablen resultiert, lässt sich der OLS-Schätzer $\hat{\beta}$ weiterhin berechnen. Jedoch lässt sich zeigen, dass die Varianzen der geschätzten Regressionskoeffizienten groß werden, womit ein Verlust an Präzision (Genauigkeit) der Schätzung einhergeht.

Um dies anhand des Regressionsmodells (4.1) zu zeigen, gehen wir der Einfachheit halber von den Abweichungen der Beobachtungswerte von ihren Mittelwerten aus:

$$(4.6) \quad \tilde{y}_t = \beta_2 \tilde{x}_{2t} + \beta_3 \tilde{x}_{3t} + u_t$$

mit

$$\tilde{y}_t = y_t - \bar{y}, \quad \tilde{x}_{2t} = x_{2t} - \bar{x}_2 \quad \text{und} \quad \tilde{x}_{3t} = x_{3t} - \bar{x}_3$$

Unter Verwendung von

$$\tilde{\mathbf{X}} = \begin{bmatrix} \tilde{x}_{21} & \tilde{x}_{31} \\ \tilde{x}_{22} & \tilde{x}_{32} \\ \vdots & \vdots \\ \tilde{x}_{2n} & \tilde{x}_{3n} \end{bmatrix} \quad \tilde{\mathbf{X}}' \tilde{\mathbf{X}} = \begin{bmatrix} \sum \tilde{x}_{2t}^2 & \sum \tilde{x}_{2t} \tilde{x}_{3t} \\ \sum \tilde{x}_{2t} \tilde{x}_{3t} & \sum \tilde{x}_{3t}^2 \end{bmatrix} \quad (\tilde{\mathbf{X}}' \tilde{\mathbf{X}})^{\text{adj}} = \begin{bmatrix} \sum \tilde{x}_{3t}^2 & -\sum \tilde{x}_{2t} \tilde{x}_{3t} \\ -\sum \tilde{x}_{2t} \tilde{x}_{3t} & \sum \tilde{x}_{2t}^2 \end{bmatrix}$$
$$|\tilde{\mathbf{X}}' \tilde{\mathbf{X}}| = \sum \tilde{x}_{2t}^2 \sum \tilde{x}_{3t}^2 - (\sum \tilde{x}_{2t} \tilde{x}_{3t})^2$$

erhält man für den OLS-Schätzer

$$\hat{\boldsymbol{\beta}} = (\tilde{\mathbf{X}}' \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}' \tilde{\mathbf{y}} \quad \text{mit} \quad \tilde{\mathbf{y}} = (\tilde{y}_1 \tilde{y}_2 \dots \tilde{y}_n)$$

die Varianz-Kovarianz-Matrix

$$(4.7) \quad \text{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\tilde{\mathbf{X}}' \tilde{\mathbf{X}})^{-1} = \frac{\sigma^2}{\sum \tilde{x}_{2t}^2 \sum \tilde{x}_{3t}^2 - (\sum \tilde{x}_{2t} \tilde{x}_{3t})^2} \begin{bmatrix} \sum \tilde{x}_{3t}^2 & -\sum \tilde{x}_{2t} \tilde{x}_{3t} \\ -\sum \tilde{x}_{2t} \tilde{x}_{3t} & \sum \tilde{x}_{2t}^2 \end{bmatrix}.$$

Die Varianzen der geschätzten Regressionskoeffizienten $\hat{\beta}_2$ und $\hat{\beta}_3$ lauten folglich

$$\text{Var}(\hat{\beta}_2) = \frac{\sigma^2 \sum \tilde{x}_{3t}^2}{\sum \tilde{x}_{2t}^2 \sum \tilde{x}_{3t}^2 - (\sum \tilde{x}_{2t} \tilde{x}_{3t})^2} \quad \text{und} \quad \text{Var}(\hat{\beta}_3) = \frac{\sigma^2 \sum \tilde{x}_{2t}^2}{\sum \tilde{x}_{2t}^2 \sum \tilde{x}_{3t}^2 - (\sum \tilde{x}_{2t} \tilde{x}_{3t})^2}$$

Man erhält eine geeignete Interpretation der Multikollinearität, wenn man hierin den Korrelationskoeffizienten r_{23} zwischen x_2 und x_3 einführt, der durch

$$r_{23} = \frac{\sum \tilde{x}_{2t} \tilde{x}_{3t}}{\sqrt{\sum \tilde{x}_{2t}^2 \sum \tilde{x}_{3t}^2}}$$

definiert ist.

Nach Quadrierung ergibt sich hieraus

$$(\sum \tilde{x}_{2t} \tilde{x}_{3t})^2 = r_{23}^2 \left(\sum \tilde{x}_{2t}^2 \sum \tilde{x}_{3t}^2 \right)$$

so dass die Varianzen von $\hat{\beta}_1$ und $\hat{\beta}_2$ in der Form

$$(4.8) \quad \text{Var}(\hat{\beta}_2) = \frac{\sigma^2}{(1 - r_{23}^2) \sum \tilde{x}_{2t}^2}$$

und

$$(4.9) \quad \text{Var}(\hat{\beta}_3) = \frac{\sigma^2}{(1 - r_{23}^2) \sum \tilde{x}_{3t}^2}$$

darstellbar sind. Bei multikollinearen Regressoren x_2 und x_3 geht r_{23} tendenziell gegen 1, so dass die Nenner von (4.8) und (4.9) klein und die Varianzen der geschätzten Regressionskoeffizienten groß werden. Eine Signifikanz eines einzelnen Regressionskoeffizienten ist dann schwieriger nachweisbar. Der Gesamtzusammenhang kann in dieser Situation auch dann statistisch gesichert sein, obwohl die geschätzten Regressionskoeffizienten insignifikant sind.

Außerdem lässt sich zeigen, dass die Korrelation zwischen den Regressionskoeffizienten $\hat{\beta}_2$ und $\hat{\beta}_3$

$$r_{\hat{\beta}_2, \hat{\beta}_3} = -r_{23}$$

beträgt. Das bedeutet, dass eine Überschätzung von $\hat{\beta}_2$ mit einer Unterschätzung von $\hat{\beta}_3$ einhergeht und umgekehrt.

4.3 Aufdeckung von Multikollinearität

Techniken zur Aufdeckung von Multikollinearität

- Einfache und erweiterte Korrelationsanalyse
- Determinante
- Konditionszahl
- Verfahren der Hilfsregressionen (auxiliary regressions)

Einfache und erweiterte Korrelationsanalyse

- Einfache Korrelationsanalyse

Aufdeckung von Multikollinearität aufgrund der paarweisen Korrelationskoeffizienten $r_{23}, r_{24}, \dots, r_{k-1,k}$.

Faustregel: Multikollinearität falls $r_{jk} > 0,8$

Das Verfahren ist hinreichend für die einfache Regressionsanalyse, deckt aber nicht wechselseitige Abhängigkeiten zwischen mehreren Einflussgrößen auf.

- Erweiterte Korrelationsanalyse

Beurteilung der Multikollinearität anhand des multiplen Korrelationskoeffizienten $R_{yx2, \dots, xk}$:

Faustregel: Multikollinearität falls $r_{jk} > R_{yx2, \dots, xk}$

Keine Aufhebung der Einschränkungen der einfachen Korrelationsanalyse, sondern Objektivierung des kritischen Wertes

- Determinante der Produktmatrix $\mathbf{X}'\mathbf{X}$

Bei Multikollinearität besteht eine näherungsweise lineare Abhängigkeit zwischen zwei oder mehreren Regressoren. Wenn die Spalten der Beobachtungsmatrix \mathbf{X} näherungsweise linear abhängig sind, wird die Produktmatrix $\mathbf{X}'\mathbf{X}$ fast singulär. Zwar lässt sich dann die für die Berechnung des OLS-Schätzers $\hat{\beta}$ benötigte Inverse $(\mathbf{X}'\mathbf{X})^{-1}$ noch berechnen, doch wird die Schätzung mit einem zunehmendem Grad der linearen Abhängigkeit instabil.

Bei exakter Multikollinearität gilt ist die Determinante der Produktmatrix $\mathbf{X}'\mathbf{X}$ gleich 0, so dass ein Wert von $|\mathbf{X}'\mathbf{X}|$ nahe 0 auf Multikollinearität hinweist:

$$|\mathbf{X}'\mathbf{X}| \approx 0 \rightarrow \text{Multikollinearität}$$

[Gleiches gilt für die Korrelationsmatrix \mathbf{R} , die man mit der standardisierten Beobachtungsmatrix \mathbf{Z} erhält: $\mathbf{R} = \mathbf{Z}'\mathbf{Z} / n$.]

- Konditionszahl der Produktmatrix $\mathbf{X}'\mathbf{X}$

Die Konditionszahl C gibt an, wie nahe eine Matrix an eine singuläre Matrix kommt.

$$C = \sqrt{\lambda_{\max} / \lambda_{\min}}$$

λ_{\max} : größter Eigenwert von $\mathbf{X}'\mathbf{X}$, λ_{\min} : kleinster Eigenwert von $\mathbf{X}'\mathbf{X}$

Mit zunehmendem Grad an Multikollinearität steigt das Verhältnis $\lambda_{\max} / \lambda_{\min}$.

Faustregel: Multikollinearität falls $C > 30$

- Verfahren der Hilfsregressionen (auxiliary regressions)

Das Verfahren der Hilfsregressionen besteht darin, dass für jede der $k-1$ (echten) exogenen Variablen eine Regression auf die verbleibenden $k-2$ Regressoren durchgeführt wird:

$$\begin{array}{rcll} \hat{x}_{2t} & = & \beta_1^2 x_{1t} & + \beta_3^2 x_{3t} + \cdots + \beta_k^2 x_{kt} \\ \hat{x}_{3t} & = & \beta_1^3 x_{1t} & + \beta_2^3 x_{2t} + \cdots + \beta_k^3 x_{kt} \\ \vdots & & \vdots & \ddots \vdots \\ \hat{x}_{kt} & = & \beta_1^k x_{1t} & + \beta_2^k x_{2t} + \cdots + \beta_{k-1}^k x_{k-1,t} . \end{array}$$

Für jede dieser Hilfsregressionen wird ein Bestimmtheitsmaß R_j^2 berechnet, bei dem der Index j angibt, dass x_j als abhängige Variable fungiert. Multikollinearität kann anhand von

- F-Tests oder
- Toleranzkoeffizienten und Varianzinflationsfaktoren

diagnostiziert werden.

a) F-Tests der Hilfsregressionen

Ein hohes R_j^2 weist eine lineare statistische Abhängigkeit der Variablen x_j mit den übrigen Regressoren aus. Die Signifikanz einer solchen Beziehung kann mit einem F-Test geprüft werden, dessen Prüfgröße

$$(4.10) \quad F_j = \frac{R_j^2 / (k - 2)}{(1 - R_j^2) / (n - k + 1)}, j = 2, 3, \dots, k$$

lautet.

Testentscheidung:

$F_j > F_{k-2; n-k+1; 1-\alpha} \rightarrow$ Signifikanter linearer statistischer Zusammenhang zwischen x_j und den übrigen Regressoren

Bei einem signifikanten Testergebnis wird auf Multikollinearität geschlossen. Das Verfahren identifiziert zugleich diejenigen unabhängigen Variablen, durch die die Multikollinearität hervorgerufen wird.

[Die Hilfsregressionen können in gleicher Form für Teilmengen der exogenen Variablen durchgeführt werden, wozu entsprechende Modifikationen des Tests vorzunehmen sind.]

b) Toleranzkoeffizienten und Varianzinflationsfaktoren

Bei einer deskriptiven Ausrichtung des Verfahrens der Hilfsregressionen wird Multikollinearität anhand von Toleranzkoeffizienten und Varianzinflationsfaktoren ermittelt. Die **Toleranzkoeffizienten** tol_j sind durch

$$(4.11) \quad \text{tol}_j = 1 - R_j^2, \quad j=2,3,\dots,k$$

definiert. Hierbei indizieren niedrige Werte der Toleranzkoeffizienten Multikollinearität.

Faustregel: Multikollinearität falls $\text{tol}_j < 0,10$

- Unter Verwendung der Toleranzkoeffizienten tol_j lassen sich die **Varianzinflationsfaktoren**

$$(4.12) \quad \text{VIF}_j = \frac{1}{\text{tol}_j}, \quad j=2,3,\dots,k$$

bilden, aus denen die durch Multikollinearität (Faustregel: $\text{VIF}_j > 10$) bedingte Inflationierung der Varianzen der geschätzten Regressionskoeffizienten hervorgeht. Genauer gibt die Kenngröße VIF_j die Erhöhung der Varianz $\text{Var}(\hat{\beta}_j)$ an, die darauf zurückgeführt werden kann, dass die Variable x_j mit den übrigen exogenen Variablen des Regressionsmodells nicht unkorreliert ist, sondern in einer durch das Bestimmtheitsmaß R_j^2 gemessenen kollinearen Beziehung steht.

Beispiel:

Mit der Geldnachfragefunktion (2.28) haben die logarithmierte Geldnachfrage ($\ln m$) durch die auf einer logarithmischen Skala gemessenen exogenen Variablen Einkommen ($\ln y$) und Zinssatz ($\ln r$) erklärt. Wir setzen jetzt die behandelten Verfahren zur Aufdeckung von Multikollinearität ein.

• Einfache und erweiterte Korrelationsanalyse

Korrelationskoeffizienten r_{23} zwischen $\ln y$ und $\ln r$: -0.689

$|r_{23}| = 0,689 < 0,8 \rightarrow$ keine gravierende Multikollinearität

$|r_{23}| = 0,689 < R_{yx2x3} = 0,984 \rightarrow$ keine gravierende Multikollinearität

• Determinante der Produktmatrix $X'X$

$$\begin{aligned} |X'X|_{3 \times 3} &= \det \begin{bmatrix} n & \sum \ln y_t & \sum \ln r_t \\ \sum \ln y_t & \sum (\ln y_t)^2 & \sum \ln y_t \cdot \ln r_t \\ \sum \ln r_t & \sum \ln r_t \cdot \ln y_t & \sum (\ln r_t)^2 \end{bmatrix} \\ &= \det \begin{bmatrix} 19,000 & 145,986 & 17,779 \\ 145,986 & 1121,937 & 135,657 \\ 17,779 & 135,657 & 23,963 \end{bmatrix} = 18,776 \gg 0 \rightarrow \text{keine grav. Multikoll.} \end{aligned}$$

- **Determinante der Korrelationsmarix R**

$$|\mathbf{R}| = \det \begin{bmatrix} 1 & 0,689 \\ 0,689 & 1 \end{bmatrix} = 1 - 0,689^2 = 0,5253 \gg 0 \rightarrow \text{keine grav. Multikollinearität}$$

- **Verfahren der Hilfsregressionen**

In dem Verfahren der Hilfsregressionen ist allein *eine* Regression durchzuführen, da es im Hinblick auf die Bestimmtheit gleichgültig ist, ob $\ln y$ auf $\ln r$ oder $\ln r$ auf $\ln y$ regressiert wird.

$$\text{Hilfsregression: } \ln y = 7,804 - 0,129 \cdot \ln r, \quad R_2^2 = 0,475$$

- F-Test**

Prüfgröße ($n=19, k=3$):

$$F_2 = \frac{R_2^2 / (k - 2)}{(1 - R_2^2) / (n - k + 1)} = \frac{0,475 / (3 - 2)}{(1 - 0,475) / (19 - 3 + 1)} = \frac{0,475}{0,923 / 17} = 15,381$$

Kritischer Wert ($\alpha=0,05$): $F_{1;17;0,95} = 4,45$

Testentscheidung:

$F_2 = 15,381 < F_{1;18;0,95} = 4,45 \rightarrow$ Signifikanter statistischer linearer
Zusammenhang zwischen x_2 und x_3
 \rightarrow Hinweis auf Multikollinearität

- Toleranzkoeffizient und Varianzinflationsfaktor

Toleranzkoeffizient:

$$\text{tol}_2 = 1 - R_2^2 = 1 - 0,475 = 0,525 > 0,10 \rightarrow \text{keine grav. Multikollinearität}$$

Varianzinflationsfaktor:

$$\text{VIF}_2 = \frac{1}{\text{tol}_2} = \frac{1}{0,525} = 1,904$$

Im Vergleich zur Situation unkorrelierter Regressoren wird die Varianz des OLS-Schätzers $\hat{\beta}_2$ (und damit auch die des OLS-Schätzers $\hat{\beta}_3$) durch die schwache lineare statistische Abhängigkeit zwischen x_2 und x_3 um den Faktor 1,904 inflationiert. Dieser Effekt einer Erhöhung der Präzision der Schätzung ist in Kauf zu nehmen, wenn das Vorhandensein von Multikollinearität vernachlässigt wird.

4.4 Überwindung von Multikollinearität

Während das Multikollinearitätsproblem für **Prognosezwecke** nicht störend zu sein braucht, da hier auf den Gesamteffekt der erklärenden Variablen auf die zu prognostizierende Variable abgestellt wird, ist es bei **ökonomischen Strukturanalysen** häufig wünschenswert, sie auszuschalten oder zumindest zu vermindern.

Verfahren zur Überwindung von Multikollinearität

- **Erweiterung der Datenbasis**

Bei einer Erweiterung der Datenbasis ist eine größere Variationsbreite der Daten zu erwarten, was tendenziell mit einer Verminderung der Multikollinearität einhergeht.

Querschnittsdaten:

Erhöhung des Stichprobenumfangs evtl. bei Querschnittseinheiten wie Konsumenten, Haushalte, Unternehmen möglich, nicht jedoch bei räumlichen Einheiten wie regionalen Arbeitsmärkten oder administrativen Regionen.

Zeitreihendaten:

Die Einbeziehung weiter zurück liegender vergangener Daten bringt die Gefahr eines Strukturbruchs mit sich.

Paneldaten:

Durch die Kombination von Querschnitts- mit Zeitreihendaten wird die Variationsbreite der Daten häufig deutlich vergrößert, wodurch das Multikollinearitätsproblem erheblich entschärft werden kann.

- **Externe Informationen**

Angenommen, es wird die Nachfrage nach einer Gütergruppe in Abhängigkeit vom Einkommen und dem Preisniveau unter Verwendung von Zeitreihendaten ökonometrisch untersucht. Da das Einkommen häufig mit dem Preisniveau korreliert sein wird, ist mit Multikollinearität zu rechnen. Wenn nun die Größenordnung des Einflusses des Einkommens auf die Nachfrage der Gütergruppe z.B. aus einer Querschnittserhebung bekannt ist, könnte auf diese **externe Information** zurückgegriffen werden, um ökonometrisch den separaten Einfluss des Preises auf die Nachfrage zu schätzen.

Problematik:

Im Rahmen der **Zeitreihenanalyse** könnte ein **langfristiger Reaktionskoeffizient** des Einkommens auf die Nachfrage gesucht sein, obwohl aus der **Querschnittsanalyse** nur ein kurzfristiger Reaktionskoeffizient bestimmbar ist.

- **Verfahren der Variablenunterdrückung**

Wenn z.B. zwei Variablen hoch korreliert sind, erfasst eine der beiden Variablen zugleich den größten Teil des Einflusses der anderen Variablen, so dass der "Erklärungsgehalt" bei der Unterdrückung einer Variablen im Großen und Ganzen erhalten bleibt. Die Multikollinearität wird dadurch nicht nur vermindert, sondern sogar völlig ausgeschaltet.

Problematik:

Fehlspezifikation der Regressionsgleichung (Störvariable fängt die Systematik auf)

• Verfahren der Differenzbildung

Beim **Verfahren der Differenzbildung** wird davon ausgegangen, dass die Multikollinearität in ökonometrischen Modellen durch einen Trend in den exogenen Variablen auftritt. Wenn man näherungsweise einen linearen Trend unterstellen kann, ist seine Ausschaltung durch eine einfache Differenzenbildung gegeben. Bei zwei exogenen Variablen (ohne Scheinvariable) lautet die Regressionsgleichung für die Periode t

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$$

Subtrahiert man hiervon die Regressionsgleichung für die Vorperiode $t-1$,

$$y_{t-1} = \beta_1 + \beta_2 x_{2,t-1} + \beta_3 x_{3,t-1} + u_{t-1}$$

so erhält man das Regressionsmodell

$$(4.13) \quad \Delta y_t = \beta_2 \Delta x_{2t} + \beta_3 \Delta x_{3t} + \Delta u_t$$

wobei z.B. $\Delta y_t = y_t - y_{t-1}$ bedeutet. In dem Maße, wie es gelingt, den Trend aus den exogenen Variablen x_2 und x_3 auszuschalten, vermindert sich die Multikollinearität gegenüber dem ursprünglichen Modell mit den Niveauvariablen.

Problematik:

- Verringerung des Streubereichs der Regressoren durch Differenzenbildung;
- Autokorrelation der Störvariablen

• Hauptkomponentenregression (principal component regression)

Die Hauptkomponentenregression ist ein multivariates Verfahren, das auf der Hauptkomponentenanalyse basiert. Mit der Hauptkomponentenanalyse werden die exogenen Variablen eines Regressionsmodells in unkorrelierter künstliche Variablen (=Hauptkomponenten) transformiert. Die Hauptkomponenten ergeben sich durch Linearkombinationen der Regressoren. Häufig erklären einige wenige Hauptkomponenten den größten Teil der Varianz der Regressoren.

Setzt man diese unkorrelierten Hauptkomponenten anstelle der beobachtbaren exogenen Variablen in die Regressionsgleichung ein, lassen sich die ihnen zugehörigen Regressionskoeffizienten trennscharf schätzen. Falls die Hauptkomponenten ökonomisch interpretierbar sind, kann sich der Ökonom genau für die Reaktionkoeffizienten interessieren. Im Allgemeinen ist das Interesse jedoch weiterhin auf die Wirkungen der originären Variablen ausgerichtet. In diesem Fall müssen die geschätzten Regressionskoeffizienten der Hauptkomponenten rücktransformiert werden.

Problematik:

- Die geschätzten Regressionskoeffizienten werden durch die Anzahl der verwendeten Hauptkomponenten beeinflusst.
- Die Rücktransformation ist mit einem Gewichtungproblem verbunden, das sich ebenfalls in den geschätzten Regressionskoeffizienten widerspiegelt.