

Kapitel 7

Multikollinearität

“God abhors a naked singularity.”
(Stephen Hawking)

7.1 Problem

Von Multikollinearität, bzw. Kollinearität, spricht man, wenn zwei oder mehrere erklärende x Variablen hoch untereinander korreliert sind.¹

Die Gauss-Markov Annahmen verlangen von der \mathbf{X} Matrix lediglich, dass sie vollen Spaltenrang hat, und dass die erklärenden Variablen exogen sind. Im Falle *perfekter Kollinearität* ist die Annahme des vollen Spaltenrangs der \mathbf{X} Matrix verletzt, d.h. zumindest eine der x Variablen lässt sich als Linearkombination von einer oder mehreren anderen x Variablen darstellen. In diesem Fall ist die Matrix $\mathbf{X}'\mathbf{X}$ singulär und kann deshalb nicht invertiert werden. Aus diesem Grund ist im Fall *perfekter Kollinearität* der OLS-Schätzer $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ nicht definiert.

Perfekte Kollinearität stellt nur selten ein Problem dar und ist auch sehr einfach zu erkennen, da der Schätzer $\hat{\beta}$ bei perfekter Kollinearität gar nicht berechnet werden kann. EViews würde z.B. mit der Fehlermeldung ‘*Near singular matrix*’ abbrechen, Stata würde die entsprechenden Variablen automatisch eliminieren.

Am häufigsten passiert dieser Fehler mit Dummy-Variablen. Zum Beispiel könnte eine Dummyvariable D^m definiert werden, die den Wert Eins hat, wenn es sich beim Untersuchungsobjekt um einen Mann handelt, und den Wert Null, wenn es sich um eine Frau handelt. Genauso könnte man eine eine Dummyvariable D^w definieren mit

$$D^w_i = \begin{cases} 1 & \text{wenn } i \text{ weiblich,} \\ 0 & \text{sonst.} \end{cases}$$

Wenn nun jemand die Auswirkungen des Geschlechts auf eine Variable y (z.B. Lohn) untersuchen möchte, könnte er auf die Idee kommen, folgende Gleichung zu schätzen

$$y_i = \hat{\beta}_1 + \hat{\beta}_2 D_i^m + \hat{\beta}_3 D_i^w + \hat{\beta}_4 x_i + \hat{\varepsilon}_i$$

¹Der Begriff wurde 1934 von Ragnar Frisch geprägt. Genau gesprochen spricht man von Multikollinearität bei mehreren linearen Abhängigkeiten, und von Kollinearität bei einer linearen Abhängigkeit zwischen den erklärenden Variablen, aber im üblichen Sprachgebrauch werden die Begriffe meist synonym gebraucht.

Die \mathbf{X} Matrix würde in diesem Fall z.B. folgendermaßen aussehen

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 1 & x_1 \\ 1 & 1 & 0 & x_2 \\ 1 & 0 & 1 & x_3 \\ 1 & 1 & 0 & x_4 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & x_n \end{pmatrix}$$

Man beachte, dass in diesem Fall die ersten drei Spalten linear abhängig sind, da die Regressionskonstante (1. Spalte, ein $\mathbf{1}$ -Vektor) immer die Summe von D^m und D^w ist (2. und 3. Spalte, d.h. $\mathbf{1} = D^m + D^w$). Aufgrund dieser linearen Abhängigkeit hat die \mathbf{X} Matrix nicht vollen Spaltenrang, die Matrix $\mathbf{X}'\mathbf{X}$ ist singulär und man kann keinen der Koeffizienten $\hat{\beta}_1, \dots, \hat{\beta}_4$ berechnen! Man nennt dies manchmal eine *Dummyvariablenfalle*. Die Lösung dieses Problems ist einfach – wir lassen einfach D^m (oder D^w) weg und das Ganze funktioniert wieder.

Wenn einfach von **Multikollinearität** gesprochen wird versteht man darunter üblicherweise, dass einzelne x Variablen zwar hoch miteinander korreliert sind, aber nicht exakt linear abhängig sind.

Wir wollen gleich anfangs festhalten, dass auch im Fall hoher Multikollinearität *keine* Gauss-Markov Annahme verletzt ist, und deshalb der OLS-Schätzer **auch in kleinen Stichproben BLU** (*best linear unbiased*) ist, d.h. unverzerrt und varianzminimal! Nicht nur die Koeffizienten werden effizient geschätzt, auch die t-Statistiken sind unverzerrt, weshalb auch Konfidenzintervalle und Teststatistiken gültig bleiben.

Wo also liegt das Problem? Nun, wir werden gleich sehen, dass in diesem Fall selbst ein ‘bestmöglicher’ (oder genauer, *effizienter*) Schätzer nicht wirklich gut ist. In solchen Fällen können die Koeffizienten zwar berechnet werden, aber nur sehr ungenau (d.h. die geschätzten Koeffizienten haben große Standardfehler). Die Konfidenzintervalle sind zwar gültig, aber sehr breit.

Man versteht Multikollinearität am besten als ein Datenproblem, das sehr stark an die Probleme einer sehr kleinen Stichprobe erinnert, denn wenn die Daten hoch korreliert sind enthalten sie häufig zu wenig Information, um die interessierenden Effekte mit der gewünschten Genauigkeit schätzen zu können.²

Der große Ökonometriker Arthur Goldberger (1930 – 2009) amüsiert sich in seinem Lehrbuch (Goldberger, 1991) über die Ökonomen, die sich von der eher technischen Bezeichnung *Multikollinearität* blenden lassen und nicht erkennen, dass es sich bei Multikollinearität im Kern um genau das gleiche Problem handelt wie bei kleinen Stichproben — die Standardfehler werden völlig korrekt gemessen, sind aber groß. Um dem Problem kleiner Stichproben einen ähnlichen Stellenwert zu geben wie der Multikollinearität schlägt er augenzwinkernd vor, ‘not having a lot of data’ *Micronumerosity* zu nennen, siehe Exkurs Seite 3.

²Ganz ähnliche Probleme entstehen auch, wenn ein oder mehrere Regressoren sehr wenig Streuung aufweisen (d.h. eine kleine Varianz haben), da auch in diesem Fall die Schätzung der betreffenden Koeffizienten sehr ungenau wird.

Exkurs: **Micronumerosity**, Arthur S. Goldberger, *A Course in Econometrics* (1991), Chapter 23.3

“Econometrics texts devote many pages to the problem of multicollinearity in multiple regression, but they say little about the closely analogous problem of small sample size in estimating a univariate mean. Perhaps that imbalance is attributable to the lack of an exotic polysyllabic name for “small sample size”. If so, we can remove that impediment by introducing the term *micronumerosity*.

Suppose an econometrician set out to write a chapter about small sample size in sampling from a univariate population. Judging from what is now written about multicollinearity, the chapter might look like this:

Micronumerosity

The extreme case, “exact micronumerosity”, arises when $n = 0$; in which case the sample estimate of μ is not unique. (Technically, there is a violation of the rank condition $n > 0$: the matrix 0 is singular.) The extreme case is easy enough to recognize. “Near micronumerosity” is more subtle, and yet very serious. It arises when the rank condition $n > 0$ is barely satisfied. Near micronumerosity is very prevalent in empirical economics.

Consequences of micronumerosity

The consequences of micronumerosity are serious. Precision of estimation is reduced. There are two aspects of this reduction: estimates of μ may have large errors, and not only that, but $V(\bar{y})$ will be large.

Investigators will sometimes be led to accept the hypothesis $\mu = 0$ because $u^0 = \bar{y}/\hat{\sigma}_{\bar{y}}$ is small, even though the true situation may be not that $\mu = 0$ but simply that the sample data have not enabled us to pick μ up. The estimate of μ will be very sensitive to sample data, and the addition of a few more observations can sometimes produce drastic shifts in the sample mean.

The true μ may be sufficiently large for the null hypothesis $\mu = 0$ to be rejected, even though $V(\bar{y}) = \sigma/n$ is large because of micronumerosity. But if the true μ is small (although nonzero) the hypothesis $\mu = 0$ may mistakenly be accepted.

Testing for micronumerosity

Tests for the presence of micronumerosity require the judicious use of various fingers. Some researchers prefer a single finger, others use their toes, still others let their thumbs rule. A generally reliable guide may be obtained by counting the number of observations. Most of the time in econometric analysis, when n is close to zero, it is also far from infinity.

Several test procedures develop critical values n^* ; such that micronumerosity is a problem only if n is smaller than n^* : But those procedures are questionable.

Remedies for micronumerosity

If micronumerosity proves serious in the sense that the estimate of μ has an unsatisfactorily low degree of precision, we are in the statistical position of not being able to make bricks without straw. The remedy lies essentially in the acquisition, if possible, of larger samples from the same population.

But more data are no remedy for micronumerosity if the additional data are simply “more of the same”. So obtaining lots of small samples from the same population will not help.

If we return from this fantasy to reality, several lessons may be drawn.

* Multicollinearity is no more (or less) serious than micronumerosity. [...]”

7.2 Drei Arten Multikollinearität graphisch darzustellen

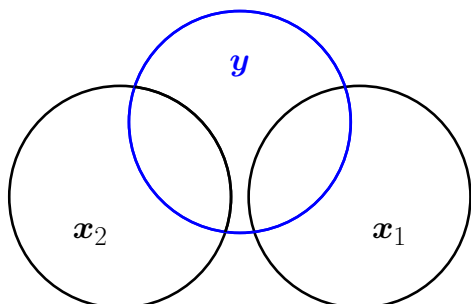
Es gibt zumindest drei verschiedene Möglichkeiten, sich das Problem der Multikollinearität graphisch vorzustellen. Die vermutlich einfachste Möglichkeit stellen Venn-Diagramme dar.

7.2.1 Venn-Diagramme

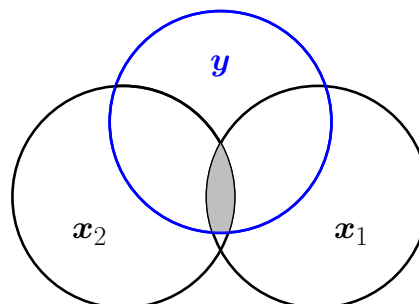
In Abbildung 7.1 symbolisiert der Kreis y die Streuung von y , und die Kreise x_1 und x_2 die Streuung von x_1 und x_2 . Die Überlappungsfläche von y und x_1 (bzw. x_2) symbolisiert die ‘gemeinsame’ Streuung, bzw. die Streuung, die durch x_1 (bzw. x_2) ‘erklärt’ wird. Wenn sich aber auch x_1 und x_2 überlappen kann dieser Überlappungsbereich nicht einer einzelnen x Variable zugerechnet werden; es gelingt nur schlecht den Einfluss einer einzelnen Variable zu isolieren.

Die Überlappungsfläche zwischen x_1 und x_2 (der grau schattierte Bereich) symbolisiert die Multikollinearität. Umso höher x_1 und x_2 korreliert sind, umso größer ist der Überlappungsbereich.

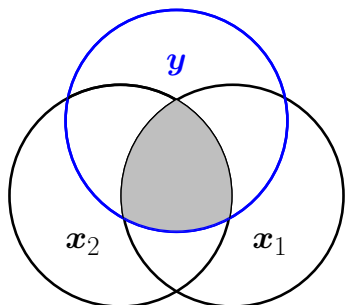
a) Keine Multikollinearität:



b) Niedrige Multikollinearität:



c) Hohe Multikollinearität:



d) Sehr hohe Multikollinearität:

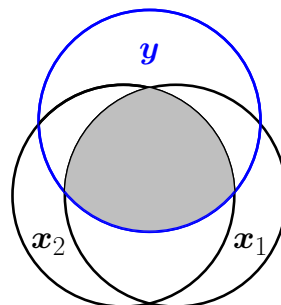


Abbildung 7.1: Multikollinearität, Venn-Diagramme nach Ballentine

Im Panel a) von Abbildung 7.1 erklären x_1 und x_2 je einen Teil der Streuung von y , aber x_1 und x_2 überlappen sich nicht, deshalb kann der Einfluss von x_1 und x_2 gut gemessen werden.

In den nächsten Panels überlappen sich x_1 und x_2 zunehmend, wodurch es schwieriger wird den Einfluss von x_1 oder x_2 isoliert zu messen.

Offensichtlich erklären x_1 und x_2 in Panel d) *gemeinsam* einen hohen Anteil der Streuung von y , aber der Einfluss kann nur in einem kleinen Ausmaß individuell den erklärenden Variablen zugerechnet werden (in der Abbildung die *nicht* grau schattierte Überlappungsfläche mit y). Dies ist auch der Grund, warum die durch hohe Multikollinearität verursachten Probleme den Problemen einer zu kleinen Stichprobe ähneln, es gelingt nur schlecht den isolierten Einfluss der einzelnen Variablen zu messen, weshalb die Standardfehler der Koeffizienten groß sind und einzelne Beobachtungen einen großen Einfluss auf die Koeffizienten haben können.

7.2.2 Grafische Darstellung im x, y Raum

Panel a) in Abbildung 7.2 zeigt zwei unkorrelierte x -Variablen x_1 und x_2 . Die schwarzen Punkte sind die beobachteten Punkte im Raum. Die (gelben) Kreise sind die dazugehörigen gefitteten Werte (\hat{y}), die *auf* der Regressionsebene liegen. Im ersten Fall haben beide Variablen eine große ‘horizontale’ Streuung. Deshalb wird die (gelbe) Regressionsebene gut ‘gestützt’; eine einzelne Beobachtung hat vermutlich keine große Auswirkung auf die Lage der Regressionsebene.

Panel b) und c) in Abbildung 7.2 zeigen Fälle mit einer größeren Korrelation zwischen x_1 und x_2 . Umso größer die Korrelation ist, umso näher liegen die Punkte in horizontaler Richtung beisammen (die horizontale Streuung ist kleiner, die Streuung in vertikaler Richtung ist in allen vier Grafiken gleich groß), und umso ‘wackeliger’ wird die Regressionsebene. Eine einzelne Beobachtung kann bei einer hohen Korrelation zwischen den x -Variablen schon beträchtliche Auswirkungen auf die Lage der Regressionsebene haben.

Im Extremfall, wenn der Korrelationskoeffizient zwischen x_1 und x_2 den Wert Eins hat (x_1 und x_2 also linear abhängig sind), ist die horizontale Streuung Null, die Werte streuen nur in vertikaler Richtung. In diesem Fall liegen alle gefitteten Werte genau auf einer Linie, und die Lage der Regressionsebene ist nicht mehr definiert, man kann sie beliebig um diese Gerade drehen, der Regressionsansatz bricht zusammen.³

7.2.3 Grafische Darstellung im Euklidischen Vektorraum

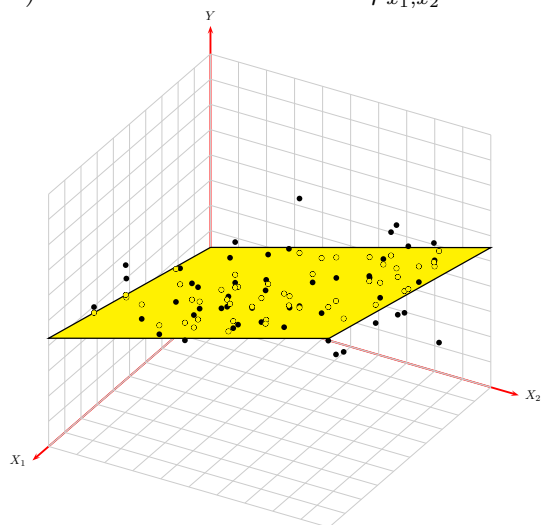
Wir erinnern uns, dass im Euklidischen Vektorraum jede x -Variable *ein* Vektor ist, und alle x -Variablen gemeinsam den Spaltenraum aufspannen, in den der y -Vektor *orthogonal* projiziert wird.

Aus der linearen Algebra ist bekannt, dass das Skalarprodukt von zwei Vektoren die Länge der Vektoren multipliziert mit dem Kosinus des eingeschlossenen Winkels ist, $\mathbf{x}'_1 \mathbf{x}_2 = \|\mathbf{x}_1\| \cdot \|\mathbf{x}_2\| \cos(\alpha)$, wobei die Länge definiert ist, d.h. $\|\mathbf{x}_1\| = \sqrt{\mathbf{x}'_1 \mathbf{x}_1}$. Daraus folgt

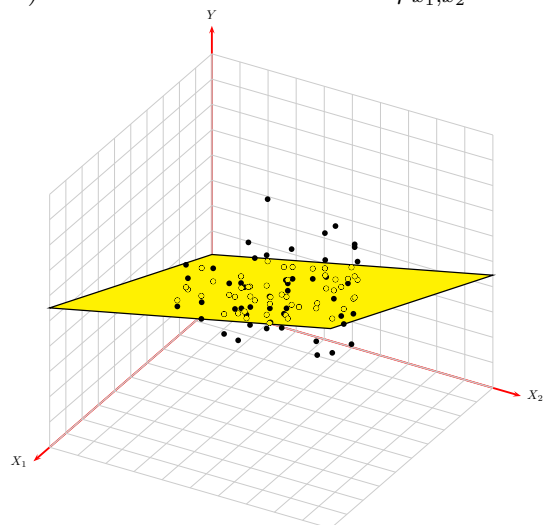
$$\cos(\alpha) = \frac{\mathbf{x}'_1 \mathbf{x}_2}{\|\mathbf{x}_1\| \cdot \|\mathbf{x}_2\|}$$

³Die Lage der Linie ist in der 3-dimensionalen Abbildung nicht sehr gut erkennbar, man stellt sich am besten vor, dass die gefitteten Werte von Panel a) einfach ‘zusammengeschoben’ werden, sodass sie alle auf einer Linie liegen.

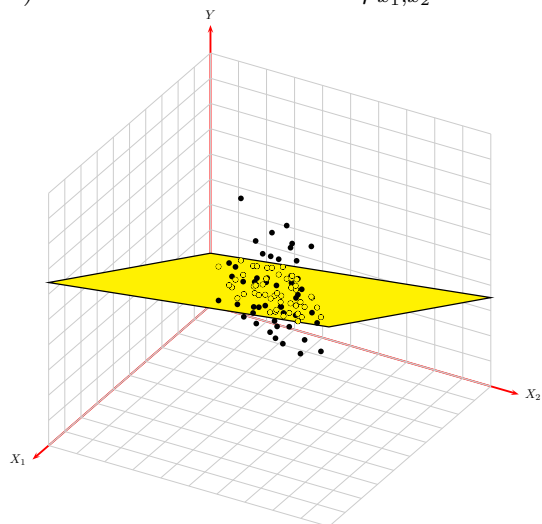
a) Keine Multikollinearität: $\rho_{x_1, x_2} = 0$



b) Mittlere Multikollinearität: $\rho_{x_1, x_2} = 0.5$



c) Hohe Multikollinearität: $\rho_{x_1, x_2} = 0.9$



d) Perfekte Multikollinearität: $\rho_{x_1, x_2} = 1$

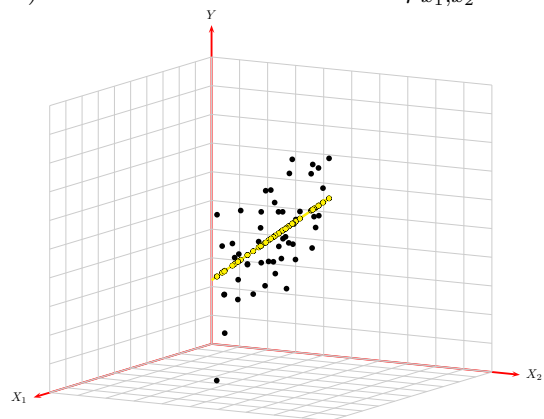


Abbildung 7.2: Multikollinearität (ρ bezeichnet den Korrelationskoeffizienten zwischen x_1 und x_2 ; die Perspektive der Abbildung d) wurde geändert um den Verlauf der Geraden besser zu veranschaulichen).

Der Kosinus von 90 Grad ist Null und der Kosinus von 0 Grad ist Eins, deshalb sind bei $\cos(\alpha) = 0$ die Vektoren orthogonal, und wenn $\cos(\alpha) = 1$ liegen die Vektoren ‘aufeinander’.

Umso höher die Korrelation zwischen zwei x Variablen ist, umso kleiner ist der eingeschlossene Winkel, denn man kann zeigen, dass der Korrelationskoeffizient zwischen zwei zentrierten Variablen (d.h. wenn von jeder Variablen der Mittelwert subtrahiert wird, z.B. $\ddot{x}_{ih} = x_{ih} - \bar{x}_h$ wobei \bar{x}_h der Mittelwert der x Variable h über alle i ist), gleich dem Kosinus des eingeschlossenen Winkels ist.

$$\cos(\alpha) = \frac{\sum_i (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)}{\sqrt{\sum_i (x_{i1} - \bar{x}_1)^2} \sqrt{\sum_i (x_{i2} - \bar{x}_2)^2}} = r$$

Wenn zwei x Vektoren hoch korreliert sind zeigen sie ‘fast’ in die gleiche Richtung. Der Spaltenraum, der von zwei hoch korrelierten Vektoren aufgespannt wird, ist

deshalb zwar exakt definiert, aber da die hoch korrelierten Vektoren ‘fast’ in die gleiche Richtung zeigen, lässt sich der individuelle Einfluss einer einzelnen Variable nur ungenau messen. Darüber hinaus können kleine Unterschiede in einer der Beobachtungen großen Einfluss auf die Lage des Spaltenraums haben, auf den projiziert wird, was den häufig starken Einfluss einzelner Beobachtungen im Fall multikollinearere Variablen erklärt (vgl. Abbildung 7.3).

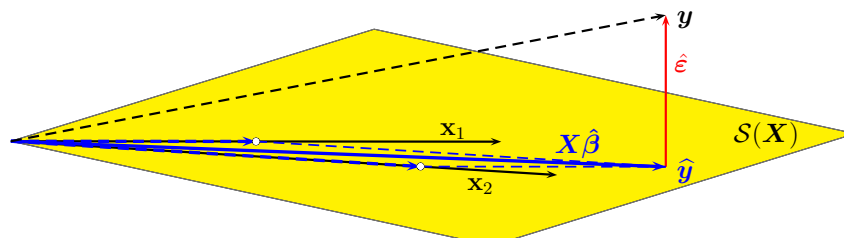


Abbildung 7.3: Bei hoher Multikollinearität ist der Winkel zwischen den erklärenden Variablen klein, die Schätzung wird ‘ungenau’.

Im Falle perfekter Kollinearität liegen die Vektoren x_1 und x_2 aufeinander, es gibt unendlich viele Kombinationen von $\hat{\beta}_2$ und $\hat{\beta}_3$, die alle zum gleichen Ergebnis führen, d.h. $\hat{\beta}_2$ und $\hat{\beta}_3$ sind nicht mehr eindeutig definiert, der Regressionsansatz bricht zusammen.

7.3 Konsequenzen der Multikollinearität

Ein Regressionskoeffizient misst die Auswirkungen auf die abhängige Variable y , wenn sich *eine* erklärende Variable *unter Konstanthaltung aller anderen x -Variablen* ändert (d.h. *ceteris paribus*). Wenn die erklärenden Variablen hoch korreliert sind, lässt sich der Einfluss einer einzelnen Variablen schlecht isolieren, was sich in großen Standardfehlern niederschlägt.

Das Problem kann anhand der einfachsten multiplen Regressionsgleichung

$$y_i = \hat{\beta}_1 + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3} + \hat{\epsilon}_i$$

dargestellt werden. Da das Interzept selten interessiert schreiben wir das Modell wieder in Abweichungsform, um die Darstellung zu vereinfachen

$$\ddot{y}_i = \hat{\beta}_2 \ddot{x}_{i2} + \hat{\beta}_3 \ddot{x}_{i3} + \hat{\epsilon}_i$$

(mit $\ddot{y}_i = y_i - \bar{y}$ usw. Das Interzept fällt bei der Differenzenbildung weg).

Die Varianz-Kovarianzmatrix von $\hat{\beta}$ ist

$$\text{var}(\hat{\beta}) = \begin{pmatrix} \text{var}(\hat{\beta}_1) & \text{cov}(\hat{\beta}_1, \hat{\beta}_2) \\ \text{cov}(\hat{\beta}_1, \hat{\beta}_2) & \text{var}(\hat{\beta}_2) \end{pmatrix} = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} = \sigma^2 \begin{pmatrix} \sum \ddot{x}_2^2 & \sum \ddot{x}_2 \ddot{x}_3 \\ \sum \ddot{x}_2 \ddot{x}_3 & \sum \ddot{x}_3^2 \end{pmatrix}^{-1}$$

Mit einiger Algebra kann man zeigen, dass⁴

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{(1 - r_{23}^2) \sum \ddot{x}_2^2} \quad \text{und} \quad \text{var}(\hat{\beta}_3) = \frac{\sigma^2}{(1 - r_{23}^2) \sum \ddot{x}_3^2}$$

⁴vgl. z.B. Gujarati, D. (1995), *Basic Econometrics*, McGraw-Hill, 3ed., p. 328.

wobei r_{23} der Korrelationskoeffizient zwischen x_2 und x_3 ist. Wenn der Korrelationskoeffizient Null ist erhalten wir die gleichen Standardfehler wie für bivariate Regressionen, wenn der Korrelationskoeffizient aber gleich Eins ist, ist der Nenner Null und die Standardfehler sind nicht definiert (bzw. unendlich groß).

Dies lässt sich für den Mehrvariablen-Fall verallgemeinern zu (vgl. Greene, 2007, 59)

$$\text{var}(\widehat{\beta}_h) = \frac{\sigma^2}{(1 - R_h^2) \sum_{i=1}^n (x_{ih} - \bar{x}_{.h})^2}$$

wobei R_h^2 das Bestimmtheitsmaß einer *Hilfsregression* ist, bei der die erklärende Variable x_h auf alle anderen x -Variablen (inkl. Regressionskonstante) regressiert wird. Solche ‘Hilfsregressionen’ (*auxiliary regressions*) spielen in der Ökonometrie eine wichtige Rolle, weshalb wir sie hier an diesem Beispiel etwas ausführlicher erläutern.

Wenn wir z.B. von der Gleichung

$$y_i = \widehat{\beta}_1 + \widehat{\beta}_2 x_{i2} + \widehat{\beta}_3 x_{i3} + \widehat{\beta}_4 x_{i4} + \widehat{\varepsilon}_i$$

ausgehen kann man folgende drei Hilfsregressionen schätzen und die entsprechenden Bestimmtheitsmaße dazu berechnen

$$\begin{aligned} x_{i2} &= \widehat{\beta}_1^a + \widehat{\beta}_2^a x_{i3} + \widehat{\beta}_3^a x_{i4} + \widehat{\varepsilon}_i^a && \rightarrow R_2^2 \\ x_{i3} &= \widehat{\beta}_1^b + \widehat{\beta}_2^b x_{i2} + \widehat{\beta}_3^b x_{i4} + \widehat{\varepsilon}_i^b && \rightarrow R_3^2 \\ x_{i4} &= \widehat{\beta}_1^c + \widehat{\beta}_2^c x_{i2} + \widehat{\beta}_3^c x_{i3} + \widehat{\varepsilon}_i^c && \rightarrow R_4^2 \end{aligned}$$

Das Bestimmtheitsmaß jeder dieser Hilfsregressionen zeigt das Ausmaß der linearen Abhängigkeiten zwischen der Variable x_h (für $h = 2, 3, 4$) und den beiden anderen x -Variablen, und diese (nicht perfekte) lineare Abhängigkeit wirkt sich – wie wir bereits gesehen haben – auf die Präzision der Schätzungen aus, denn

$$\text{var}(\widehat{\beta}_h) = \frac{\sigma^2}{(1 - R_h^2) \sum_{i=1}^n (x_{ih} - \bar{x}_{.h})^2}$$

Man kann in dieser Darstellungsform sehr schön die Determinanten der Varianz von Koeffizienten im multiplen Regressionsmodell erkennen:

1. Die Varianz der Störterme der Grundgesamtheit σ^2 : Die Varianzen der Koeffizienten $\widehat{\sigma}_{\widehat{\beta}_h}^2$ sind umso größer, je größer die Varianz der Grundgesamtheit σ^2 ist.
2. Die Streuung der x -Variablen ($= \sum_i (x_i - \bar{x})^2$): Die Varianzen der Koeffizienten $\widehat{\sigma}_{\widehat{\beta}_h}^2$ sind umso kleiner, je größer die Streuung der x -Variablen ist.
3. Die Anzahl der Beobachtungen n : Die Varianzen der Koeffizienten $\widehat{\sigma}_{\widehat{\beta}_h}^2$ nehmen mit n ab, da der Nenner mit n größer wird.
4. Die Korrelation (bzw. das Ausmaß linearer Abhängigkeit) zwischen den x -Variablen R_h^2 : Die Varianzen der Koeffizienten $\widehat{\sigma}_{\widehat{\beta}_h}^2$ nehmen ceteris paribus mit der Korrelation zwischen den x -Variablen (oder genauer: mit dem Bestimmtheitsmaß der Hilfsregression R_h^2) zu. Abbildung 7.4 zeigt diesen Zusammenhang.

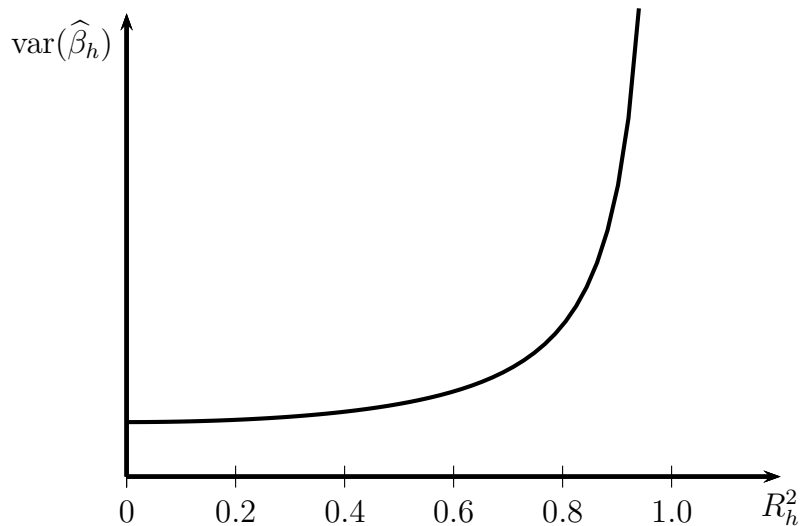


Abbildung 7.4: Multikollinearität: Die Varianz eines Koeffizienten $\hat{\beta}_h$, d.h. $\text{var}(\hat{\beta}_h)$, als Funktion des Bestimmtheitsmaßes einer Hilfsregression R_h^2

Wenn die x -Variablen untereinander völlig unkorreliert sind (d.h. wenn $R_h^2 = 0$) erhalten wir die gleichen Koeffizienten und Varianzen, die man bei getrennten bivariaten Regressionen von y auf die einzelnen x_h erhalten würde.

Wenn unter den x -Variablen eine lineare Abhängigkeit existiert ($R_h^2 \rightarrow 1$), können die Koeffizienten und Varianzen der Koeffizienten nicht berechnet werden, da der Nenner $(1 - R_h^2) \sum_i (x_{ih} - \bar{x}_{.k})$ Null wird.

Wenn die x -Variablen *nicht* exakt linear abhängig sind, aber hoch korreliert sind, wird die Varianz der Koeffizienten sehr groß, d.h. die Schätzungen werden sehr ‘ungenau’. Mit anderen Worten, mit zunehmender Multikollinearität nimmt die *Power* der Tests ab, die Nullhypothese kann häufig nicht mehr verworfen werden.

Abbildung 7.5 zeigt die Konfidenzintervalle und Konfidenzellipsen⁵ für $\hat{\beta}_2$ und $\hat{\beta}_3$ der Abbildung 7.2 zugrundeliegenden Daten. Offensichtlich werden die Konfidenzellipsen umso schmaler, je höher die Korrelation zwischen x_1 und x_2 ist – achten Sie auf die Achsenskalierung!!!

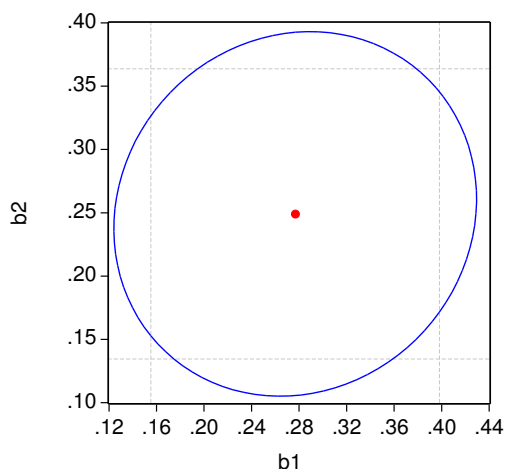
Wie schon eingangs erwähnt ist der OLS-Schätzer bei Multikollinearität auch in kleinen Stichproben erwartungstreu und effizient, also BLU (*best linear unbiased*).

Die geschätzten Standardfehler werden bei Multikollinearität zwar häufig sehr groß, aber sie sind ebenfalls unverzerrt. Deshalb bleiben die üblichen Tests gültig und können interpretiert werden. Das bedeutet, wenn bei Multikollinearität ein Koeffizient signifikant ist, kann er ganz normal interpretiert werden.

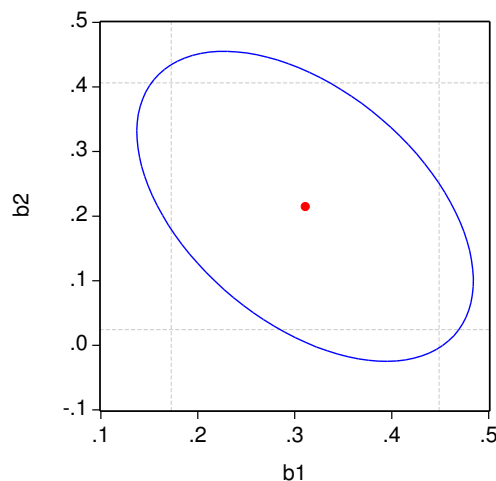
Selbst wenn einzelne Koeffizienten nicht signifikant von Null verschieden sind, so liefert ein F -Test auf *gemeinsame* Signifikanz multikollinearere Variablen häufig ein hoch signifikantes Ergebnis (vgl. Abbildung 7.5).

⁵Zur Erinnerung: die Fläche innerhalb der Konfidenzellipsen entspricht dem Annahmehereich der Nullhypothese, die Fläche außerhalb dem Verwerfungsbereich.

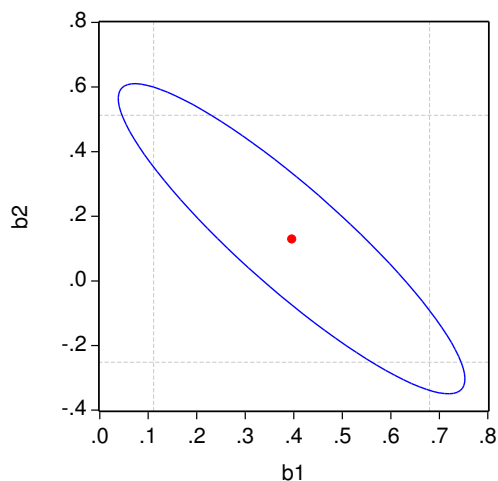
a) Keine Multikollinearität: $\rho_{x_1,x_2} = -0.08$



b) Mittlere Multikollinearität: $\rho_{x_1,x_2} = 0.5$



c) Hohe Multikollinearität: $\rho_{x_1,x_2} = 0.9$



d) Hohe Multikollinearität: $\rho_{x_1,x_2} = 0.99$

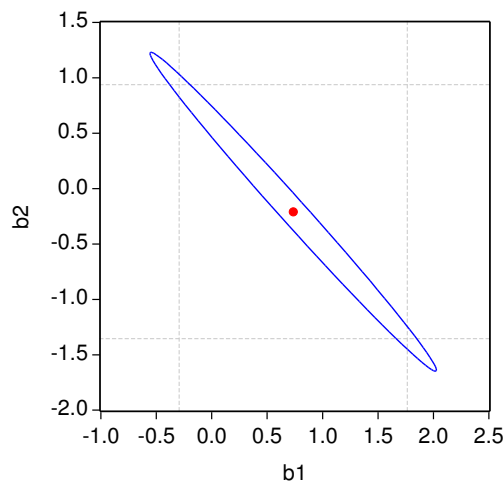


Abbildung 7.5: Multikollinearität und Konfidenzellipsen

Multikollinearität ist in erster Linie ein Problem, wenn das Ziel der Regression eine *Erklärung* ist. Wie man auch aus dem Venn-Diagramm (siehe Abbildung 7.1) erkennen kann, wird es bei Multikollinearität sehr schwierig den Einfluss einer einzelnen Variable zu isolieren, deshalb ist es schwer die Abhängigkeit von einzelnen Variablen zu erkennen, auch wenn die *gemeinsame* Abhängigkeit leicht messbar ist.

Multikollinearität ist in der Regel ein deutlich kleineres Problem, wenn das Ziel eine *Prognose* ist. Die Qualität einer Prognose wird durch Multikollinearität kaum betroffen, und selbstverständlich bleiben Prognosen auch bei Multikollinearität erwartungstreu!

Wie bereits früher erwähnt ähnelt das Problem der Multikollinearität dem Problem bei zu kleinen Stichproben, die Daten enthalten einfach zu wenig Informationen, um den Einfluss einzelner Variablen isolieren zu können. Deshalb stellt Multikollinearität eher ein Datenproblem als ein Schätzproblem dar, da die OLS-Schätzungen trotz Multikollinearität effizient sind.

Wenn ein Koeffizient signifikant ist stellt auch hohe Multikollinearität kein Problem

dar, der Koeffizient kann ganz normal interpretiert werden. Das Problem besteht einzig und allein darin, dass Koeffizienten bei hoher Multikollinearität häufig nicht signifikant sind, da die erklärenden Variablen aufgrund der hohen Korrelation untereinander häufig nicht genügend Informationen enthalten, was in den großen Standardfehlern zum Ausdruck kommt.

7.4 Erkennung von Multikollinearität

Einen statistischen Test zur Erkennung von Multikollinearität kann es ebenso wenig geben wie einen Test zur Erkennung einer ‘zu kleinen’ Stichprobe.

Aus dem bisher gesagten sollte aber eigentlich schon klar sein, wie Multikollinearität erkannt werden kann, deshalb hier nur noch eine kurze Aufzählung:

- Hohes R^2 und wenige signifikante Koeffizienten.
- Schätzergebnisse schwanken stark in Abhängigkeit von der gewählten Spezifikation und/oder bei Ausschluss einzelner Beobachtungen (wenn z.B. einzelne Koeffizienten bei Spezifikationsänderungen das Vorzeichen wechseln).
- Hohe paarweise Korrelationen zwischen Regressoren (z.B. größer 0.7 – 0.8). In der Korrelationsmatrix der x -Variablen sind aber nur bivariate Korrelationen erkennbar. Wenn es lineare Abhängigkeiten zwischen mehreren Variablen gibt, ist dies in der einfachen Korrelationsmatrix nicht mehr unbedingt erkennbar. Hilfsregressionen vermeiden diese Schwäche.
- Das Mittel der Wahl sind vermutlich **Hilfsregressionen** (vgl. Seite 8), bei denen die erklärenden Variablen aufeinander regressiert werden. Umso höher das R_h^2 dieser Hilfsregression ist, umso größer ist das Multikollinearitätsproblem. Es gibt keine feste Regel wie hoch das R_h^2 einer Hilfsregression maximal sein darf, aber als grobe Richtgröße können folgende Faustregeln gelten:
 - **Klien’s Faustregel:** Multikollinearität stellt ein ernstes ‘Problem’ dar, wenn das R_h^2 einer Hilfsregression größer ist als das R^2 der ursprünglichen Regression. Die Frage ist allerdings, welche Konsequenzen dies hat. Wenn der interessierende Standardfehler trotz Multikollinearität klein genug und der Koeffizient deshalb signifikant ist, dann stellt die Multikollinearität kein Problem dar. Wenn hingegen der Standardfehler groß und der Koeffizient deshalb *nicht* signifikant ist, dann ist guter Rat teuer.
 - Manche Computerprogramme (z.B. SPSS) geben für jeden Regressionskoeffizienten einen ‘**Variance Inflation Factor**’ (VIF) aus

$$\text{VIF}_h = \frac{1}{1 - R_h^2}$$

wobei R_h^2 wieder das Bestimmtheitsmaß der Hilfsregression (s.o.) ist. Welchen VIF man akzeptiert bleibt dem Benutzer überlassen, aber manchmal findet man in der Literatur Hinweise, dass man ab einem VIF größer 10 von einem Multikollinearitätsproblem sprechen könne. Natürlich ist die Frage, ab wann ein Multikollinearitätsproblem ‘zu groß’ ist, ebenso schwer zu beantworten wie die Frage, ab wann eine Stichprobe ‘zu klein’ ist.

- Ein etwas aufwändiger zu berechnender Indikator für das Vorliegen von Multikollinearität ist der **Condition Index** CI (bzw. *Condition Number*) von Belsley et al. (1980). Man berechnet dazu die Eigenwerte der Matrix $(\mathbf{X}'\mathbf{X})$; der CI ist dann die Wurzel aus dem Verhältnis vom größten zum kleinsten Eigenwert

$$\text{CI} = \sqrt{\frac{\text{maximaler Eigenwert von } (\mathbf{X}'\mathbf{X})}{\text{minimaler Eigenwert von } (\mathbf{X}'\mathbf{X})}}$$

Ein $\text{CI} > 20$ (30) ist ein Hinweis auf (starke) Multikollinearität.⁶ Der praktische Wert des Condition Index ist umstritten.

7.5 Ein Monte Carlo Beispiel

Das folgende Beispiel soll die Auswirkungen von Multikollinearität verdeutlichen. Wir generieren zuerst zwei normalverteilte erklärende Variablen x_2 und x_3 , die miteinander korreliert sind. Wir beginnen mit einem Korrelationskoeffizienten von 0.1 (d.h. $\text{corr}(x_2, x_3) = 0.1$) und erzeugen damit entsprechend der folgenden Gleichung die abhängige Variable y

$$y_i = 5 + 5x_{i2} + 5x_{i3} + \varepsilon_i \quad \text{mit} \quad \varepsilon_i \sim \text{N}(0, 100) \quad (7.1)$$

Diese Gleichung gibt das ‘wahre’ Modell wieder, und Spalte (1) in Tabelle 7.1 zeigt eine Schätzung (Realisation) dieses Modells. Spalten (2), (3) und (4) zeigen Schätzungen, wenn x_2 und x_3 stärker korreliert sind, nämlich Spalte (2) für $\text{corr}(x_2, x_3) = 0.6$, Spalte (3) und (4) für $\text{corr}(x_2, x_3) = 0.9$ und $\text{corr}(x_2, x_3) = 0.99$.

Wir sehen, dass die Schätzungen – insbesondere für $\hat{\beta}_3$ – mit zunehmender Korrelation ungenauer werden, und teilweise nicht mehr signifikant von Null verschieden sind.

Im täglichen Geschäft haben wir meist nur eine Realisation zur Verfügung, hier können wir dieses Gedankenexperiment aber beliebig oft wiederholen, also eine Monte Carlo Simulation durchführen. Konkret schätzen wir jede dieser vier Gleichungen tausend Mal, und erhalten somit tausend Schätzungen für $\hat{\beta}_2$ und $\hat{\beta}_3$.

Abbildung 7.6 zeigt die Boxplots dieser Schätzungen, wobei der Punkt den Mittelwert, der Strich in der Box den Median, und die ‘Box’ das obere bzw. untere Quartil zeigt. Auf der horizontalen Achse sind die Korrelationen zwischen x_2 und x_3 aufgetragen. Das linke Panel zeigt Realisationen von $\hat{\beta}_2$, das rechte Panel Realisationen von $\hat{\beta}_3$. Offensichtlich ist der OLS Schätzer auch bei hoher Multikollinearität erwartungstreu, aber wenig erstaunlich werden die Schätzungen mit zunehmender Multikollinearität ungenauer. Dieses Beispiel zeigt auch, wie sich die Streuung der erklärenden Variablen auf die Genauigkeit der Schätzung auswirkt; in diesem Beispiel ist die Streuung von x_2 viel kleiner als die Streuung von x_3 , deshalb wird x_2 viel ungenauer gemessen als x_3 .

⁶Achtung, es sind verschiedene Verfahren den CI zu berechnen gebräuchlich. Manchmal werden die x Variablen zentriert (d.h. der Mittelwert subtrahiert oder anderweitig skaliert), manchmal wird er mit nicht skalierten Variablen berechnet. In STATA sind mit dem ado-Befehl `collin` diagnostische Kennzahlen verfügbar. In EViews sind VIF's mit `eqname.varinf` verfügbar; in R im Packet ‘car’ mit `vif(eqname)`.

Tabelle 7.1: Vier Schätzungen für den datengenerierenden Prozess $y_i = 5 + 5x_{i2} + 5x_{i3} + \varepsilon_i$, $\varepsilon_i \sim N(0, 100)$, wobei x_2 und x_3 unterschiedlich stark korreliert sind (siehe Gleichung (7.1)). Man beachte, dass die Varianz von x_2 deutlich kleiner ist als die Varianz von x_3 ($\text{var}(x_2) = 1$, $\text{var}(x_3) = 100$), deshalb wird x_3 auch bei hoher Multikollinearität genauer geschätzt.

	(1)	(2)	(3)	(4)
Dep. Var.	y	y	y	y
Const.	4.708* (1.767)	2.763 (1.395)	7.339*** (1.469)	4.576*** (1.255)
x_2	3.380* (1.616)	4.996** (1.741)	11.654** (3.620)	-16.996 (8.989)
x_3	4.958*** (0.200)	5.251*** (0.175)	4.424*** (0.373)	7.173*** (0.899)
R^2	0.932	0.971	0.971	0.977
F	320.060	794.218	796.467	976.512
p -value	0.000	0.000	0.000	0.000
n	50	50	50	50
$\text{corr}(x_2, x_3)$	0.1	0.6	0.9	0.99

Daraus wird auch eine große Gefahr von Multikollinearität ersichtlich, aufgrund der großen Streuung der Schätzungen haben ‘Datenschürfer’ leichtes Spiel, und auch charakterstarke Ökonometrikerinnen werden bei der Auswahl der zu publizierenden Ergebnisse in Versuchung geführt.

7.6 Maßnahmen bei Multikollinearität

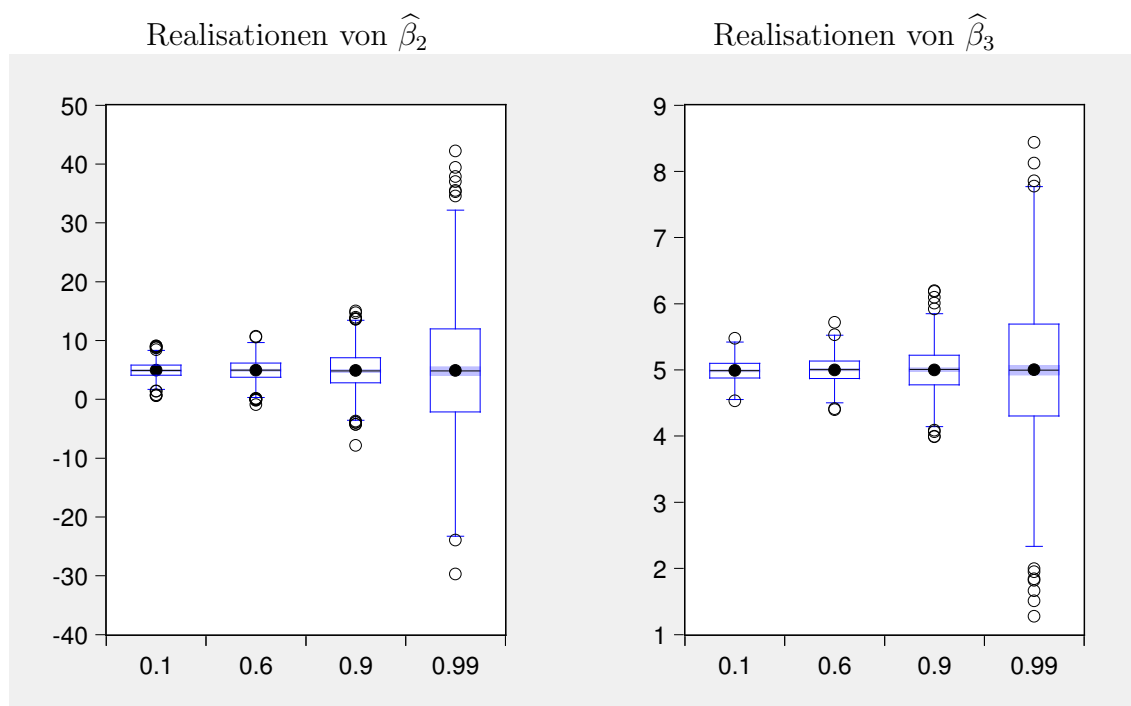
Wie bereits erwähnt stellt Multikollinearität für Prognosen kaum ein Problem dar. Wenn das Ziel ausschließlich eine Prognose ist – und angenommen werden kann, dass die Korrelation auch in der Zukunft in der gleichen Form erhalten bleiben wird – sind in der Regel keine Maßnahmen erforderlich.

Etwas schwieriger ist die Situation, wenn das Ziel darin besteht, den Einfluss einer erklärenden Variable auf die abhängige Variable zu isolieren (*Erklärung*, Strukturanalysen). Wie schon mehrfach erwähnt ist der OLS-Schätzer auch in diesem Fall effizient, aber die *Power* ist sehr niedrig.

In der Literatur werden für diesen Fall verschiedene Maßnahmen vorgeschlagen:

Zusätzliche Daten erheben: dies ist für ökonomische Datensätze selten möglich, und wenn die Korrelation in der Grundgesamtheit besteht auch nicht sehr zielführend, da wir nur mehr vom Gleichen erhalten.

Zusätzliche externe Information erheben und verwerten: zum Beispiel Zeitreihen- und Querschnittsdaten kombinieren, oder Berücksichtigung theoretischer Restriktionen.



horizontale Achse: Korrelation zwischen x_2 und x_3

Abbildung 7.6: Tausend Schätzungen der SRF $y = \hat{\beta}_1 + \hat{\beta}_2x_2 + \hat{\beta}_3x_3 + \hat{\varepsilon}$ der PRF $y = 5 + 5x_2 + 5x_3 + \varepsilon$ (‘wahres’ Modell) mit unterschiedlichen Korrelationen zwischen x_2 und x_3 ($n = 50$). Man beachte, dass die Streuung von x_2 deutlich kleiner ist als die Streuung von x_3 ($\text{var}(x_2) = 1, \text{var}(x_3) = 100$), deshalb wird der Koeffizient von x_2 deutlich *ungenauer* gemessen (beachten Sie die Skala)! Auch bei hoher Multikollinearität sind die Koeffizienten erwartungstreu.

Transformation der Variablen: zum Beispiel Bildung erster Differenzen bei trendbehafteten Zeitreihen. Mehr dazu später.

Variablen “weglassen”: dies ist eine sehr verbreitete, allerdings auch sehr gefährliche Methode. Für Prognosen bringt das ‘Weglassen’ von Variablen kaum Vorteile, aber wenn es darum geht den Einfluss einer Variable zu isolieren, kann das ‘Weglassen’ zu völlig falschen Schlussfolgerungen und Fehlinterpretationen führen (*‘omitted variable bias’*). Ein einfaches Beispiel soll dies illustrieren: angenommen, sie sollten testen, ob ein gewisses Mundspülwasser Krebs erzeugt, und Sie wissen, dass dieses Mundspülwasser sehr häufig von Rauchern verwendet wird. Das bedeutet, Rauchen und der Gebrauch dieses Mundspülwassers werden hoch korreliert sein, deshalb es ist nicht unwahrscheinlich, dass in einer Regression mit den erklärenden Variablen ‘Rauchen’ und ‘Verwendung des Mundspülwassers’ zumindest eine der Variablen nicht signifikant von Null verschieden ist. Wenn Sie sich nun entschließen die Variable ‘Rauchen’ wegzulassen, erhalten Sie vermutlich einen signifikanten Wert für die krebserregende Wirkung des Mundspülwassers, weil Sie die krebserregende Wirkung des Rauchens irrtümlich dem Mundspülwasser zurechnen. Ob das Mundspülwasser tatsächlich eine krebserregende Wirkung hat, können Sie durch das Weglassen der Variable Rauchen ganz sicher nicht besser erkennen, dies verleitet nur zu

falschen Schlussfolgerungen.

Wir haben bereits festgestellt, dass der OLS-Schätzer auch im Fall von hoher Multikollinearität unverzerrt und effizient ist. Dagegen werden wir in dem Kapitel über *Spezifikation* zeigen, dass der OLS-Schätzer im Fall *fehlender relevanter Variablen* (*‘omitted variables’*) weder erwartungstreu noch konsistent ist! Daher scheint es wenig Gründe für das Weglassen von multikollinearen Variablen zu geben.

Wenn das ‘Weglassen’ trotzdem häufig geschieht, dann vermutlich nicht zuletzt deshalb, weil individuell nicht signifikante Koeffizienten häufig ein Problem für den verwöhnten Geschmack von Referees sind, die an signifikante Koeffizienten gewöhnt sind. Besser wäre es in solchen Fällen oft mit einem *F*-Test auf die gemeinsame Signifikanz der Variablen zu testen.

Allerdings sei auch angemerkt, dass das Weglassen von multikollinearen Variablen vermutlich kein allzu großes Problem darstellt, solange davon nicht die eigentlich interessierenden Variablen betroffen sind, sondern nur Variablen, die ausschließlich einbezogen wurden, um für weniger interessierende Zusammenhänge zu kontrollieren.

Ridge Regression Das ‘Problem’ bei hoher Multikollinearität besteht hauptsächlich in der niedrigen Power der Tests, verursacht durch die großen Standardfehler der Koeffizienten. Vor allem in der statistischen Literatur wird für solche Fälle manchmal ein Schätzer empfohlen, der eine ‘Stabilisierung’ der geschätzten Koeffizienten erreichen soll. Konkret wird für die Ridge Regression folgende Zielfunktion minimiert

$$\sum_i (y_i - \mathbf{x}'_i \boldsymbol{\beta})^2 + \lambda \sum_{j=1}^k \beta_j^2$$

was zu folgender Lösung führt

$$\hat{\boldsymbol{\beta}}^{\text{ridge}} = (\mathbf{X}'\mathbf{X} - \lambda \mathbf{I})^{-1} \mathbf{X}'\mathbf{y}$$

Man beachte die Ähnlichkeit mit dem OLS Schätzer, allerdings werden die Hauptdiagonalelemente von $\mathbf{X}'\mathbf{X}$ um den Betrag λ reduziert.

Die damit geschätzten Ridge Koeffizienten sind in der Regel verzerrt, und die Wahl eines geeigneten λ ist alles andere als trivial (häufig wird dazu eine Kreuzvalidierung empfohlen).

In der klassischen theoriegetriebenen Ökonometrie sind diese Schätzer etwas außer Mode gekommen und werden eher selten angewandt.

Häufiger werden solche Methoden für eine Variablenselektion in hochdimensionalen Modellen und *data-mining* angewandt, für eine Übersicht siehe Belloni et al. (2014).

Sich damit abfinden – manchmal nicht die schlechteste Lösung!

Literaturverzeichnis

Belloni, A., Chernozhukov, V. and Hansen, C. (2014), ‘High-dimensional methods and inference on structural and treatment effects’, *Journal of Economic Perspectives* **28**(2), 29–50.

URL: <http://www.aeaweb.org/articles?id=10.1257/jep.28.2.29>

Belsley, D. A., Kuh, E. and Welsch, R. E. (1980), *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity (Wiley Series in Probability and Statistics)*, 1 edn, Wiley-Interscience.

Goldberger, A. S. (1991), *A Course in Econometrics*, Harvard University Press.

Greene, W. H. (2007), *Econometric Analysis*, 6th edn, Prentice Hall.