

Oscar A. G. Treyer

Business Forecasting

Anwendungsorientierte Theorie quantitativer Prognoseverfahren

Haupt Verlag
Bern · Stuttgart · Wien

Oscar A. G. Treyer, Dr. oec. publ. et dipl. Handelslehrer,

studierte Betriebswirtschaftslehre, Finanzwissenschaft und Volkswirtschaftslehre an der Universität Zürich und promovierte 1985 zum Doktor der Wirtschaftswissenschaft aufgrund der Dissertation 'Amerikanisches Financial Accounting (Theorien und Methoden)'.

Wirtschaftspraxis: Tätigkeiten bei der Realisierung von Management-Informationssystemen, Restrukturierungen und Finanzbeschaffungen als Abteilungsleiter Control/EDV eines Unternehmens der Baustoff-Branche, Geschäftsführer des Bereichs Finanz/Control einer schweizerischen Informatik-Gruppe, Gruppenbereichsleiter Gruppen-Controlling einer international, europaweit führenden Unternehmensgruppe der Sanitärtechnik, operativer Finanzchef und Mitglied der Gruppenleitung eines internationalen Casino-Unternehmens sowie selbstständiger Berater.

Lehrtätigkeit: Lehrbeauftragter an der Universität Zürich, Dozent an der Kammerschule Zürich, Professor an der FHBB Fachhochschule beider Basel Nordwestschweiz, Dozent an der Schweizerischen Akademie für Wirtschaftsprüfung in Zürich, Kursleiter im CFA® Exam Preparation Course in Zug, Approved Tutor für das MBA-Programm der Edinburgh Business School (Harrow Watt University), Modulleiter und Referent im MAS Corporate Finance CFO der Fachhochschule Nordwestschweiz (Hochschule für Wirtschaft). Seit 2004: Ständiger Dozent für Accounting an der Universität St. Gallen.

Publikationen: Diverse Fachartikel in der Schweiz und Deutschland auf dem Gebiet des Management und Financial Accounting sowie Business Forecasting. Lehrbuch zu 'Business Statistik'.



1. Auflage 2010

Bibliografische Information der *Deutschen Nationalbibliothek*
Die *Deutsche Nationalbibliothek* verzeichnet diese Publikation
in der Deutschen Nationalbibliografie; detaillierte bibliografische
Angaben sind im Internet über <http://dnb.d-nb.de> abrufbar.

ISBN: 978-3-8252-3365-5

Alle Rechte vorbehalten.

Copyright © 2010 by Haupt, Berne

Jede Art der Vervielfältigung ohne Genehmigung des Verlages ist unzulässig.

Einbandgestaltung: Atelier Reichert, Stuttgart

Satz: Verlag Die Werkstatt, Göttingen

Printed in Germany

www.haupt.ch

UTB-Bestellnummer: 978-3-8252-3365-5

Bei der Berechnung dieses p-Wertes setzt man die berechnete Prüfgröße bzw. Teststatistikprüfgröße als Kritischen Wert ein, und liest aus der entsprechenden Tabelle die «Restwahrscheinlichkeit» heraus, d.h. die Wahrscheinlichkeit, bei welcher dieser neue Kritische Wert gerade erreicht und überschritten (bzw. bei negativen Werten unterschritten) wird.

6.6 Korrelationsanalyse

In den Kapiteln 6.4 und 6.5 ging es darum, Rückschlüsse auf Grundgesamtheitsparameter anhand von Stichprobenparametern vorzunehmen. In den folgenden Ausführungen verlagert sich der Fokus auf die Fragestellung, wie ein möglicher Zusammenhang zwischen zwei (oder mehreren Variablen) untersucht werden kann. Dabei bedient man sich einer mathematischen Gleichung, die es erlaubt, eine Variable aufgrund von anderen zu schätzen. Es geht also darum, statistisch einen Zusammenhang zwischen Ursache 'X' und Wirkung 'Y' aufzuzeigen. Mit der Korrelationsanalyse¹⁵⁴ bezeichnet man Methoden und Techniken, welche die «Güte» der Beziehung bzw. den Zusammenhang von Variablen misst.

6.6.1 Überblick

Damit überhaupt ersichtlich ist, ob ein Zusammenhang besteht und wie groß dieser ist, bedient man sich in einem ersten Schritt des Hilfsmittels 'Streudiagramm'¹⁵⁵. Dabei ist es üblich, dass auf der Abszisse (Horizontalen) – auch mit 'X' bezeichnet – die unabhängige Variable aufgezeichnet wird und auf der Ordinate (Vertikalen) – auch mit 'Y' bezeichnet – die abhängige Variable.

Die in Abbildung 16 aufgeführten Streudiagramme¹⁵⁶ sollen beispielhaft verschiedenen «starke» sowie «positive» und «negative» Korrelationen bzw. Zusammenhänge aufzeigen.

154 correlation analysis

155 scatterplot

156 vgl. auch (Moore & McCabe, 1998, S. 130 – Figure 2.11)

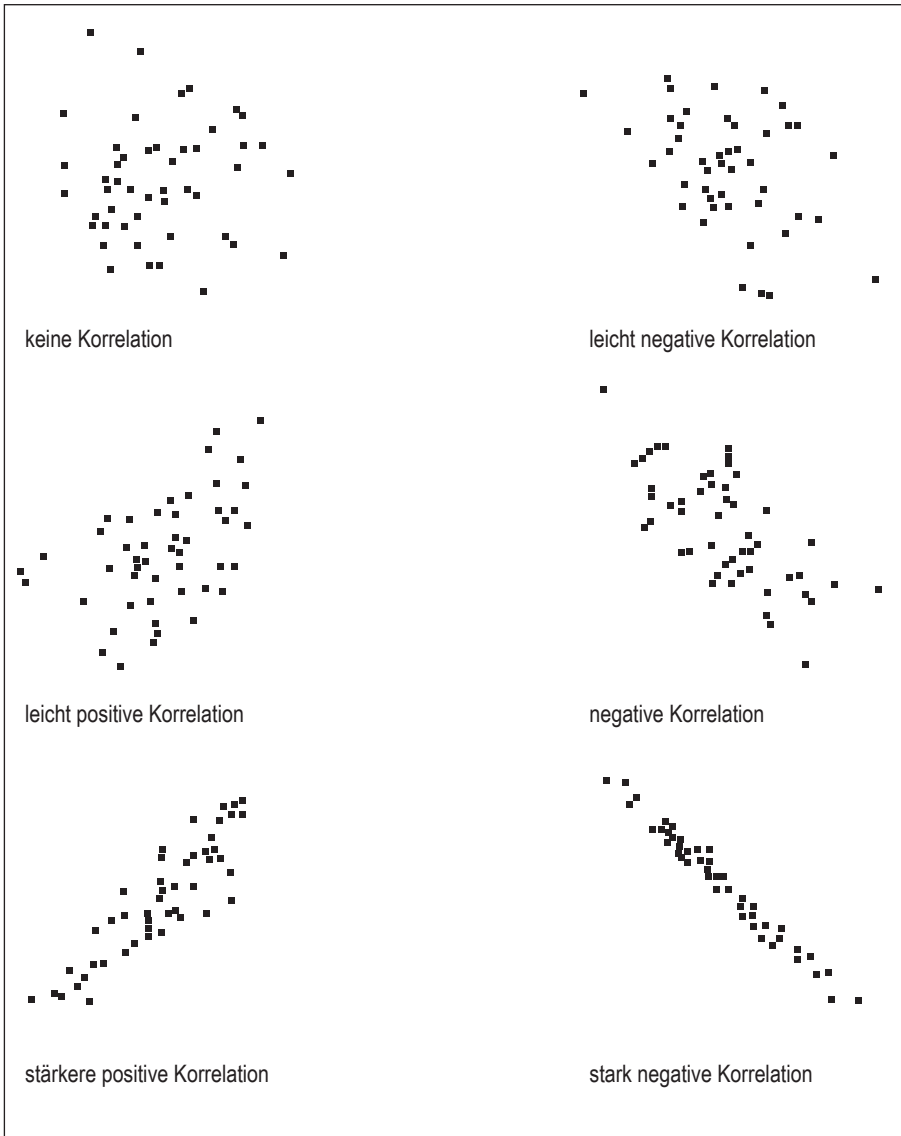


Abbildung 16: Beispiele unterschiedlicher, positiver und negativer Korrelationsstärken

Dabei ist zu beachten, dass beim Vorhandensein einer starken Korrelation zwischen zwei Variablen dies lediglich heißt, dass eine starke «Beziehung» zwischen den beiden Variablen besteht, nicht aber, dass dadurch ein Ursache-Wirkung-Zusammenhang hergeleitet bzw. «bewiesen» ist. Letzterer ist allein durch

eine ökonomisch logische Interpretation möglich, nicht aber durch statistische Kennzahlen.

6.6.2 Kennzahlen der Korrelation

Üblicherweise wird unterstellt, dass zwischen der unabhängigen Variablen 'X' und der abhängigen Variablen 'Y' ein linearer Zusammenhang besteht.¹⁵⁷

Wie in Abbildung 16 zu sehen ist, wird eine «starke» Korrelation durch eine kleine Streuung der beobachteten Werte um die unterstellte, theoretische Gerade gekennzeichnet. In Anlehnung an die Kennzahl 'Standardabweichung' bzw. 'Varianz' wurden Kennzahlen bezüglich Korrelation «kreiert», die folgende drei Eigenschaften aufweisen:

- Die Kennzahl kann die Werte zwischen minus '1' und plus '1' annehmen.
- Eine perfekte negative Korrelation weist den Wert '-1' auf, eine perfekte positive Korrelation den Wert '+1'.
- Keine Korrelation wird durch den Wert '0' wiedergegeben, d.h., alle Werte in der Nähe von '0' kennzeichnen eine schwache Korrelation.

Korrelationskoeffizient

Die wohl bekannteste Kennzahl der Korrelation ist der Korrelationskoeffizient nach Pearson.^{158,159} Dieser Produkt-Moment-Korrelationskoeffizient¹⁶⁰ wird nach Karl Pearson (1857-1936) benannt, obwohl nicht Letzterer, sondern Sir Francis Galton (1822–1911) diesen im Jahre 1888 «einführte».¹⁶¹

157 vgl. hierzu Kapitel 17.2.1.2 im Anhang

Falls kein linearer Zusammenhang festgestellt wird, kann durch Logarithmieren z.B. eine geometrische Funktion und eine Exponentialfunktion in eine «linearisierte» Form übergeleitet werden.

158 coefficient of correlation

159 Beim Korrelationskoeffizient nach Pearson wird unterstellt, dass ein linearer Zusammenhang zwischen 'x' und 'y' besteht, d.h., es wird das Vorhandensein einer (Regressions-)Gerade vorausgesetzt. Falls dem nicht so ist bzw. Ungewissheit darüber besteht, ist es besser, den Rangkorrelationskoeffizienten nach Spearman zu verwenden.

160 product moment correlation coefficient

161 vgl. hierzu auch Fußnote 97

Die Berechnung des Korrelationskoeffizienten erfolgt nach Formel 15.

Formel 15: Korrelationskoeffizient nach Pearson

$$r = \frac{\sum [(X - \bar{X}) \cdot (Y - \bar{Y})]}{\sqrt{\sum (X - \bar{X})^2 \cdot \sum (Y - \bar{Y})^2}} = \frac{\text{cov}_{X,Y}}{s_X \cdot s_Y}$$

r steht für den Korrelationskoeffizienten

X ist der jeweilige X-Wert eines beobachteten Wertepaares

\bar{X} ist das arithmetische Mittel aller X-Werte

Y ist der jeweilige Y-Wert eines beobachteten Wertepaares

\bar{Y} ist das arithmetische Mittel aller Y-Werte

s_X Standardabweichung der X-Werte bzw. $s_X = \sqrt{\frac{\sum (X - \bar{X})^2}{n - 1}}$

s_Y Standardabweichung der Y-Werte bzw. $s_Y = \sqrt{\frac{\sum (Y - \bar{Y})^2}{n - 1}}$

$\text{cov}_{X,Y}$ Kovarianz bzw. $\text{cov}_{X,Y} = \frac{\sum [(X - \bar{X}) \cdot (Y - \bar{Y})]}{n - 1}$

Der Ausdruck des Zählers von 'r' gleicht der Kovarianz¹⁶² – mit Ausnahme der Division durch 'n – 1'. Die Kovarianz gibt Auskunft darüber, ob sich tendenziell mit der Änderung von 'X' auch 'Y' ändert, bzw. ob im Durchschnitt zu größeren (kleineren) X-Werten größere (kleinere) Y-Werte gehören. Die Kovarianz wird vor allem auf dem Gebiet des «Corporate Finance» und im «Portfolio Management» verwendet.

Determinationskoeffizient

Der oben erwähnte Korrelationskoeffizient zeigt aber noch nicht auf, wie groß der Anteil der Streuung von 'Y' durch die Regressionsgeraden «bestimmt» oder «erklärt» wird. Diese Eigenschaft zeigt erst das Quadrat des Korrelationskoeffizienten nach Pearson, nämlich der sogenannte Determinationskoeffizient¹⁶³. Letzterer berechnet sich entsprechend der Formel 16.

¹⁶² covariance

¹⁶³ coefficient of determination

Der Determinationskoeffizient wird auch 'Bestimmtheitsmaß' genannt.

Was genau der Determinationskoeffizient aussagt, soll in Abbildung 17 verdeutlicht werden. Formel 16 zeigt die dazugehörige Berechnungsformel.¹⁶⁴

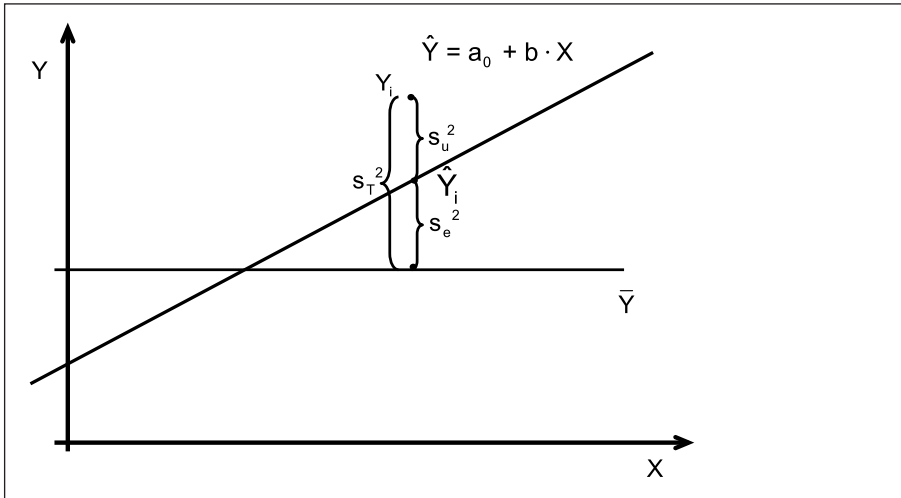


Abbildung 17: Aussagekraft des Determinationskoeffizienten

Formel 16: Determinationskoeffizient

$$r^2 = \frac{s_e^2}{s_T^2} = \frac{\text{erklärte Varianz}_{\text{von } Y}}{\text{Varianz Total}_{\text{von } Y}} = \frac{\frac{1}{n-1} \cdot \sum (\hat{Y} - \bar{Y})^2}{\frac{1}{n-1} \cdot \sum (Y - \bar{Y})^2}$$

$$\Rightarrow \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}}$$

wobei $s_T^2 = s_e^2 + s_u^2 \Rightarrow \text{Varianz total} = \text{erklärte Varianz} + \text{unerklärte Varianz}$

SSR (Sum of Squared Regressions)

ist die Summe der quadrierten Regressionsabweichungen bzw. $\sum (\hat{Y} - \bar{Y})^2$

SSE (Sum of Squared Errors or residuals)

ist die Summe der quadrierten Residuen(abweichungen) bzw. $\sum (Y - \hat{Y})^2$

SST (Sum of Squared Totals) ist die Summe der quadrierten Totalabweichungen bzw. $\sum (Y - \bar{Y})^2$

Wie aus Abbildung 17 und Formel 16¹⁶⁵ ersichtlich ist, misst der Determinationskoeffizient, welcher (Prozent-)Anteil der Streuung anhand der unterstellten (Re-

¹⁶⁴ vgl. hierzu auch Kapitel 17.2.1.2 im Anhang

¹⁶⁵ Die Herleitung von Formel 16 auf der Basis des quadrierten Korrelationskoeffizienten nach Pearson (vgl. Formel 15) kann z. B. bei (Bohley, 1996, S. 246) nachgesehen werden.

gressions-)Geraden¹⁶⁶ «bestimmt» ist. Die Ergänzung auf '1' bzw. 100 % zeigt dementsprechend auf, welcher (Prozent-)Anteil der Streuung **nicht** «erklärt» wird. Dabei ist wichtig zu wissen, dass die «Basis» der Messung des (Prozent-)Anteils der Streuung jeweils das arithmetische Mittel von 'Y' ist. Die Begründung dafür ist, dass jede «effiziente» Prognoseschätzung (bzw. die (Regressions-)Gerade) die arithmetischen Mittel von 'Y' und 'X' (bzw. \bar{Y} und \bar{X}) enthalten muss.

Aus den Ausführungen ergibt sich, dass der Korrelationskoeffizient 'r' eine sehr gute Maßzahl für die «Güte» der Beziehung zwischen zwei Variablen darstellt. Wenn es aber darum geht, über den prozentualen Anteil der Korrelation etwas auszusagen, ist es korrekter, den Determinationskoeffizienten 'r²' – der immer einen geringeren absoluten Wert aufweist – zu verwenden.

Rangkorrelationskoeffizient

Falls kein linearer Zusammenhang zwischen 'X' und 'Y' besteht, bzw. wenn man nicht sicher ist, dass ein linearer Zusammenhang effektiv vorhanden ist, respektive eine nicht parametrische Beziehung unterstellt wird, ist es nicht korrekt, den Korrelationskoeffizienten nach Pearson zu verwenden. In all diesen Fällen ist es besser, den Rangkorrelationskoeffizienten¹⁶⁷ – nach Charles Spearman (1863–1945) benannt – einzusetzen. Letzterer wird mit 'ρ' [rho] bezeichnet und berechnet sich nicht anhand der beobachteten Werte, sondern anhand ihrer Rangordnungen. Konkret werden sowohl alle 'X'-Werte als auch alle 'Y'-Werte (unabhängig voneinander) in eine absolute, ganzzahlige Rangnummer umgewandelt. Die 'XY'-Beziehung wird dabei nicht verändert! Erst anhand dieser beiden Rangordnungen wird nun der Korrelationskoeffizient (nach Pearson) berechnet.

166 vgl. auch Fußnote 159

167 Spearman's rank correlation coefficient

De facto ist der Rangkorrelationskoeffizient nach Spearman ein Spezialfall des Produkt-Moment-Korrelationskoeffizienten nach Pearson.

rechts, noch oberhalb, noch unterhalb des «wahren» Grundgesamtheitsparameters aufweisen. Es geht hier nicht primär darum, den genauen Grundgesamtheitsparameter zu ermitteln, sondern dass sich dieser (tendenziell) in der Mitte der Parameterschätzungen befindet.

- **Konsistenz:** Eine gute Schätzung sollte konsistent sein. Die Parameterschätzung sollte mit zunehmendem Stichprobenumfang immer näher zum «wahren» Grundgesamtheitsparameter hin tendieren, bzw. die Streuung der Parameterschätzung sollte immer kleiner werden.
- **Kleinste Varianz:** Eine gute Schätzung – unabhängig von der Größe des Stichprobenumfanges – die kleinstmögliche Streuung um den «wahren» Grundgesamtheitsparameter aufweisen.

Werden die beiden Eigenschaften 'Unverzerrtheit' bzw. 'Erwartungstreue' und 'Kleinste Varianz' von einer Parameterschätzung erfüllt, so wird dies als 'Effizienz'⁵¹¹ bezeichnet.

17.2 Prognoseschätzung

Im Gegensatz zur Parameterschätzung (vgl. Kapitel 17.1 im Anhang), bei der es de facto um eine Punktschätzung statistischer Kennzahlen einer Grundgesamtheit geht, versucht man bei der Prognoseschätzung – unter Zuhilfenahme der Überlegungen der Parameterschätzung – eine Vorhersage eines Zusammenhangs bzw. einer Entwicklung zu ermitteln. Diese kann in einem Trial-and-Error-Verfahren gefunden werden, was aber eventuell sehr umständlich sein kann. Aus diesem Grund verwendet man hierzu zwei Hauptmethoden, nämlich die Methode der kleinsten Quadrate sowie die Maximum-Likelihood-Methode. Diese Methoden werden in den Kapiteln 17.2.1 und 17.2.2 im Anhang genauer umschrieben.

17.2.1 Methode der kleinsten Quadrate

17.2.1.1 Einleitung

Die Methode der kleinsten Quadrate⁵¹² kann als eine Methode der Datenanpassung ausgelegt werden. Die beste Anpassung im Kontext der kleinsten Quadrate weist dasjenige Modell auf, bei welchem die Summe der quadrierten Residu-

⁵¹¹ Bei linearen Schätzungen wird dies auch als 'BLUE' [Best Linear Unbiased Estimator] bezeichnet.

vgl. hierzu auch (Urban & Mayerl, 2008, S. 120ff)

⁵¹² method of least squares or least-squares method

en(abweichungen)^{513,514} (SSE) den kleinsten Wert aufweist. Diese Methode wurde von Johann Carl Friedrich Gauß (1777–1855) um 1794 zum ersten Mal umschrieben.⁵¹⁵

Die konkrete Vorgehensweise bei der Anwendung der Methode-der-kleinsten-Quadrate ist in Abbildung 103 schematisch aufgeführt. Dabei ist ersichtlich, dass die Methode-der-kleinsten-Quadrate ganz konkrete Modellvoraussetzungen bei der Grundgesamtheit unterstellt, damit die Parameterschätzungen aufgrund der Stichprobe ihre Gültigkeit haben (vgl. hierzu auch Kapitel 17.2.1.4 im Anhang).

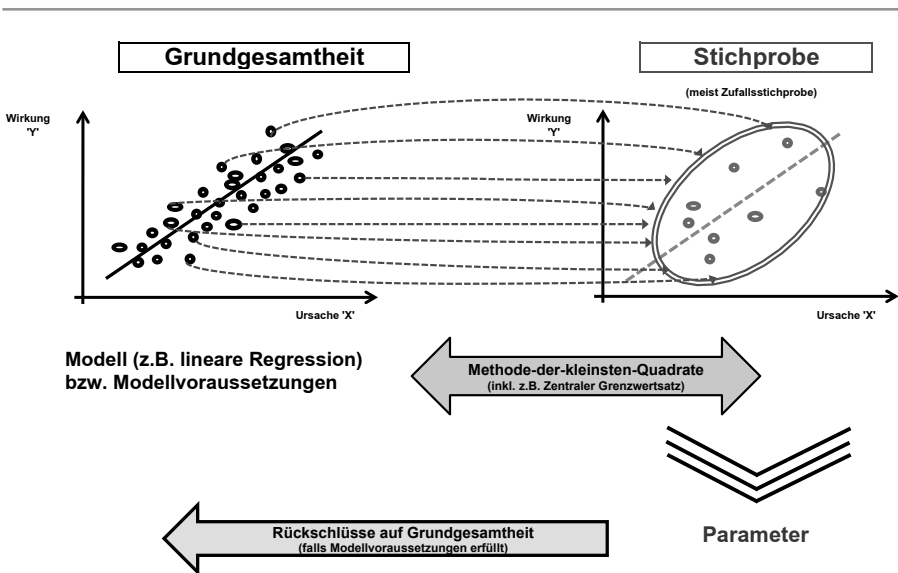


Abbildung 103: Vorgehensstruktur der Methode-der-kleinsten-Quadrate

513 residuals

Die Residuen(abweichungen) 'e' sind die jeweilige Differenz zwischen den effektiv gemessenen 'Y'-Werten und den diesbezüglichen (Prognose-)Werten 'Ŷ'.

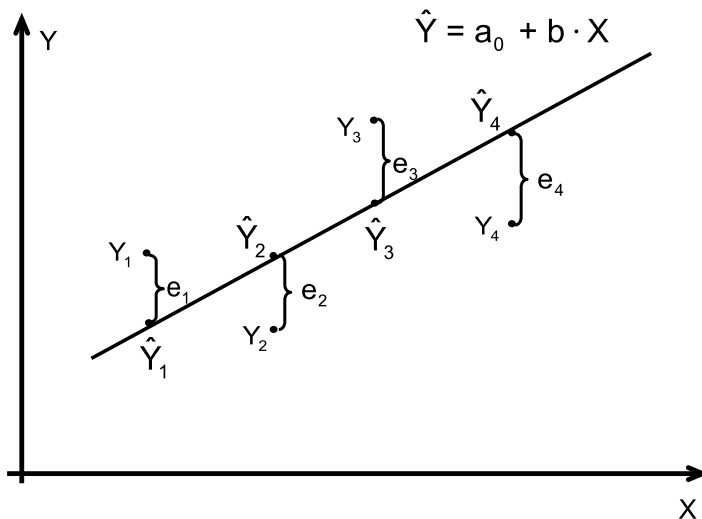
514 Sum of Squared Errors or residuals (SSE)

vgl. Formel 24

515 Falls die Residuen(abweichungen) normal verteilt sind, entspricht die Methode-der-kleinsten-Quadrate zugleich auch der Maximum-Likelihood-Methode (vgl. auch Kapitel 17.2.2 im Anhang).

Im Gegensatz zur Freihandmethode⁵¹⁶, Extremwertmethode⁵¹⁷ bzw. Schichtkostenverfahren⁵¹⁸ und Mittelwertmethode weist die Methode-der-kleinsten-Quadrate⁵¹⁹ eine klare Berechnungsgrundlage auf. Bei der Methode-der-kleinsten-Quadrate werden die vertikalen Abstandsquadrate zwischen den beobachteten 'Y'-Werten und den diesbezüglichen Werten auf der Regressionsgeraden minimiert.⁵²⁰

Abbildung 104 illustriert die Vorgehensweise der Berechnung der Methode-der-kleinsten-Quadrate.



Methode-der-kleinsten-Quadrate:

$$= \sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2 \Rightarrow \text{minimieren!}$$

Abbildung 104: Berechnungsvorgehensweise der Methode-der-kleinsten-Quadrate

516 scatterplot method

vgl. z. B. (Däumler & Grabe, 1996, S. 67f)

517 high-low method

vgl. z. B. (Horngren, Datar, Foster, Rajan, & Ittner, 2008, S. 372f)

518 vgl. z. B. (Nadig, 2000, S. 134ff)

519 least-squares method or normal least squares (NLS) or ordinary least squares (OLS)

520 Diese vertikalen Abstände werden auch als Residuen(abweichungen) bzw. Störgrößen 'e' bezeichnet.

Der Nachteil dieser Methode besteht darin, dass sogenannte «Ausreißer»⁵²¹ durch die Quadrierung ein überproportionales Gewicht bei der Berechnung erhalten. Aus diesem Grund ist es in den meisten Statistiksoftwarepaketen möglich, dass «große» Ausreißer aus der Berechnung ausgeschlossen werden können.

17.2.1.2 Lineare Regressionsfunktionen

Wird eine Gerade unterstellt, lautet diese «einfache» lineare Regressionsfunktion wie in Formel 57 aufgeführt.

Formel 57: Einfache lineare Regressionsfunktion bzw. Geradengleichung⁵²²

$$\hat{Y} = a_0 + b \cdot X$$

\hat{Y} steht für den geschätzten Wert der abhängigen Y-Variable
bezüglich eines gewählten X-Wertes

a_0 ist der Achsenabschnitt der linearen Gleichung bei einem X-Wert von '0'

b ist die Steigung (auch Regressionskoeffizient genannt) der Geraden,

d.h. wie viel sich durchschnittlich der geschätzte Wert von \hat{Y} durch eine Veränderung einer Einheit von X verändert

X ist jeder gewählte Wert der unabhängigen Variablen 'X'

Damit nun die beiden Parameter, nämlich der Achsenabschnitt⁵²³ ' a_0 ' und die Steigung⁵²⁴ bzw. der Regressionskoeffizient⁵²⁵ ' b ', bei einem beobachteten Datensatz an X- und Y-Werten nach der Methode-der-kleinsten-Quadrate berechnet werden können, sind die beiden in Formel 58 aufgeführten Gleichungen simultan nach den beiden Parametern aufzulösen.⁵²⁶

521 outlier

522 Die beobachteten Y-Werte können somit auch wie folgt bestimmt werden:

$$Y = a_0 + b \cdot X + e$$

Y steht für die beobachteten abhängigen Y-Variablen

e steht für die Residuen(abweichungen) bzw. Störgrößen

523 intercept

524 slope

525 regression coefficient

526 vgl. (Spiegel, Schaum's Outline of Theory and Problems of Statistics, 1992, S. 267) und (Bohley, 1996, S. 212f)

Formel 58: Normalgleichungen der Methode-der-kleinsten-Quadrate

$$\sum Y = a_0 \cdot n + b \cdot \sum X$$

$$\sum Y \cdot X = a_0 \cdot \sum X + b \cdot \sum X^2$$

n steht für die Anzahl der beobachteten Wertepaare 'X' und 'Y'

Die Auflösung der beiden Gleichungen ergibt die in Formel 59 aufgeführten (direkten) Berechnungen der beiden Parameter.

Formel 59: Parameterberechnung der einfachen linearen Regressionsfunktion nach der Methode-der-kleinsten-Quadrate

$$b = \frac{\sum [(X - \bar{X}) \cdot (Y - \bar{Y})]}{\sum (X - \bar{X})^2}$$

oder

$$b = \frac{n \cdot (\sum X \cdot Y) - (\sum X) \cdot (\sum Y)}{n \cdot (\sum X^2) - (\sum X)^2}$$

oder

$$b = \frac{\text{COV}_{X,Y}}{s_X^2}$$

$$a_0 = \bar{Y} - b \cdot \bar{X}$$

\bar{X} steht für das arithmetische Mittel der X-Werte

\bar{Y} steht für das arithmetische Mittel der Y-Werte

$$s_X \quad \text{Standardabweichung der X-Werte bzw. } s_X = \sqrt{\frac{\sum (X - \bar{X})^2}{n - 1}}$$

$$\text{cov}_{X,Y} \quad \text{Kovarianz bzw. } \text{cov}_{X,Y} = \frac{\sum [(X - \bar{X}) \cdot (Y - \bar{Y})]}{n - 1}$$

Wichtig zu erwähnen ist auch, dass die lineare Regressionsfunktion, welche nach der Methode-der-kleinsten-Quadrate ermittelt wurde, durch das arithmetische Mittel der (jeweiligen) X-Werte sowie das arithmetische Mittel der Y-Werte verläuft.

Wird nun dieser lineare Zusammenhang auf mehr als eine unabhängige Variable 'X' ausgedehnt, so erhält man eine multiple lineare Regressionsfunktion (vgl. Formel 60).

Formel 60: Multiple lineare Regressionsfunktion⁵²⁷

$$\hat{Y} = a_0 + b_1 \cdot X_1 + b_2 \cdot X_2 \dots + b_k \cdot X_k$$

\hat{Y} steht für den geschätzten Wert der abhängigen Y-Variable
bezüglich der gewählten X-Werte [$X_1, X_2 \dots X_k$]

a_0 ist der Achsenabschnitt der linearen Gleichung, wenn
alle X-Werte gleich '0' sind

$b_1, b_2 \dots b_k$ ist die jeweilige Steigung (auch Regressionskoeffizient
genannt), d.h. wie viel sich durchschnittlich der
geschätzte Wert von \hat{Y} durch eine Veränderung lediglich
einer Einheit von X_1 bzw. X_2 bzw. ... X_k verändert

$X_1, X_2 \dots X_k$ ist jeder gewählte Wert der unabhängigen
Variablen ' X_1 ' bzw. ' X_2 ' bzw. ... ' X_k '

Hier ist festzuhalten, dass jeder (partielle) Regressionskoeffizient⁵²⁸ ' b_k ' diejenige durchschnittliche Veränderung der abhängigen Variablen 'Y' aufzeigt, die aufgrund einer Veränderung einer Einheit des diesbezüglichen Regressionskoeffizienten ' b_k ' entsteht. Dabei sind alle übrigen Regressionskoeffizienten konstant zu halten. Für die etwas komplizierten Parameterberechnungen der multiplen linearen Regressionsfunktion sei z. B. auf (Salvatore, 1982, S. 142ff) verwiesen.

17.2.1.3 Standardfehler der Schätzung (SEE)

Da die beobachteten Y-Werte kaum den anhand der Methode-der-kleinsten-Quadrate «geschätzten» \hat{Y} -Werten entsprechen, ergibt sich ein Fehler der Schätzung. Anders ausgedrückt: Es entstünde kein Fehler der Schätzung, wenn die berechnete Regressionsgerade genau durch die beobachteten Y-Werte verlaufen würde.

Da somit üblicherweise ein «Fehler der Schätzung» vorhanden ist, benötigt man eine Kennzahl, die aussagt, wie groß (bzw. klein) dieser Fehler der Schät-

⁵²⁷ Die beobachteten Y-Werte können somit auch wie folgt bestimmt werden:

$$Y = a_0 + b_1 \cdot X_1 + b_2 \cdot X_2 \dots + b_k \cdot X_k + e$$

Y steht für die beobachteten abhängigen Y-Variablen

e steht für die Residuen(abweichungen) bzw. Störgrößen

⁵²⁸ partial or net regression coefficient

zung ist. In Anlehnung an das Streuungsmaß 'Standardabweichung' wird diese Kennzahl auch als «Standardfehler der Schätzung»⁵²⁹ bezeichnet und – wie in Formel 61 für die einfache lineare Regressionsfunktion aufgeführt – berechnet:

Formel 61: Standardfehler der Schätzung (SEE)

$$SEE = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n - 2}}$$

Der Standardfehler der Schätzung gleicht stark der Berechnung der Standardabweichung einer Stichprobe. Der Unterschied besteht darin, dass der Standardfehler der Schätzung die (durchschnittliche) vertikale Streuung um die Regressionsgeraden misst. Da die Regressionsgerade zwei Parameter, nämlich ' a_0 ' und ' b ', aufweist und diese grundsätzlich nicht bekannt sind, wird im Nenner durch ' $n-2$ ' dividiert.

Bei der multiplen Regressionsanalyse lautet der Nenner nicht ' $n-2$ ', sondern ' $n-k-1$ ', wobei ' k ' der Anzahl Regressionskoeffizienten entspricht und ' 1 ' dem Achsenabschnitt.

Anhand dieser Kennzahl ist es z. B. möglich, die «Güte» bzw. die Anpassungsgenauigkeit von verschiedenen Regressionsgeraden (bei gleichem Datensatz) miteinander zu vergleichen: Ein kleiner Standardfehler der Schätzung bedeutet, dass die beobachteten Werte nahe bei der Regressionsgeraden liegen, ein großer Standardfehler der Schätzung weist auf eine schlechte «Anpassung» hin. Zudem benötigt man den Standardfehler der Schätzung auch für die Berechnung des Vertrauensintervalls der Parameter der Regressionsgeraden sowie des Prognoseintervalls.

17.2.1.4 Modellannahmen der Methode-der-kleinsten-Quadrate

Damit die Parameterschätzungen für lineare Regressionsfunktionen anhand der Methode-der-kleinsten-Quadrate die Eigenschaft 'Effizienz'⁵³⁰ aufweisen, müssen folgende sieben (Modell-)Annahmen erfüllt sein:⁵³¹

529 standard error of estimate (SEE)

In gewissen Programmen wird der SEE auch als «standard error of regression» (z. B. (EViews 6 for Windows, 2008)) oder «sigma» (z. B. (SAS 9.1.3 for Windows, 2008)) genannt.

530 vgl. Kapitel 17.1.3 im Anhang sowie Fußnote 511

531 vgl. (Backhaus, Erichson, Plinke, & Weiber, 2006, S. 79f) Die 7. Modellannahme ist für die Eigenschaft 'Effizienz' zwar wünschenswert, nicht aber Voraussetzung.

1. Das Modell ist korrekt spezifiziert, d.h., es besteht ein linearer Zusammenhang in den Regressionsparametern, es enthält die relevanten unabhängigen Variablen 'X' und die Zahl der zu schätzenden Parameter ist kleiner als die Anzahl der Beobachtungen.
2. Die Residuen(abweichungen) haben einen Erwartungswert von '0'.
3. Es besteht kein Zusammenhang bzw. keine Korrelation zwischen der/den unabhängigen Variablen 'X' und den Residuen(abweichungen) 'e'.
4. Die Residuen(abweichungen) weisen eine konstante Varianz auf (Homoskedastizität⁵³²).
5. Die Residuen(abweichungen) weisen keinen Zusammenhang unter sich auf, bzw. sind unkorreliert (keine Autokorrelation⁵³³).
6. Zwischen den unabhängigen Variablen 'X' besteht keine lineare Abhängigkeit (keine (perfekte) Multikollinearität⁵³⁴).
7. Die Residuen(abweichungen) sind normal verteilt.⁵³⁵

17.2.1.5 Überprüfung der Modellannahmen

Um sicherzustellen, dass die Methode-der-kleinsten-Quadrate zur Parameterschätzung des linearen Regressionsmodells verwendet werden darf, sind deren Modellannahmen⁵³⁶ zu überprüfen.⁵³⁷ Dies wird meist anhand ausgewählter Tests durchgeführt.

Linearität

Das lineare Regressionsmodell basiert auf der Annahme der Linearität in den Parametern. Falls eine geometrische oder exponentielle Beziehung zwischen der abhängigen Variablen 'Y' und der/den unabhängigen Variablen 'X' vorhanden ist, kann diese meist durch «Transformation»⁵³⁸ (z. B. durch Logarithmieren⁵³⁹) in eine lineare umgewandelt werden. Die transformierte lineare Beziehung wird alsdann zur Parameterschätzung anhand der Methode-der-kleinsten-Quadrate verwendet. Anschließend können die damit berechneten Prognosewerte durch korrek-

532 homoscedasticity

533 serial correlation or autocorrelation

534 multicollinearity

535 normality of residuals

536 vgl. Kapitel 17.2.1.4 im Anhang

537 vgl. (Backhaus, Erichson, Plinke, & Weiber, 2006, S. 80ff), (Menzefricke, 1995, S. 449ff) und (Urban & Mayerl, 2008, S. 177ff)

538 Beispiele hierzu können (Bohley, 1996, S. 219) entnommen werden.

539 Gemeint ist hier der natürliche Logarithmus 'ln' mit der Basis 'e' (Eulersche Zahl).

te «Rücktransformation» in die ursprüngliche nicht lineare Beziehung überführt werden.

Eine nicht entdeckte Nichtlinearität führt zu einer Verzerrung der Parameterschätzung trotz steigendem Stichprobenumfang.

Der einfachste diesbezügliche «Test» besteht darin, die Beziehung zwischen der abhängigen Variablen 'Y' und der/den unabhängigen Variablen 'X' visuell anhand eines Streudiagramms⁵⁴⁰ aufzuzeigen.

Erwartungswert der Residuen(abweichungen) gleich '0'

Falls diese 2. Modellannahme verletzt wird, ist ein systematischer Messfehler die Folge, welcher sich im Achsenabschnitt ' a_0 ' niederschlägt. Dieser wird somit konstant zu groß oder zu klein ausgewiesen. Da in den meisten Anwendungen der Achsenabschnitt nicht von großer Bedeutung ist, fällt diese Verzerrung nicht sonderlich ins Gewicht. Wird aber ein lineares Regressionsmodell ohne Konstante (Achsenabschnitt) unterstellt, kann dies zu einer Verzerrung aller Regressionskoeffizienten führen. Aus diesem Grund sollte nur in sehr begründeten Fällen im linearen Regressionsmodell die Konstante weggelassen werden.

Vorhandensein aller relevanten unabhängigen Variablen

Diese Modellannahme (siehe 1. Modellannahme) ist überaus schwierig zu überprüfen. Wenn sie nicht eingehalten wird, dafür aber die 3. Modellannahme⁵⁴¹, entsteht lediglich die gleiche Verzerrung wie wenn der Erwartungswert der Residuen(abweichungen) ungleich '0' wäre.

Meist zeigt sich eine Verletzung dieser Modellannahme darin, dass die Vorzeichen der Regressionskoeffizienten nicht dem ökonomischen bzw. logischen Ursache-Wirkung-Zusammenhang entsprechen.

Keine bedingte Heteroskedastizität

Bedingte Heteroskedastizität⁵⁴² besteht dann, wenn die Residuen(abweichungen)⁵⁴³ entlang der unabhängigen Variablen 'X' (bzw. der prognostizierten abhängigen Variablen 'Ŷ') variieren.

540 scatterplot

541 Es besteht kein Zusammenhang bzw. Korrelation zwischen der/den unabhängigen Variablen 'X' und den Residuen(abweichungen).

542 conditional heteroscedasticity

543 residuals

Die Residuen(abweichungen) 'e' sind die jeweilige Differenz zwischen den effektiv gemessenen 'Y'-Werten und den diesbezüglichen (Prognose-)Werten 'Ŷ'.

gigen Variablen ' \hat{Y} ') systematisch größer oder kleiner werden.⁵⁴⁴ Dadurch wird die 3. und 4. Modellannahme verletzt.

Durch die Berechnungsmethodik der Methode-der-kleinsten-Quadrate erhalten die «großen» Residuen(abweichungen) mehr Gewicht und beeinflussen damit – rein systemtechnisch bedingt – die Lage der Regressionsgeraden. Diese ist somit nicht mehr verlässlich. Die Verletzung dieser 4. Modellannahme führt dazu, dass die Standardfehler der Regressionskoeffizienten sowie deren Signifikanz verfälscht bzw. ungenau werden.⁵⁴⁵

Am einfachsten wird das mögliche Vorhandensein der bedingten Heteroskedastizität via ein Streudiagramm⁵⁴⁶ der 'X'-Werte (bzw. der prognostizierten abhängigen Variablen ' \hat{Y} ') und der Residuen(abweichungen) getestet: Falls die Streuung der Residuen(abweichungen) entlang der X-Achse systematisch zu- oder abnimmt, besteht bedingte Heteroskedastizität.

In Abbildung 105 ist die bedingte Heteroskedastizität visualisiert.

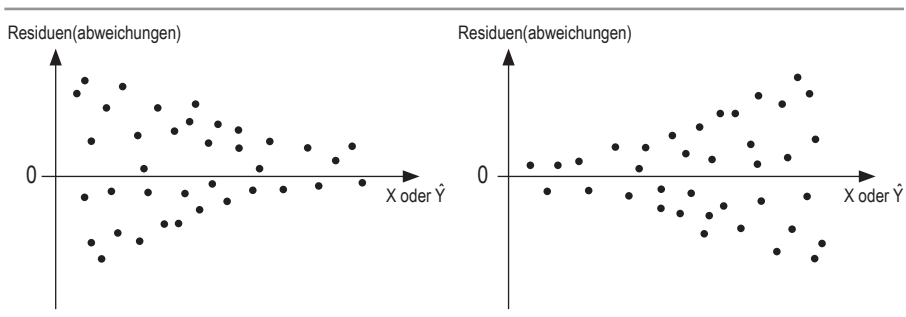


Abbildung 105: Bedingte Heteroskedastizität

Nebst dem visuellen Test bezüglich der Residuen(abweichungen) ist vor allem der sogenannten «Breusch-Pagan-Test»⁵⁴⁷ und «White-Test»⁵⁴⁸ verbreitet. Beide

544 Wenn dieses Größer- und Kleinerwerden der Residuen(abweichungen) nicht systematisch geschieht, spricht man lediglich von Heteroskedastizität, nicht aber von einer bedingten Heteroskedastizität. Die Auswirkungen sind in jenem Fall normalerweise nicht von wesentlicher Bedeutung.

545 Die Konsequenzen des Vorhandenseins von Heteroskedastizität, das diesbezügliche Testen sowie mögliche Korrekturverfahren sind bei (Wooldridge, 2008, S. 264ff) ausführlich dargelegt.

546 scatterplot

547 vgl. (Breusch & Pagan, 1979) und z.B. (Pindyck & Rubinfeld, 1997, S. 154ff), (Defusco, McLeavey, Pinto, & Runkle, 2001, S. 447f)

548 vgl. (White, 1980, S. 821ff) und z.B. (Pindyck & Rubinfeld, 1997, S. 156f)

Tests untersuchen die Nullhypothese 'Homoskedastizität' (bzw. keine Heteroskedastizität). Wenn die Teststatistik, welche eine Chiquadratverteilung⁵⁴⁹ mit einem Freiheitsgrad von 'k' (Anzahl Regressionskoeffizienten) aufweist, größer ist als der Kritische Wert⁵⁵⁰, dann wird die Nullhypothese abgelehnt, d. h., das Vorliegen von Heteroskedastizität wird akzeptiert.

Liegt eine bedingte Heteroskedastizität vor, kann diese z. B. durch die 'Methode-der-gewichteten-kleinsten-Quadrate'⁵⁵¹ umgangen werden.⁵⁵² Diese Methode versucht die Residuen(abweichungen) so zu gewichten, dass der systemtechnische Fehler der Berechnungsmethodik der Methode-der-kleinsten-Quadrate möglichst klein gehalten wird. Dies kann z. B. dadurch erreicht werden, indem jede Gewichtung so festgelegt wird, dass sie dem Kehrwert der Varianz der Residuen(abweichungen) entspricht.

Da in den meisten Untersuchungen nicht so sehr die Regressionsfunktion an und für sich von Interesse ist, sondern vielmehr die korrekte Aussage über die Signifikanz der Regressionskoeffizienten, wurden entsprechende «Korrekturvverfahren» entwickelt.⁵⁵³

Die beispielhafte Quartalsumsatzentwicklung eines Unternehmens (vgl. Abbildung 106) sei zur Verdeutlichung der bedingten Heteroskedastizität und deren Auswirkung auf die lineare Regression anhand der Methode-der-kleinsten-Quadrate sowie auf den «Lösungsansatz» anhand der Methode-der-gewichteten-kleinsten-Quadrate herangezogen.

549 vgl. Chiquadratverteilung in Kapitel 17.5 im Anhang

550 Der p-Wert zur entsprechenden Statistik ist dann auch kleiner als das festgelegte Signifikanzniveau.

551 weighted least squares (WLS)

vgl. dazu z. B. die Ausführungen bei (Pindyck & Rubinfeld, 1997, S. 148f & 158f)

552 Die WLS-Methode gehört zur «Familie» der 'generalisierten Methoden der kleinsten Quadrate' [generalized least squares (GLS)]

vgl. dazu z. B. die Ausführungen bei (Pindyck & Rubinfeld, 1997, S. 170ff)

553 vgl. (White, 1980) mit einer Lösung anhand einer Heteroskedastizitäts-konsistenten Kovarianzmatrix; (Newey & West, 1987) mit einer Lösung anhand einer Heteroskedastizitäts- und Autokorrelations-konsistenten Kovarianzmatrix

Beide «Lösungsansätze» verändern nicht die Werte der Regressionskoeffizienten, sondern «korrigieren» lediglich den Standardfehler der Regressionskoeffizienten. Dadurch kann die Verlässlichkeit der 'Signifikanz' wieder sichergestellt werden.

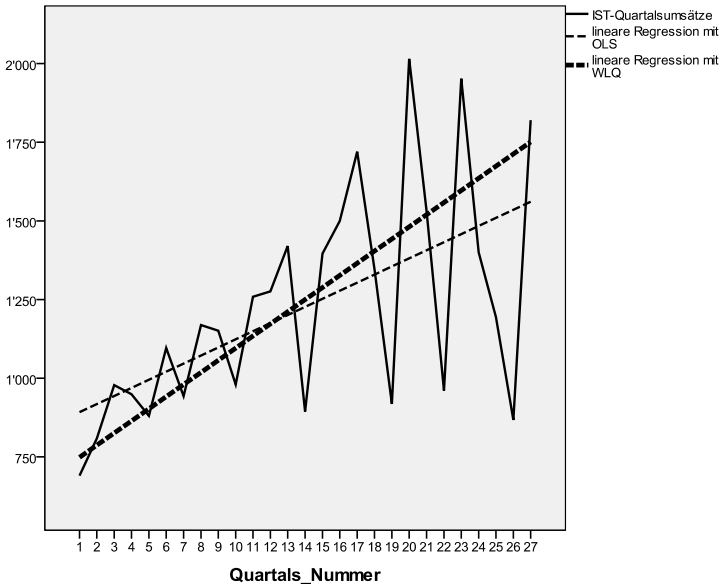


Abbildung 106: Beispiel der Anwendung der Methode-der-gewichteten-kleinsten Quadrate [mit (PASW Statistics 17.0, 2009): 'Analyze – Regression – Linear' (für OLS); 'Analyze – Regression – Weight Estimation' verbunden mit 'Analyze – Regression Linear' und 'WLS Weight' (für WLS)]

Die ausgezogene Linie in Abbildung 106 zeigt die IST-Quartalsumsätze auf. Wird eine lineare Regression nach der Methode-der-kleinsten-Quadrate⁵⁵⁴ in diesen Datensatz «gelegt», ergibt sich die schmal gestrichelte Gerade. Durch die im Zeitverlauf steigende Streuung der IST-Quartalsumsätze werden die Residuen (abweichungen) systematisch größer (vgl. Abbildung 107). Durch die zusätzlich tendenziell stärker nach unten ausschlagenden Quartalsumsätze gegen Ende der Zeitreihe wird die Gerade durch die Berechnungsmethodik der kleinsten Quadrate «systembedingt» im Uhrzeigersinn gedreht. Anhand der linearen Regression nach der Methode-der-gewichteten-kleinsten-Quadrate⁵⁵⁵ wird diese Verzerrung korrigiert: Die dick gestrichelte WLS-Gerade hat sich nun – gegenüber der OLS-Geraden – im Gegenuhrzeigersinn adjustiert.

554 least-squares method or normal least squares (NLS) or ordinary least squares (OLS)

555 weighted least squares (WLS)

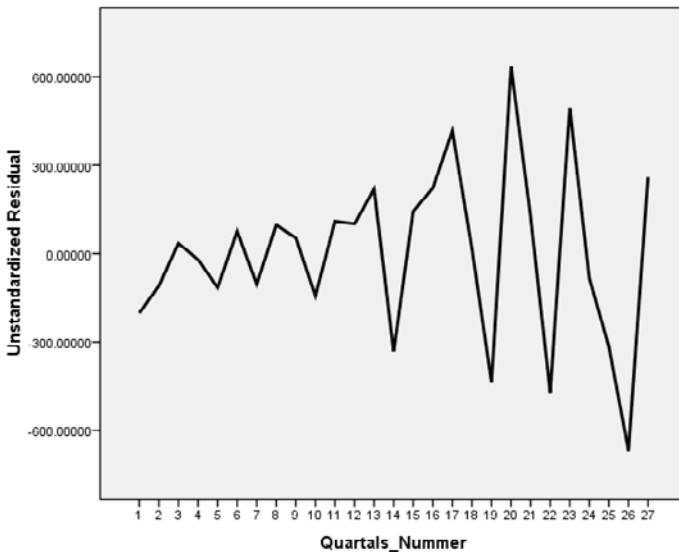


Abbildung 107: Systematisch größer werdende Residuen(abweichungen) zu Beispiel in Abbildung 106 [mit (PASW Statistics 17.0, 2009): 'Analyze – Regression – Linear' (zur Berechnung der Residuen(abweichungen)); 'Analyze – Forecasting – Sequence Charts' (für Darstellung)]

Keine Autokorrelation

Autokorrelation⁵⁵⁶ besteht dann, wenn die Residuen(abweichungen) untereinander korrelieren, bzw. die beobachteten 'Y'-Werte einen Zusammenhang (bzw. Abhängigkeit) untereinander aufweisen. Die Verletzung dieser Modellannahme tritt vor allem bei Zeitreihen auf. Nicht von Bedeutung ist sie, wenn kein sachlicher Zusammenhang unter den beobachteten 'Y'-Werten besteht, z. B. (Kosten-)Vergleiche von Sparten, Filialen usw. Wie bereits bei der Heteroskedastizität führt die Verletzung dieser 5. Modellannahme dazu, dass die Standardfehler der Regressionskoeffizienten sowie deren Signifikanz verfälscht bzw. ungenau werden.

Am einfachsten wird auch hier das mögliche Vorhandensein von Autokorrelation via ein Streudiagramm⁵⁵⁷ der 'X'-Werte (bzw. der prognostizierten abhängigen Variablen 'Y') und der Residuen(abweichungen) getestet: Wenn die Residuen(abweichungen) entlang der 'X'-Werte (bzw. der prognostizierten abhängigen Variablen 'Y') nahe, gleichgerichtet beieinanderliegen, spricht man von positiver Auto-

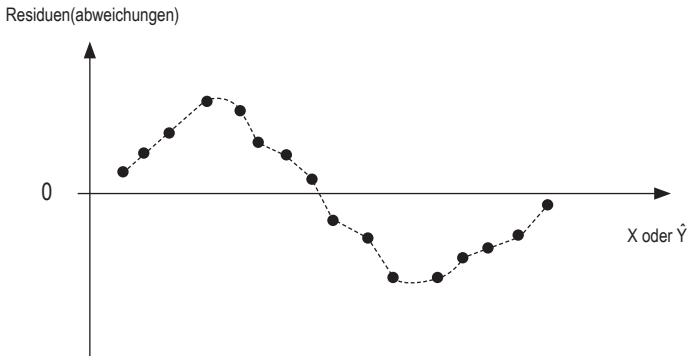
556 autocorrelation or serial correlation

557 scatterplot

korrelation; wenn sie hingegen auf- und absteigen, spricht man von negativer Autokorrelation.

In Abbildung 108 sind je ein Beispiel für das Vorhandensein von positiver und negativer Autokorrelation visualisiert.

positive Autokorrelation



negative Autokorrelation

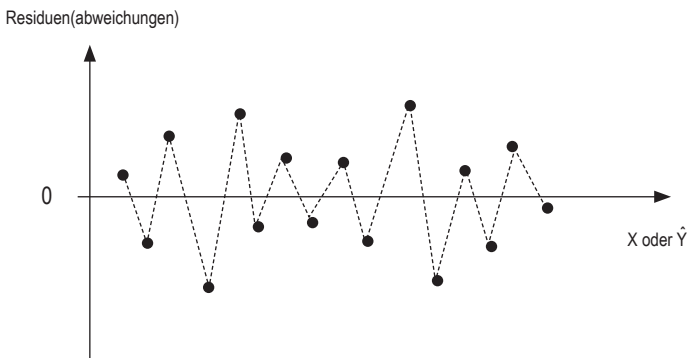


Abbildung 108: Beispiele für positive und negative Autokorrelation

Nebst dem visuellen Test bezüglich der Residuen(abweichungen) ist vor allem die Durbin-Watson-Statistik⁵⁵⁸ (DW) zur Abklärung der Autokorrelation⁵⁵⁹ bekannt. Diese definiert sich wie in Formel 62 aufgeführt.

558 vgl. (Durbin & Watson, Testing for Serial Correlation in Least Squares Regression. I | II | III, 1950 | 1951 | 1971)

559 Genau genommen «testet» die DW-Statistik, ob unter den Residuen(abweichungen) ein autoregressives Modell erster Ordnung bzw. ein AR(1) besteht oder nicht.

Formel 62: Durbin-Watson-Statistik

$$DW = \frac{\sum (e_t - e_{t-1})^2}{\sum e_t^2}$$

e_t Residuen(abweichungen) bzw. $(Y - \hat{Y})$

Unter Annahme eines entsprechenden Signifikanzniveaus kann für die Nullhypothese 'wahrer Autokorrelationskoeffizient = 0' ein oberer Grenzwert (d_U) und für die Nullhypothese 'wahrer Autokorrelationskoeffizient > 0' ein unterer Grenzwert (d_L) des Konfidenz- bzw. Vertrauensintervalls berechnet werden.⁵⁶⁰ Werden die beiden Grenzwerte mit der berechneten DW-Statistik in Beziehung gesetzt, können folgende Schlussfolgerungen gezogen werden:

Wert der DW-Statistik	Schlussfolgerung
$0 < DW < d_L$	positive Autokorrelation
$d_L < DW < d_U$	keine statistisch signifikante Aussage möglich
$DW = 2$	(gar) keine Autokorrelation
$(4 - d_U) < DW < (4 - d_L)$	keine statistisch signifikante Aussage möglich
$(4 - d_L) < DW < 4$	negative Autokorrelation

Die oberen und unteren Grenzwerte sind von (Durbin & Watson, Testing for Serial Correlation in Least Squares Regression. II, 1951, S. 173-175) bezüglich 'n' und 'k' berechnet und tabellarisch zusammengefasst worden.⁵⁶¹ Dabei bedeutet 'n' die Anzahl Beobachtungen (bzw. der Stichprobenumfang) und 'k' die Anzahl Regressionskoeffizienten.

Liegt Autokorrelation vor, kann diese Verletzung der 5. Modellannahme z. B. durch (einfache) Differenzenbildung «gelöst» oder zumindest reduziert werden. Eine weitere Möglichkeit besteht in der Anwendung des Cochrane-Orcutt-Algorithmus⁵⁶² (bzw. des Prais-Winsten-Lösungsansatzes⁵⁶³). Dieses Autoregressions-

560 vgl. (Durbin & Watson, Testing for Serial Correlation in Least Squares Regression. II, 1951, S. 173-175)

561 siehe dazu die ausgewählten DW-Statistik-Werte in Kapitel 17.7 im Anhang

562 vgl. (Cochrane & Orcutt, 1949)

Mit diesem Verfahren werden die «wahren» Regressionskoeffizienten bei Zeitreihenwerten mit autokorrelierenden Residuen erster Ordnung (AR(1)) geschätzt.

563 vgl. (Prais & Winsten, 1954)

verfahren erweitert die lineare Regressionsfunktion durch eine Komponente von autoregressiven Residuen(abweichungen) erster Ordnung (AR(1)).⁵⁶⁴ Dadurch wird jedoch die Durbin-Watson-Statistik als Prüfgröße hinfällig.

Eine weitere Möglichkeit besteht darin, die – durch Autokorrelation und bedingter Heteroskedastizität ausgelöst – «verfälschten» Standardfehler durch z. B. die Hansen-Korrektur⁵⁶⁵ auszugleichen.⁵⁶⁶ Dabei sollte auch abgeklärt werden, ob es nicht vielleicht besser wäre, ein komplett anderes Prognosemodell (z. B. ein ARIMA-Modell) anzuwenden.

Die beispielhafte Quartalsumsatzentwicklung von zwei Unternehmen (vgl. Abbildung 109 und Abbildung 110) sei zur Verdeutlichung der positiven bzw. negativen Autokorrelation und deren Auswirkung auf die lineare Regression anhand der Methode-der-kleinsten-Quadrate sowie unter Anwendung des Cochrane-Orcutt-Algorithmus herangezogen.

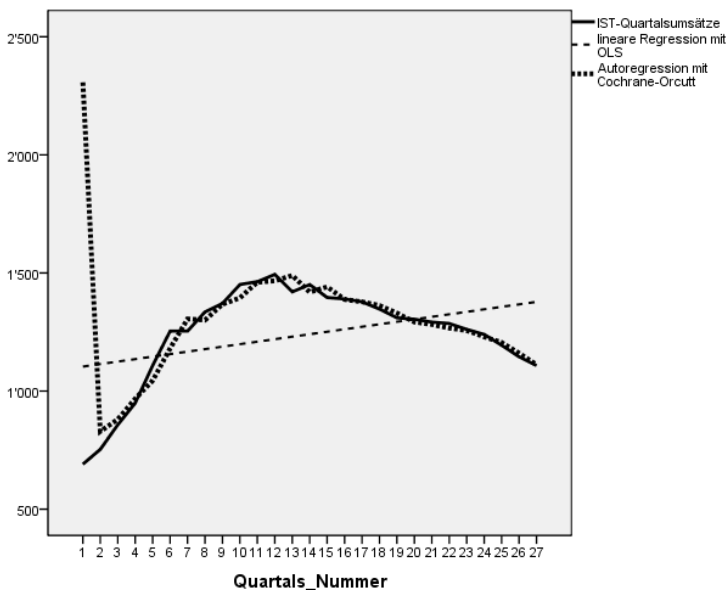


Abbildung 109: Beispiel der Anwendung des Cochrane-Orcutt-Lösungsansatzes bei positiver Autokorrelation [mit (PASW Statistics 17.0, 2009): 'Analyze – Regression – Linear' (für OLS); mit dem Befehl 'AREG' und Unterbefehl 'CO' (für die Autoregression)]

564 Der optimale, zu schätzende Parameter des autoregressiven Modells erster Ordnung wird üblicherweise mit 'RHO' bezeichnet.

565 vgl. (Hansen, 1982)

566 vgl. dazu auch Fußnote 553

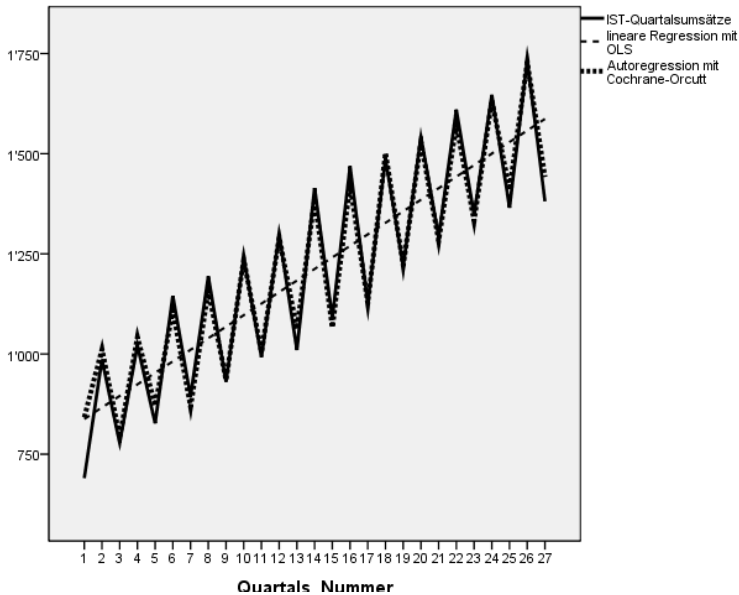


Abbildung 110: Beispiel der Anwendung des Cochrane-Orcutt-Lösungsansatzes bei negativer Autokorrelation [mit (PASW Statistics 17.0, 2009): 'Analyze – Regression – Linear' (für OLS); mit dem Befehl 'AREG' und Unterbefehl 'CO' (für die Autoregression)]

Die ausgezogenen Linien in Abbildung 109 und Abbildung 110 zeigen jeweils die IST-Quartalsumsätze auf. Wird jeweils eine lineare Regression nach der Methode der kleinsten Quadrate⁵⁶⁷ in die Datensätze «gelegt», ergeben sich die schmal gestrichelten Geraden, die sich jeweils in der Mitte positionieren, den effektiven Verlauf der IST-Quartalsumsätze aber nur sehr rudimentär reflektieren und deshalb auch die entsprechenden Residuen (abweichungen) aufweisen (vgl. Abbildung 111). Wird nun der Cochrane-Orcutt-Algorithmus für die beiden Zeitreihen angewandt, werden die autoregressiven Residuen (abweichungen) erster Ordnung (AR(1)) ausgeschaltet, und es entsteht eine korrigierte Regressionsgerade (dick gestrichelte Linie), die sich dem Kurvenverlauf anpasst.

567 least-squares method or normal least squares (NLS) or ordinary least squares (OLS)

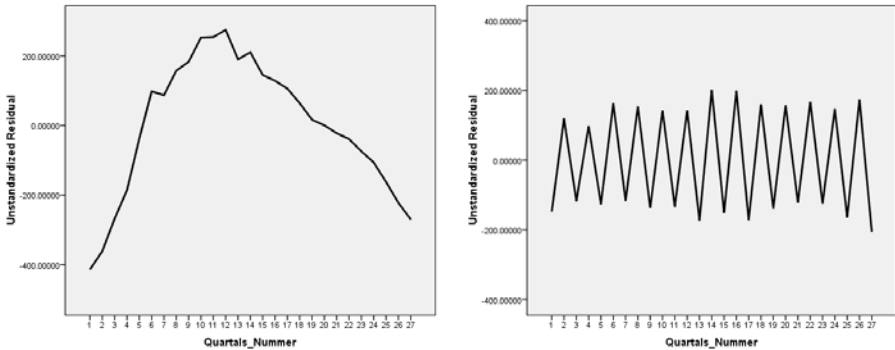


Abbildung 111: Residuen(abweichungen) bei positiver und negativer Autokorrelation zu den Beispielen in Abbildung 109 und Abbildung 110

[mit (PASW Statistics 17.0, 2009): 'Analyze – Regression – Linear' (zur Berechnung der Residuen(abweichungen)); 'Analyze – Forecasting – Sequence Charts' (für Darstellung)]

Keine Multikollinearität

Multikollinearität⁵⁶⁸ besteht dann, wenn eine hohe Korrelation zwischen zwei oder mehreren Regressionskoeffizienten⁵⁶⁹ besteht, d. h., zwei oder mehrere Regressionskoeffizienten weisen eine gegenseitige Abhängigkeit unter sich auf.⁵⁷⁰ Man kann somit auch von einer «Überspezifizierung» des (Regressions-)Modells sprechen. Durch die Verletzung dieser 6. Modellannahme entsteht die Problematik, dass die Standardfehler der Regressionskoeffizienten zu groß ausfallen. Bei Multikollinearität wird die Eindeutigkeit der Ursache-Wirkung-Kette verletzt, d. h., die Streuungen der unabhängigen Variablen überschneiden sich und weisen deshalb neben einem direkten Einfluss auf die abhängige Variable auch noch einen indirekten Einfluss auf, ausgelöst durch eine Abhängigkeit zu einer anderen unabhängigen Variablen.

Ein typischer Hinweis für Multikollinearität besteht, wenn die einzelnen t-Teststatistiken der Regressionskoeffizienten unter dem Kritischen Wert liegen (d. h. keine signifikante Unterscheidung von '0' aufweisen), die F-Teststatistik⁵⁷¹ hingegen für das ganze Modell einen signifikanten Wert aufzeigt. Ein weiterer Hinweis für Multikollinearität besteht dann, wenn sich die Vorzeichen der Parameter der Regressionskoeffizienten gegenüber der logischen Erwartung entgegengesetzt verhalten.

568 multicollinearity

569 regression coefficient

570 vgl. dazu z. B. die Ausführungen bei (Pindyck & Rubinfeld, 1997, S. 95ff)

571 vgl. auch Kapitel 13.2.2 sowie die F-Verteilung in Kapitel 17.6 im Anhang

Bevor das Problem der Multikollinearität gelöst werden kann, ist deren Vorhandensein festzustellen. Dies geschieht am einfachsten durch das Aufstellen einer Korrelationsmatrix⁵⁷², bei welcher der Zusammenhang zwischen den Variablen anhand des Korrelationskoeffizienten⁵⁷³ gemessen wird.⁵⁷⁴

In Abbildung 112 ist beispielhaft eine Korrelationsmatrix für die drei unabhängigen Variablen 'X1', 'X2' und 'X3' aufgeführt. Wie ersichtlich ist, weisen die unabhängigen Variablen 'X1' und 'X2' eine hohe Korrelation von 0.991 unter sich auf, die bei einem Signifikanzniveau von 1 % statistisch signifikant ist. Dies lässt auf eine Multikollinearität zwischen 'X1' und 'X2' schließen: Eine der beiden unabhängigen Variablen ist aus dem linearen Regressionsmodell auszuschließen. Die Abhängigkeiten zwischen 'X1' und 'X3' sowie 'X2' und 'X3' sind weder hoch, noch statistisch signifikant, weshalb hier keine Multikollinearität angenommen werden muss.

Correlations

		X1	X2	X3
X1	Pearson Correlation	1	.991**	.464
	Sig. (2-tailed)		.000	.071
	N	16	16	16
X2	Pearson Correlation	.991**	1	.445
	Sig. (2-tailed)	.000		.084
	N	16	16	16
X3	Pearson Correlation	.464	.445	1
	Sig. (2-tailed)	.071	.084	
	N	16	16	16

** . Correlation is significant at the 0.01 level (2-tailed).

Abbildung 112: Beispiel einer Korrelationsmatrix

[mit (PASW Statistics 17.0, 2009): Analyze – Correlation – Bivariate]

572 correlation

573 coefficient of correlation

vgl. Kapitel 6.6.2

Z.B. eignet sich der Korrelationskoeffizient nach Pearson zur Aufdeckung linearer Zusammenhänge, der (Rang-)Korrelationskoeffizient nach Spearman für nicht lineare und nicht parametrische Beziehungen.

574 Im einfachsten Fall wird eine sogenannte 'bivariate' Korrelation gemessen. Möchte man aber eine «Grenzkorrelation» messen, ist eine sogenannte 'partielle' Korrelation zu berechnen. Dabei werden eine oder mehrere (unabhängige Variablen) «eingefroren», damit deren Einfluss auf den Zusammenhang das Ergebnis nicht verfälscht.

Ein weitere, sehr ähnliche Möglichkeit zur Aufdeckung vorhandener Multikollinearität besteht in der Berechnung des sogenannten «Variance Inflation Factors» (VIF), der je Regressionskoeffizient (wie in Formel 63 aufgeführt) berechnet wird: Je größer der VIF ist, desto eher liegt Multikollinearität vor.⁵⁷⁵

Formel 63: Variance Inflation Factor (VIF)

$$VIF_j = \frac{1}{1 - R_j^2}$$

R_j^2 ist der Determinationskoeffizient der Regression der unabhängigen Variable 'X_j' gegenüber den verbleibenden unabhängigen Variablen 'X_{k-1}'

Falls Multikollinearität vorliegt, sind die «überspezifizierenden» Regressionskoeffizienten aus dem multiplen Regressionsmodell zu entfernen. Damit nun die richtige Auswahl der unabhängigen Regressionskoeffizienten ermittelt werden kann, wurden verschiedene Verfahren entwickelt.⁵⁷⁶ Als Auswahlkriterium wird meist der Adjustierte Determinationskoeffizient⁵⁷⁷ verwendet, bekannt sind aber auch das sogenannte «Akaike Information Criterion» (AIC) und das «Bayesian Information Criterion» (BIC) bzw. «Schwarz Information Criterion» (SIC).⁵⁷⁸ Je kleiner das AIC, BIC bzw. SIC ist, desto besser ist die Modellanpassung.

Normalität der Residuen(abweichungen)

Unter dem Begriff «Normalität der Residuen(abweichungen)»⁵⁷⁹ versteht man, dass die Residuen(abweichungen) normal verteilt sind. Falls diese 7. Modellannahme nicht erfüllt ist, weisen die Parameterschätzungen bei linearen Regressionsfunktionen anhand der Methode-der-kleinsten-Quadrate trotzdem die Eigenschaft 'Effizienz'⁵⁸⁰ auf. Doch bei der Durchführung von Signifikanztests bezüglich der Parameterschätzungen (bzw. deren Treffsicherheit und Güte) sowie dem

575 Üblicherweise geht man davon aus, dass ein 'VIF > 2' auf eine starke Multikollinearität hinweist, ein 'VIF > 10' auf eine signifikante Multikollinearität.

576 Bekannt sind folgende Verfahren: «schrittweise Auswahl» [stepwise] [vgl. z. B. (Häcke & Wichern, 2008, S. 304ff)], «rückwärtige Elimination» [backward elimination] und «Vorwärtsselektion» [forward selection]. [vgl. z. B. (Siegel A. F., 2002, S. 547f)]

577 vgl. Kapitel 13.2.1

578 vgl. hierzu Kapitel 8.5

579 normality of residuals

580 vgl. Kapitel 17.1.3 im Anhang sowie Fußnote 511

F-Test⁵⁸¹ zur Überprüfung des Regressionsmodells als Ganzes ist die Erfüllung dieser Modellannahme wichtig: "Da die Störgröße [...] die gemeinsame Wirkung sehr vieler und im Einzelnen relativ unbedeutender Einflußfaktoren repräsentiert, die voneinander weitgehend unabhängig sind, lässt sich die Annahme der Normalverteilung durch den "zentralen Grenzwertsatz" der Statistik stützen."⁵⁸² Dies gilt insbesondere bei einem Stichprobenumfang 'n' von weniger als '40'.

Die Normalität der Residuen(abweichungen) kann z.B. anhand einer Häufigkeitsverteilung⁵⁸³ der Residuen(abweichungen) überprüft werden. Ist diese in etwa normal verteilt, kann davon ausgegangen werden, dass die 7. Modellannahme 'Normalität der Residuen(abweichungen)' erfüllt ist (vgl. Beispiel in Abbildung 113).

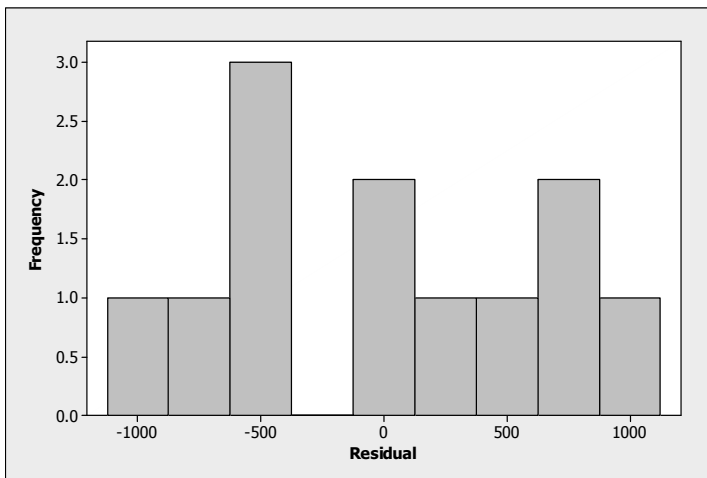


Abbildung 113: Beispiel einer Häufigkeitsverteilung der Residuen(abweichungen)
[mit (Minitab 15.1 Statistical Software, 2008): Stat – Regression – Graphs – Histogram of residuals]

Das Histogramm der Residuen(abweichungen) in Abbildung 113 zeigt aber nicht eindeutig auf, ob Letztere normal verteilt sind oder nicht. Deshalb bedient man sich einer weiteren Möglichkeit, indem man die Residuen(abweichungen) «normiert»⁵⁸⁴ und in einer Grafik aufzeichnet: Die «normierten» Residuen(abweichungen) sollten bei perfekter Normalverteilung auf einer Geraden liegen. Größere Abweichungen der Residuen(abweichungen) von der Geraden zeigen entsprechend

581 vgl. auch Kapitel 13.2.2 sowie die F-Verteilung in Kapitel 17.6 im Anhang

582 (Backhaus, Erichson, Plinke, & Weiber, 2006, S. 80)

583 frequency distribution

584 De facto werden die kumulierten Häufigkeiten logarithmiert.

auch eine Abweichung von der Normalverteilung auf.⁵⁸⁵ Durch das Aufzeichnen eines oberen und unteren Verlaufs der Konfidenz- bzw. Vertrauensgrenzwerte (z. B. für ein Vertrauensniveau von 95 %), kann besser verdeutlicht werden, ob die Abweichungen signifikant sind oder nicht. Weiter kann anhand z. B. der «Anderson-Darling-Statistik»⁵⁸⁶ (AD) ein Test bezüglich Normalverteilung durchgeführt werden: Je kleiner die Anderson-Darling-Statistik ist, desto besser entsprechen die Residuen(abweichungen) einer Normalverteilung. Falls der zur Anderson-Darling-Statistik dazugehörige p-Wert kleiner als das vorgegebene Signifikanzniveau ' α ' (z. B. 5 %) ausfällt, dann muss die Normalität der Residuen **abgelehnt** werden. In Abbildung 114 ist eine entsprechende Grafik mit Vertrauensgrenzwerte bei einem Niveau von 95 % beispielhaft aufgeführt.

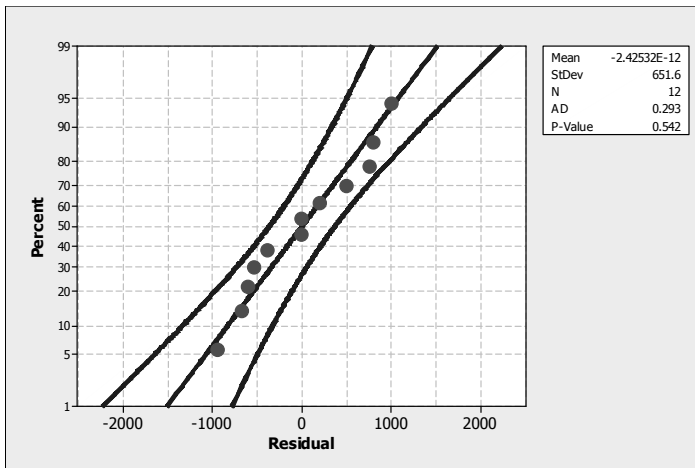


Abbildung 114: Beispiel von approximativ normal verteilten Residuen(abweichungen)

[mit (Minitab 15.1 Statistical Software, 2008): Stat – Basic Statistics – Normality Test | Edit Distribution Fit – Options – Show Confidence Interval]

585 Bei einer kleinen Stichprobe ($n < 50$) ist es – trotz vorhandener Normalität der Residuen(abweichungen) – möglich, dass die «normierten» Residuen(abweichungen) von der Geraden leicht abweichen.

586 vgl. (Anderson & Darling, 1952)

Ein ähnlicher Test wie die Anderson-Darling-Statistik bezüglich «Normalität» ist die sogenannte «Kolmogorov-Smirnov-Statistik» (KS). Letztere weist aber eine geringere Treffsicherheit aus [vgl. (Rayner & Best, 1989, S. 64 & 93)].

Weiter bekannt ist die «Jarque-Bera-Statistik» [vgl. (Jarque & Bera, 1980)], welche eine Chiquadrat-Verteilung mit ' $df = 2$ ' unter der Nullhypothese der Normverteilten Residuen(abweichungen) aufweist. Auch hier wird – wie bei der AD-Statistik – ein möglichst großer p-Wert angestrebt, bzw. ein p-Wert, der größer ist als das vorgegebene Signifikanzniveau.