UNIVERSITY OF APPLIED SCIENCES

THM
TECHNISCHE HOCHSCHULE MITTELHESSEN

Kai Bruchlos

Zweidimensionale lineare Regressionsanalyse mit Fehlern in den Variablen

THM-Hochschulschriften Band 1

Kai Bruchlos

Zweidimensionale lineare Regressionsanalyse mit Fehlern in den Variablen

THM-Hochschulschriften Band 1

THM-Hochschulschriften Band 1

Herausgeber der THM-Hochschulschriften:
Der Präsident der Technischen Hochschule Mittelhessen

Einzelne Hochschulschriften sind auch online abrufbar:
www.thm.de/bibliothek/hochschulschriften

ISSN (Print) 1439-1112
ISSN (Online) 2568-3020

Vorwort

Ziel dieser Abhandlung ist es, Voraussetzungen und Aussagen verschiedener Modelle der linearen Regressionsanalyse genau anzugeben, die Zusammenhänge zwischen den Modellen aufzuzeigen und die Anwendungsmöglichkeiten darzustellen.

Die eigentliche Aufgabe der Regressionsanalyse ist eine Optimierungsaufgabe: Suche aus einer Menge von Funktionen die heraus, die „am besten“ zu einer gegebenen Menge von Punkten P_i passt. — Wir beschränken uns hier auf den einfachsten Fall: Bestimme die Gerade, die „am nächsten“ zu den Punkten $(x_1, y_1), \ldots, (x_n, y_n)$ liegt. — Erstaunlich ist, dass mir keine Quelle bekannt ist, in der im Zusammenhang mit der Regressionsanalyse von dieser Optimierungsaufgabe die Rede ist. Die Formulierung der Aufgabe als Optimierungsproblem ist deshalb so wichtig, da sie der Ausgangspunkt der mathematischen Entwicklung der verschiedenen Modellvarianten ist. Wir beginnen in Kapitel 1 und 2 auch mit dem Optimierungsproblem und nennen in Abgrenzung zur schließenden Statistik den Abschnitt ganz bewusst Regressionsrechnung als Teilgebiet der beschreibenden Statistik.

Ebenso erstaunlich ist, dass höchst selten auf eine zentrale Voraussetzung für die Modelle der gewöhnlichen Regressionsanalyse hingewiesen wird: Die beobachteten Werte $x_1, \ldots, x_n$ der unabhängigen Variable sind nicht nur fest gewählt, sondern müssen bei jeder Wiederholung der Stichprobenziehung dieselben sein.[1] Derart deutlich schreiben dies in Lehrbüchern nach meiner Kenntnis nur Fuller 1987 und Gujarati 1988.[2] Diese Voraussetzung ist in aller Regel bei Zeitreihen erfüllt, bei anderen Merkmalen, insbesonderen bei ökonomischen meist nicht[3]. Die beobachteten Werte $y_1, \ldots, y_n$ der abhängigen Variable können sich hingegen bei Wiederholung der Stichprobe ändern, sind mit zufälligen Schwankungen versehen.

Bemerkenswert ist auch, dass das mathematische Grundmodell, welches den Übergang von der beschreibenden zur schließenden Statistik ermöglicht, nur in einer mir bekannten neueren Quelle dargestellt wird, nämlich in Fisz

[1] Die Verletzung dieser Voraussetzung hat fundamentale Folgen: Vergleiche Gujarati 1988, S. 417.

[2] Fuller 1987, S. 1.Gujarati 1988 wählt auf S. 19 folgende Formulierung: *The explanatory variables, on the other hand, are assumed to have fixed values (in repeated sampling), ...*

[3] Bezüglich des Einsatzes des Fehler-in-den-Variablen-Modells vergleiche Schönfeld 1971, S. 106.

1989, S. 117 f.[4] Andere Autoren verwenden zumindest implizit das mathematische Grundmodell, sprechen aber bestenfalls von bedingten Erwartungswerten.[5] Argumentiert wird meist mit den Eigenschaften der Verteilungen, die zusammengefasst das Modell ergeben.[6] Dabei lassen sich die verschiedenen Modelle, Modellvarianten sehr gut mit Hilfe des Grundmodells darstellen.

Darüber hinaus ist es rätselhaft, dass während der Entwicklungsphase des Fehler-in-den-Variablen-Modells in den 1940er bis 1960er Jahren die Arbeit von Kummell 1879 nicht entsprechend berücksichtigt worden ist, obwohl Deming 1948 sein Ergebnis anführt.[7] Gerade die erst 1969 gelösten Probleme beim funktionalen Modell hätte es so nicht gegeben.

Das Fehler-in-den-Variablen-Modell (FVM) behandelt die beiden Merkmale (Variablen) X und Y gleich: Die Merkmalswerte beider Merkmale werden mit zufälligen Abweichungen gemessen. Hieraus ergibt sich ein Vorteil des FVM: Es besteht die Möglichkeit der symmetrischen Schätzung der Regressiongeraden, d. h. es spielt keine Rolle, ob x die unabhängige Variable ist und y die abhängige oder umgekehrt.[8] Auch hier ist erstaunlich, dass diese Eigenschaft in der Literatur sehr selten erwähnt wird. Dies trifft auch auf die Unabhängigkeit bezüglich der Skalierung (Vielfache von Einheiten) zu.

Ist der Quotient der Fehlervarianzen σ_δ^2 und σ_ε^2 bekannt, dann ist das funktionale Modell mit Abstand zu bevorzugen (Lemma 2.2.3, Satz 2.2.20, Satz 3.0.4):

- Das Modell ist die Verallgemeinerung der gewöhnlichen Regression.
- Das Modell ist symmetrisch.
- Die Schätzer sind konsistent.
- Das Modell ist unabhängig bezüglich der Skalierung.

Die hiesige Beschränkung auf das Zweidimensionale, also das Betrachten von nur zwei Merkmalen ist von keiner besonderen Bedeutung. Wie die Ökonometrie zeigt, lassen sich die Ergebnisse relativ leicht ins Mehrdimensionale verallgemeinern. Der Formalismus wird allerdings aufwendiger.

In der Regel sind die beobachteten Werte $x_1, \ldots, x_n$ alle unterschiedlich genau so wie die beobachteten Werte $y_1, \ldots, y_n$. Dies muss aber nicht so

[4] Eine ältere Quelle ist Cramér 1946, S. 270 f.

[5] Vgl. Miller 1996, S. 204.

[6] Vgl. Fuller 1987, S. 30 ff.; Kendall und Stuart 1979, S. 403 ff.; Schach und Schäfer 1978, S. 155 ff.

[7] Deming 1948, S. 184. – Madansky 1959 verweist auf S. 181 nur auf Deming 1948, um die verschiedenen Namen eines Verfahrens zu belegen. Das Ergebnis von Kummel erwähnen Madansky 1959, S.202 und Lindley 1947, S. 241. Die Hinweise auf die Ergebnisse von Kummell 1879 scheinen aber keine Folgen gehabt zu haben.

[8] Vgl. Schach und Schäfer 1978, S. 159 f., Fußnote +).

sein. Ob alle beobachteten Werte $x_1, \ldots, x_n$ oder $y_1, \ldots, y_n$ gleich sein dürfen, hängt vom Modell ab. Bei der gewöhnlichen Regressionsanalyse müssen mindestens zwei der $x_1, \ldots, x_n$ unterschiedlich sein. Sonst gilt schon Satz 1.1.1 nicht.[9]

Bedanken möchte ich mich bei meinem Kollegen Prof. Dr. Manfred Börgens für die vielen Hinweise und Anmerkungen insbesondere zur Unabhängigkeit bezüglich der Skalierung.

Brietlingen, im September 2017 Kai Bruchlos

[9] Beachte Georgii 2004, S. 317, Beispiel 12.1.

Inhaltsverzeichnis

1 Gewöhnliche Regressionsanalyse

In diesem Kapitel werden die grundlegenden Eigenschaften der gewöhnlichen Regression zur Verfügung gestellt, auf die bei der Erweiterung zum Fehler-in-den-Variablen-Modell Bezug genommen wird. Der Ausgangspunkt der Regressionsanalyse, das Optimierungsproblem und eine mögliche Lösung wird in Abschnitt 1.1 formuliert. In Anschnitt 1.2 wechseln wir mit der Darstellung des wahrscheinlichkeitstheoretischen Grundmodells in die schließende Statistik, um dann in Abschnitt 1.3 Aussagen zur Qualität der KQ-Schätzer zu erhalten.

1.1 Regressionsrechnung

Die Regressionrechnung[1] ist ein Teilgebiet der beschreibenden Statistik und geht davon aus, dass Beobachtungswerte $x_1, \ldots, x_n$ und $y_1, \ldots, y_n$ zweier metrisch skalierter Merkmale X und Y vorliegen, wobei es einen funktionalen Zusammenhang der Art $f(x) = y$ geben soll. Die Wahrscheinlichkeitstheorie wird nicht verwendet.

Wir betrachten hier nur den linearen Zusammenhang zweier Merkmale X und Y, mithin die Gerade[2]

$$y = \alpha + \beta \cdot x \ .$$

Ist $n > 2$, sind also mehr als zwei Punkte (x_i, y_i) gegeben, dann liegt das überbestimmte lineare Gleichungssystem

$$\begin{aligned} y_1 &= \alpha + \beta \cdot x_1 \\ &\vdots \\ y_n &= \alpha + \beta \cdot x_n \end{aligned}$$

vor. In aller Regel kann dieses Gleichungssystem nicht exakt gelöst werden. Es ist somit eine lineare Gleichung

$$y = a + b \cdot x \ ,$$

[1] Vgl. Schwarze 2005, S. 127 ff.

[2] Mit Blick auf die Notation des weiter unten angeführten Modells der linearen Regressionsrechnung müsste hier eigentlich $\eta = \alpha + \beta \cdot x$ stehen. Diese Notation ist aber in diesem Zusammenhang nicht üblich.

die sogenannte **Regressionsfunktion**, hier **Regressionsgerade** zu bestimmen, zu der die Punkte (x_i, y_i) einen möglichst kleinen „Abstand" haben. Dies ist eine Optimierungsaufgabe. Zu deren Lösung wird meist die **Methode der kleinsten Quadrate**[3] herangezogen, welche die Summe der Quadrate der einzelnen Abweichungen

$$e_i := y_i - (a + b \cdot x_i)$$

minimiert. Wir suchen also das Minimum der Funktion

$$S : \mathbb{R} \times \mathbb{R} \to \mathbb{R},\ (a, b) \mapsto \sum_{i=1}^{n} (y_i - a - b \cdot x_i)^2 \ .$$

Die Differentialrechnung liefert den Tiefpunkt mit den Koordinaten[4]

$$b = \frac{\sum\limits_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2} \qquad \text{und} \qquad a = \bar{y} - b \cdot \bar{x}$$

a und b heißen **Kleinste-Quadrate-Schätzer**, kurz **KQ-Schätzer**. Insgesamt haben wir:

Satz 1.1.1:[5] *Seien $x_1, \ldots, x_n$ und $y_1, \ldots, y_n$ Beobachtungswerte der Merkmale X und Y an statistischen Einheiten einer Grundgesamtheit (statistischen Masse). Wird der Zusammenhang $y = \alpha + \beta \cdot x,\ \alpha \in \mathbb{R},\ \beta \in \mathbb{R} \backslash \{0\}$ angenommen, dann sind*

$$b := \frac{\sum\limits_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2} \qquad \textit{und} \qquad a := \bar{y} - b \cdot \bar{x}$$

die KQ-Schätzer für β und α.

Korollar 1.1.2:[6] *Der Punkt $(\bar{x}, \bar{y})$ liegt auf der Regressionsfunktion $y = a + bx$.*

Bemerkung 1.1.3: (i) Die Methode der kleinsten Quadrate geht wegen der Minimierung der Abweichungen

$$e_i := y_i - (a + b \cdot x_i) \iff y_i = a + b \cdot x_i + e_i$$

von zwei Annahmen aus:

[3] Vgl. Lexikon der Statistik 2004, Stichwort: Methode der kleinsten Quadrate, S. 154 f.
[4] Vgl. Hartung [u.a.] 2009, S. 575.
[5] Pestman 1998, S. 185, Proposition IV.1.1. Vgl. Schwarze 2005, S. 130 f.; Hartung [u.a.] 2009, S. 574 f.
[6] Vgl. Schwarze 2005, S. 132, R 4.3.11.

- $x_1, \ldots, x_n$ sind feste Werte, würden sich bei einer Wiederholung der Stichprobenziehung nicht ändern.

- $y_1, \ldots, y_n$ können von den Werten $a + b \cdot x_1, \ldots, a + b \cdot x_n$ abweichen, wobei $(x_1, y_1), \ldots, (x_n, y_n)$ zur selben Stichprobe gehören.[7]

Dies kann so interpretiert werden, dass x_i fehlerfrei gemessen ist, y_i fehlerbehaftet.

(ii) Oft wird die Notation $b = \hat{\beta}$ und $a = \hat{\alpha}$ an dieser Stelle verwendet. Diese Schreibweise wird hier bewusst gemieden, da wir uns noch nicht in der schließenden Statistik befinden.

(iii) Die Regressionsrechnung hat zum Ziel, dass die Parameter eines gegebenen Funktionstyps $f(x) = y$ „möglichst gut" an die Punkte $(x_1, y_1), \ldots, (x_n, y_1)$ angepasst werden. Unter „möglichst gut" wird in der Regel die Bestimmung eines Minimums mit Hilfe der Differentialrechnung verstanden. Speziell bei der zweidimensionalen linearen Regressionrechnung wird bei der Verwendung der Methode der kleinsten Quadrate nur das Minimum von S bestimmt. Wie gut dieses Minimum die Optimierungsaufgabe löst, ist bisher nicht bekannt. Oder in der Sprache der schließenden Statistik: Welche Eigenschaften die KQ-Schätzer haben, von welcher Qualität sie sind, ist bis jetzt offen. Dafür benötigen wir ein wahrscheinlichkeitstheoretisches Modell zur Regression.

Bisher sind wir davon ausgegangen, dass x die unabhängige Variable[8] und y die abhängige Variable[9] ist. Es kann aber auch der umgekehrte Fall vorliegen. Folgende Bezeichnungen sind üblich:

Definition 1.1.4:[10] Seien $x_1, \ldots, x_n$ und $y_1, \ldots, y_n$ Beobachtungswerte der Merkmale X und Y an statistischen Einheiten einer Grundgesamtheit.

(i) Wird der funktionale Zusammenhang $f(x) = y$ angenommen, dann liegt eine **Regression von y auf x** vor.

(ii) Wird der funktionale Zusammenhang $f(y) = x$ angenommen, dann liegt eine **Regression von x auf y** vor.

Es gibt Fälle, in denen die Abhängigkeit der Merkmale X und Y sachlich fest steht, also sachlich begründet ist, ob eine Regression von y auf x vorliegt oder eine Regression von x auf y. Aber es gibt auch Fälle, in denen die Abhängigkeit nicht sachlich zu begründen oder gar unsinnig ist. Für diese Fälle wäre es vorteilhaft, wenn die Regressionsfunktion $x = a' + b'y$

[7] Aus einer anderen Stichprobe $(x_1, \tilde{y}_1), \ldots, (x_n, \tilde{y}_n)$ berechnen sich in aller Regel auch andere KQ-Schätzer als a und b, etwa $\tilde{a}$ und $\tilde{b}$. Und dann weichen natürlich $\tilde{y}_1, \ldots, \tilde{y}_n$ von den Werten $\tilde{a} + \tilde{b} \cdot x_1, \ldots, \tilde{a} + \tilde{b} \cdot x_n$ ab.

[8] Andere Bezeichnungen: Erklärende Variable, Regressor, Prädiktor, exogene Variable.

[9] Andere Bezeichnungen: Erklärte Variable, Regressand, Prädiktand, endogene Variable.

[10] Vgl. Schwarze 2005, S. 128, D 4.3.2.

(im y-x-Koordinatensystem) der Regressionsfunktion $y = a + bx$ (im x-y-Koordinatensystem) entsprechen würden. Es gilt nun:

$$y = a + bx \iff x = -\frac{a}{b} + \frac{1}{b}y$$

Insbesondere muss also $b' = 1/b$ sein. Mit Blick auf die beiden Schätzer

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \text{und} \qquad b' = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

steht fest, dass $b' = 1/b$ im Allgemeinen nicht gilt.[11] Wir haben also:

Bemerkung 1.1.5: (i) Bei der gewöhnlichen linearen Regression (mit KQ-Schätzern) führt eine Regression von y auf x im Allgemeinen zu einer anderen Regressionsgeraden als eine Regression von x auf y. Die gewöhnliche lineare Regression ist **asymmetrisch**.[12]

(ii) Die Asymmetrie der gewöhnlichen linearen Regression hat ihren Ursprung darin, dass die KQ-Methode die unabhängige und die abhängige Variable unterschiedliche verwendet: Die unabhängige Variable ist fehlerfrei gemessen, die abhängige nicht – vergleiche Bemerkung 1.1.3.

Darstellung des Modells der linearen Regressionsrechnung:

$x_1, \ldots, x_n$	feste Werte des Merkmales X, ändern sich bei Wiederholung der Stichprobenziehung nicht.
$y_1, \ldots, y_n$	Werte des Merkmales Y
$y_i = a + b \cdot x_i + e_i$	Regressionsgleichungen
$y = a + b \cdot x$	Regressionsfunktion (Regressionsgeraden)
$\eta = \alpha + \beta \cdot x$	„wahrer“ Zusammenhang

1.2 Wahrscheinlichkeitstheoretisches Grundmodell

Um über die Qualität der KQ-Schätzer aus Satz 1.1.1 etwas aussagen zu können, soll ein wahrscheinlichkeitstheoretisches Modell herangezogen werden. Das einzige mir bekannte wahrscheinlichkeitstheoretische Modell der Regressionsanalyse ist von Fisz 1989 und wird auf den Seiten 117 f. dargelegt:

Sei (X, Y) eine zweidimensionale Zufallsvariable mit stetiger Dichte $f(x, y)$ und stetigen Randdichten f_X und f_Y.

[11] Vgl. Schach und Schäfer 1978, S. 159, Fußnote +.

[12] Vgl. Miller 1996, S. 156; Schach und Schäfer 1978, S. 159, Fußnote +; Madansky 1959, S. 175.

Satz 1.2.1:[13] *Für die bedingten Erwartungswerte gilt*

$$\mathbb{E}(X|Y=y) = \int\limits_{-\infty}^{\infty} x \cdot \frac{f(x,y)}{f_Y(y)}\, dx \quad \textit{sowie} \quad \mathbb{E}(Y|X=x) = \int\limits_{-\infty}^{\infty} y \cdot \frac{f(x,y)}{f_X(x)}\, dy\ .$$

Sind X und Y stochastisch unabhängig, dann gilt $\mathbb{E}(X|Y=y) = \mathbb{E}(X)$ und $\mathbb{E}(Y|X=x) = \mathbb{E}(Y)$.

Definition 1.2.2:[14] Die Punktmenge in der x-y-Ebene mit den Koordinaten

$$\big(\mathbb{E}(X|Y=y), y\big)$$

heißt **Regressionskurve der Zufallsvariablen X bezüglich Y**, die Punktmenge mit den Koordinaten

$$\big(x, \mathbb{E}(Y|X=x)\big)$$

heißt **Regressionskurve der Zufallsvariablen Y bezüglich X**.

Bemerkung 1.2.3: *Im allgemeinen überdecken sich die beiden Regressionskurven nicht.*[15] Die Regression ist also grundsätzlich asymmetrisch, vergleiche Bemerkung 1.1.5.

Die Regressionskurve der Zufallsvariablen X bezüglich Y bzw. die Regressionskurve der Zufallsvariablen Y bezüglich X entspricht in der Regressionsrechnung der Regression von x auf y bzw. der Regression von y auf x. Die Verbindung zwischen den beiden Begriffen wollen wir im nächsten Abschnitt herstellen.

1.3 Eigenschaften der Schätzer

Wir betrachten im folgenden nur die Regressionskurve der Zufallsvariablen Y bezüglich X und stellen die Verbindung zur Regression von y auf x mit der

<u>Annahme 1</u> $\qquad \mathbb{E}(Y|X=x) = \alpha + \beta \cdot x,\ \alpha \in \mathbb{R},\ \beta \in \mathbb{R}\backslash\{0\}$

her. – Das Vorgehen bei der Regressionskurve der Zufallsvariablen X bezüglich Y erfolgt analog und die Ergebnisse sind die gleichen.

Seien $x_1, \ldots, x_n$ fest gegebene Realisationen von X, die sich auch bei Wiederholung der Stichprobenziehung nicht ändern, und $Y_{x_1}, \ldots, Y_{x_n}$ Zufallsvariablen, deren Verteilung die bedingte Verteilung von Y unter $X = x_i$

[13] Fisz 1989, S. 117, (3.7.1'), S. 118.
[14] Fisz 1989, S. 118, Definition 3.7.1.
[15] Fisz 1989, S. 118.

ist. Somit liegt das überbestimmte lineare Gleichungssystem

$$\begin{array}{rcl} \mathbb{E}(Y_{x_1}) & = & \alpha + \beta \cdot x_1 \\ & \vdots & \\ \mathbb{E}(Y_{x_n}) & = & \alpha + \beta \cdot x_n \end{array}$$

vor.

Satz 1.3.1:[16] *Seien $y_1, \ldots, y_n$ Realisationen von $Y_{x_1}, \ldots, Y_{x_n}$. Die KQ-Schätzer a und b für α und β sind erwartungstreu und Linearkombinationen von $Y_{x_1}, \ldots, Y_{x_n}$, sogenannte* ***lineare*** *Schätzer von $Y_{x_1}, \ldots, Y_{x_n}$.*

Bemerkung 1.3.2: (i) Ein Schätzer ist eine Zufallsvariable und a sowie b sind dies nicht. Allerdings sind a und b Realisationen der entsprechenden Schätzer. Da der Formalismus zur Definition der entsprechenden Schätzer hier von untergeordneter Bedeutung ist, nehmen wir die Realisation eines Schätzers als Schätzer.

(ii) Können sich die Realisationen $x_1, \ldots, x_n$ bei einer Wiederholung der Stichprobenziehung ändern, etwa wenn sie nicht fehlerfrei gemessen werden können, dann ist es schwierig, eine Aussage zur Erwartungstreue der KQ-Schätzer a und b zu treffen, da $\mathbb{E}(b)$ nicht wie im Satz 1.3.1 berechnet werden kann.[17] Bei der Konsistenz der KQ-Schätzer ist es gerade umgekehrt: Der Nenner von b ist n mal die Stichprobenvarianz. Diese konvergiert stochastisch gegen die Varianz[18], welche hier den Wert 0 hat, da $x_1, \ldots, x_n$ fest gegebene Werte sind und keine Realisationen von Stichprobenvariablen $X_1, \ldots, X_n$.

Für weitergehende Aussagen zu den KQ-Schätzern benötigen wir die folgenden Annahmen:

Annahme 2	$Y_{x_1}, \ldots, Y_{x_n}$ sind stochastisch unabhängig.
Annahme 3	Es gilt $\mathbb{V}(Y_{x_1}) = \cdots = \mathbb{V}(Y_{x_n}) =: \sigma^2$.

Bemerkung 1.3.3: (i) Ab Annahme 2 könnte die Modellierung auch über die Fehler e_i erfolgen: Die Annahmen werden für die Zufallsvariablen $E_{x_i} := Y_{x_i} - (\alpha + \beta \cdot x_i)$ formuliert. Hartung [u.a.] 2009 geht auf den Seiten 576 ff. so vor.

(ii) Annahme 3 bedeutet, dass die Streuung bei der Datenmessung immer dieselbe ist. Diese Eigenschaft wird in der Ökonometrie **Homoskedastizität** genannt.[19]

[16] Pestman 1998, S. 187, Proposition IV.1.2.
[17] Gujarati 1988 behauptet auf S. 417, dass die KQ-Schätzer nicht erwartungstreu sind.
[18] Fisz 1989, S. 430.
[19] Vgl. Fahrmeir [u.a.] 2011, S. 478; Gujarati 1988, S. 316 ff.

Satz 1.3.4:[20] *Unter den getroffenen Annahmen 1 bis 3 gilt:*

$$\mathbb{V}(b) = \frac{\sigma^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\mathbb{V}(a) = \frac{\left(\sum\limits_{i=1}^{n}(x_i - \bar{x})^2 + n \cdot (\bar{x})^2\right) \cdot \sigma^2}{n \cdot \sum\limits_{i=1}^{n}(x_i - \bar{x})^2}$$

$$Cov(a, b) = -\frac{\bar{x} \cdot \sigma^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}$$

Satz 1.3.5:[21] **von Gauß-Markov** *Mit den Annahmen 1 bis 3 haben die KQ-Schätzer a und b unter den erwartungstreuen linearen Schätzern die geringste Varianz und sind mit dieser Eigenschaft die einzigen.*

Bemerkung 1.3.6: Die KQ-Schätzer a und b heißen auch **BLUE-Schätzer** (best linear unbiased estimator).[22]

Annahme 4 Für alle x_i ist $Y_{x_i} \sim N(\alpha + \beta \cdot x_i, \sigma^2)$.

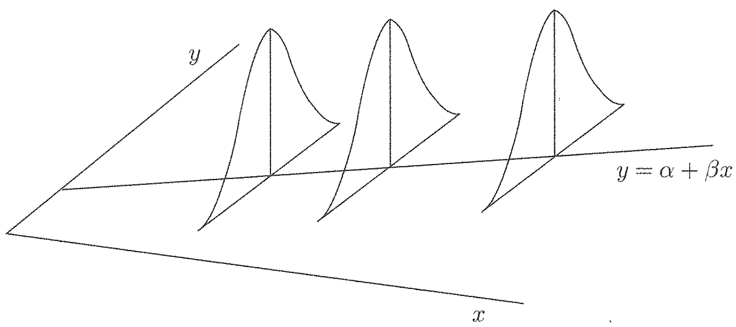
Bemerkung 1.3.7: (i) Mit Annahme 4 sind Annahme 1 und 3 gegenstandslos, können also entfallen.

(ii) Gilt Annahme 4, so wird von **normaler Regressionsanalyse** gesprochen.[23]

Satz 1.3.8:[24] *Mit den Annahmen 1 bis 4 sind die KQ-Schätzer a und b auch die ML-Schätzer für α und β.*

Die folgende Grafik[25] soll einen Eindruck vermitteln, was normale Regressionsanalyse bedeutet. Die Verteilung der Y_{x_i} unterscheiden sich im Wesentlichen nur in der Größe des Erwartungswertes. Und die Grafik zeigt, wie sehr die Realisationen y_i vom Erwartungswert $\mathbb{E}(Y_{x_i})$ abweichen können, um ihn streuen:[26]

[20] Pestman 1998, S. 187, Proposition IV.1.2.
[21] Pestman 1998, S. 189, Proposition IV.1.3. Vgl. Gujarati 1988, S. 63.
[22] Vgl. Gujarati 1988, S. 63.
[23] Vgl. Pestman 1998, S. 194.
[24] Pestman 1998, S. 198, Proposition IV.3.4.
[25] Die Grafik ist Fahrmeir [u.a.] 2011, S. 478 entnommen.
[26] Statt $y = \alpha + \beta \cdot x$ muss es $\eta = \alpha + \beta \cdot x$ heißen.



Bevor wir die Darstellung des Modells angeben können, müssen wir noch die Abweichungen, die Fehler für $i = 1, \ldots, n$ definieren:

$$E_{x_i} := Y_{x_i} - (\alpha + \beta \cdot x_i)$$

Es ist $E_{x_i} \sim N(0, \sigma^2)$ und $E_{x_1}, \ldots, E_{x_n}$ sind stochastisch unabhängig.

Darstellung des linearen normalen Regressionsmodells:

(X, Y)	zweidimensionale Zufallsvariable
$x_1, \ldots, x_n$	$\star$)
$y_1, \ldots, y_n$	$\star\star$)
$Y_{x_i} = \alpha + \beta \cdot x_i + E_{x_i}$	Regressionsgleichungen
$y = a + b \cdot x$	Regressionsfunktion
$\eta = \mathbb{E}(Y_x) = \mathbb{E}(Y \mid X = x) = \alpha + \beta \cdot x$	„wahrer“ Zusammenhang
$Y_{x_1}, \ldots, Y_{x_n}$	sind stochastisch unabhängig
$Y_{x_i} \sim N(\alpha + \beta \cdot x_i, \sigma^2)$	
$E_{x_i} \sim N(0, \sigma^2)$	

$\star$): Fest gegebene Realisationen der Zufallsvariablen X, ändern sich bei Wiederholung der Stichprobenziehung nicht.
$\star\star$): Realisationen der Zufallsvariablen $Y_{x_1}, \ldots, Y_{x_n}$

2 Fehler-in-den-Variablen-Modell

Ausgangspunkt der gewöhnlichen Regressionsanalyse sind Beobachtungswerte $x_1, \ldots, x_n$ und $y_1, \ldots, y_n$ zweier Merkmale X und Y, für die es einen funktionalen Zusammenhang der Art $f(x) = y$ geben soll. Nehmen wir den Zusammenhang $y = \alpha + \beta \cdot x$ an, dann kann das daraus resultierende Optimierungsproblem mit der Methode der kleinsten Quadrate gelöst werden. Die Methode der kleinsten Quadrate geht davon aus, dass $x_1, \ldots, x_n$ ganz bestimmte feste Werte sind, insbesondere fehlerfrei gemessen, hingegen $y_1, \ldots, y_n$ bei einer Wiederholung der Stichprobenziehung sich ändern können, mithin fehlerbehaftet sind – vergleiche Bemerkung 1.1.3 (i).

Ist X ein ökonomisches Merkmal, dann sind $x_1, \ldots, x_n$ meist genauso fehlerbehaftet wie $y_1, \ldots, y_n$. Und dies gilt nicht nur für ökonomische Merkmale, sondern für sehr viele andere — die meisten? — auch.[1] Das Modell der gewöhnlichen Regressionsanalyse muss also bezüglich der Fehlerfreiheit von $x_1, \ldots, x_n$ geändert werden, damit das Optimierungsproblem auch bei fehlerbehafteten $x_1, \ldots, x_n$ mit der Methode der kleinsten Quadrate gelöst werden kann. Das Modell, bei dem sowohl $x_1, \ldots, x_n$ also auch $y_1, \ldots, y_n$ bei einer Wiederholung der Stichprobenziehung sich ändern können, heißt **Fehler-in-den-Variablen-Modell**, kurz **FVM**.

Das Ziel dieses Kapitels ist es, die Folgen der Änderung der Modellannahme bezüglich der $x_1, \ldots, x_n$ aufzuzeigen. Wir beginnen in Abschnitt 2.1 zunächst wieder mit der Regressionsrechnung.

Idealerweise ist das Fehler-in-den-Variablen-Modell eine Verallgemeinerung des gewöhnlichen Regressionsmodells oder anders ausgedrückt das gewöhnliche Regressionsmodell ein Spezialfall des Fehler-in-den-Variablen-Modells. Es wird sich in Abschnitt 2.2 zeigen, dass eine Variante des Fehler-in-den-Variablen-Modell, das sogenannte funktionale Modell, die Verallgemeinerung des gewöhnlichen Regressionsmodells ist (Lemma 2.2.3).

2.1 Regressionsrechnung

Wir beginnen wie bei der gewöhnlichen Regressionsrechnung mit Beobachtungswerten $\xi_1, \ldots, \xi_n$ und $y_1, \ldots, y_n$ zweier metrisch skalierter Merkmale X

[1] Vgl. Miller 1996, S. 203.

und Y. Alle Beobachtungen sind fehlerbehaftet, d.h. es gibt „zufällige Fehler“[2] $\delta_1, \ldots, \delta_n, \varepsilon_1, \ldots, \varepsilon_n$ und die fehlerfreien Werte $x_1, \ldots, x_n, \eta_1, \ldots, \eta_n$, die nicht beobachtet werden können, so dass

$$\xi_1 = x_1 + \delta_1, \ldots, \xi_n = x_n + \delta_n \quad \text{sowie} \quad y_1 = \eta_1 + \varepsilon_1, \ldots, y_n = \eta_n + \varepsilon_n$$

gilt.[3]

Wieder betrachten wir den linearen Zusammenhang

$$y = \alpha + \beta \cdot \xi \quad \text{bzw.} \quad \eta = \alpha + \beta \cdot x \ .$$

Ist $n > 2$, dann liegt somit das überbestimmte lineare Gleichungssystem

$$\begin{array}{rcl} y_1 & = & \alpha + \beta \cdot \xi_1 \\ & \vdots & \\ y_n & = & \alpha + \beta \cdot \xi_n \end{array}$$

vor. In aller Regel kann dieses Gleichungssystem nicht exakt gelöst werden. Es ist somit die Regressionsgerade

$$y = a + b \cdot \xi \ ,$$

zu bestimmen, zu der die Punkte (ξ_i, y_i) einen möglichst kleinen „Abstand“ haben. Die Methode der kleinsten Quadrate minimiert ja die Summe der Quadrate der einzelnen Abweichungen

$$e_i := y_i - (a + b \cdot \xi_i) \ .$$

An dieser Stelle kommt es jetzt zur entscheidenden Änderung gegenüber der gewöhnlichen Regression. Werden a und b mit den fehlerfreien Werten x_i und η_i bestimmt, dann gilt $\eta_i = \alpha + \beta \cdot x_i = a + b \cdot x_i$, und somit ist[4]

$$e_i = \eta_i + \varepsilon_i - a - b \cdot (x_i + \delta_i) = \eta_i - (a + b \cdot x_i) + (\varepsilon_i - b \cdot \delta_i) = \varepsilon_i - b \cdot \delta_i \ .$$

Die Methode der kleinsten Quadrate hängt mit der Steigung b der Regressionsgeraden, die mit Hilfe der Methode berechnet werden soll, zusammen. Und dies hat unangenehme Folgen für die Eigenschaften der KQ-Schätzer a und b. – Zum Vergleich ist dies bei der gewöhnlichen Regression nicht der Fall:

$$e_i = \eta_i + \varepsilon_i - a - b \cdot x_i = \eta_i - (a + b \cdot x_i) + \varepsilon_i = \varepsilon_i$$

[2] Oft wird zwischen groben Fehlern, systematischen Fehlern und zufälligen Fehlern als Fehlerarten unterschieden. Ein grober Fehler liegt bei mangelnder Sorgfalt oder mangelnder Konzentration beim Messen vor. Der systematische Fehler umfasst Umwelteinflüsse und Gerätefehler. Der zufällige Fehler gilt als unvermeidlich und entsteht durch die Unvollkommenheit der Messinstrumente und der menschlichen Wahrnehmung.

[3] $x_1, \ldots, x_n$ sind weiter die fehlerfreien Werte. Dies erhöht die Lesbarkeit der nachfolgenden Notationen deutlich.

[4] Vgl. Hartung [u.a.] 2009, S. 601; Schach und Schäfer 1978, S. 153.

Formal können wir wie bei der gewöhnlichen Regression weiter vorgehen und erhalten entsprechend Satz 1.1.1 die Berechnungsformeln für a und b. Mit dem KQ-Schätzer ist das FVM natürlich dann auch asymmetrisch.
Darstellung des Modells der linearen Regressionsrechnung:

$x_1, \ldots, x_n$	fehlerfreie Werte des Merkmales X
$\eta_1, \ldots, \eta_n$	fehlerfreie Werte des Merkmales Y
$\xi_i := x_i + \delta_i,\ i = 1, \ldots, n$	beobachtet Werte des Merkmales X
$y_i := \eta_i + \varepsilon_i,\ i = 1, \ldots, n$	beobachtet Werte des Merkmales Y
$y_i = a + b \cdot \xi_i + e_i$	Regressionsgleichungen
$y = a + b \cdot \xi$	Regressionsfunktion
$\eta = \alpha + \beta \cdot x$	„wahrer“ Zusammenhang

2.2 Wahrscheinlichkeitstheoretische Modelle

Die Änderung der Modellannahme, dass x_i nicht mehr fehlerfrei gemessen werden kann, hat erhebliche Auswirkung auf die Schätzung der Parameter a und b der Regressionsgeraden. Zunächst erweist sich der KQ-Schätzer als nicht konsistent. Dann gab es bis 1969 nur ein sehr spezielles Modell — das strukturelle Modell mit zusätzlicher Information —, in dem überhaupt eine Qualitätsschätzung, nämlich eine konsistente möglich war.[5] Dies erklärt wohl auch, warum in den meisten Lehrbüchern immer noch nur das strukturelle Modell präsentiert wird. Seit Mitte der 1970er Jahre gibt es auch einen konsistenten Schätzer im funktionalen Modell mit Zusatzinformation.

Wir werden jetzt das funktionale und das strukturelle Modell einführen, um dann als erstes im Abschnitt 2.2.1 die Inkonsistenz des KQ-Schätzers für beide Modelle zu zeigen. Danach betrachten wir im Abschnitt 2.2.2 die KQ-Schätzung in beiden Modellen. Funktionales wie strukturelles Modell setzen das Grundmodell aus Abschnitt 1.2 voraus und verbinden die Regressionskurve mit der Regressionsrechnung.

Ausgehend von der zweidimensionale Zufallsvariable (X, Y) des Grundmodells sei die bedingte Verteilung von Y unter $X = x$ eine Normalverteilung, genauer

$$Y_{|X=x} \sim N(\alpha + \beta \cdot x, \sigma_\varepsilon^2)$$

mit $\alpha \in \mathbb{R},\ \beta \in \mathbb{R}\backslash\{0\},\ \sigma_\varepsilon^2 > 0$. Insbesondere gilt[6]

$$\mathbb{E}(Y_{|X=x}) = \mathbb{E}(Y|X = x) = \alpha + \beta \cdot x$$

[5] Vgl. Schach und Schäfer 1978, S. 165. Beachte auch Schönfeld 1971, S. 106. Siehe Satz 2.2.14.
[6] Pestman 1998, S. 30, Theorem I.6.2.

und[7]

$$\mathbb{V}(Y_{|X=x}) = (1 - \varrho^2_{X,Y}) \cdot \sigma^2_Y \ .$$

Hierbei ist $\varrho^2_{X,Y}$ der Korrelationskoeffizient von X und Y und σ^2_X die Varianz von X.

Seien $x_1, \ldots, x_n$ fest gegebene Realisationen von X, die sich auch bei wiederholter Stichprobenziehung nicht ändern, aber nicht beobachtet werden können und $Y_{x_1}, \ldots, Y_{x_n}$ stochastisch unabhängige Zufallsvariablen, deren Verteilung die bedingte Verteilung von Y unter $X = x_i$ ist, also

$$Y_{x_i} \sim N(\alpha + \beta \cdot x_i, \sigma^2_\varepsilon) \ .$$

Die Fehler der abhängigen Variable sind

$$E_{x_i} := Y_{x_i} - (\alpha + \beta \cdot x_i)$$

für $i = 1, \ldots, n$. Es gilt $E_{x_i} \sim N(0, \sigma^2_\varepsilon)$ mit $\sigma^2_\varepsilon = (1 - \varrho^2_{X,Y}) \cdot \sigma^2_Y$ und $E_{x_1}, \ldots, E_{x_n}$ sind stochastisch unabhängig.[8] Es fehlen noch die Fehler der unabhängigen Variable: Seien $D_{x_1}, \ldots, D_{x_n}$ stochastisch unabhängige Zufallsvariablen definiert auf den Messräumen von X mit den folgenden Eigenschaften:

1. $D_{x_1}, \ldots, D_{x_n}, E_{x_1}, \ldots, E_{x_n}$ sind stochastisch unabhängig.

2. $D_{x_i} \sim N(0, \sigma^2_\delta)$ mit $\sigma^2_\delta := (1 - \varrho^2_{X,Y}) \cdot \sigma^2_X$ für $i = 1, \ldots, n$.

Die spezielle Wahl der Varianz der Verteilung, σ^2_δ garantiert die Symmetrie des funktionalen Modells.

Es fehlen nun noch die fehlerbehafteten beobachteten Werte der unabhängigen Variable. Sei $V_{x_i} := x_i + D_{x_i}$ für $i = 1, \ldots, n$. Es gilt $V_{x_i} \sim N(x_i, \sigma^2_\delta)$.[9] Damit liegt das funktionale Modell vor:

[7] Fisz 1989, S. 192, (5.11.5). Vgl. Pestman 1998, S. 479, Aufgabe 17.
[8] Vgl. Bruchlos 2015, S. 90, Lemma 1.
[9] Pestman 1998, S. 30, Theorem I.6.1 (i).

Definition 2.2.1:[10] Das Modell

(X, Y)	zweidimensionale Zufallsvariable
$x_1, \ldots, x_n$	fehlerfreie Werte der Zufallsvariable X
$y_1, \ldots, y_n$	beobachtete Realisationen von $Y_{x_1}, \ldots, Y_{x_n}$
$V_{x_i} := x_i + D_{x_i}$	Zufallsvariable beobachteter Werte von X
$v_1, \ldots, v_n$	beobachtete Realisationen von $V_{x_1}, \ldots, V_{x_n}$
$Y_{x_i} = \alpha + \beta \cdot x_i + E_{x_i}$	Regressionsgleichungen
$y = \hat{\alpha} + \hat{\beta} \cdot v$	Regressionsfunktion
$\eta = \mathbb{E}(Y \vert X = x) = \alpha + \beta \cdot x$	„wahrer“ Zusammenhang
$Y_{x_i} \sim N(\alpha + \beta \cdot x_i, \sigma_\varepsilon^2)$	
$E_{x_i} \sim N(0, \sigma_\varepsilon^2)$	
$D_{x_i} \sim N(0, \sigma_\delta^2)$	
$D_{x_1}, \ldots, D_{x_n}, E_{x_1}, \ldots, E_{x_n}$	stochastisch unabhängig

heißt **funktionales Modell**.

Bemerkung 2.2.2: $v_1, \ldots, v_n$ entspricht im Modell der linearen Regressionsrechnung $\xi_1, \ldots, \xi_n$.

Die normale Regressionsanalyse ist ein Spezialfall des funktionalen Modells:

Lemma 2.2.3:[11] *Setzen wir im funktionalen Modell $D_{x_i} \equiv 0$, dann liegt das normale Regressionsmodell vor.*

Für das strukturelle Modell seien die Zufallsvariablen $X_1, \ldots, X_n$, definiert auf den Messräumen von X, normalverteilt, genauer[12]

$$X_i \sim N(x_i, \sigma_\delta^2/2),\ i = 1, \ldots, n\ .$$

Die stochastisch unabhängigen Zufallsvariablen $\tilde{D}_{x_1}, \ldots, \tilde{D}_{x_n}$ seien die Fehler der unabhängigen Variable, definiert auf den Messräumen von X, mit den folgenden Eigenschaften:

1. $\tilde{D}_{x_1}, \ldots, \tilde{D}_{x_n}, E_{x_1}, \ldots, E_{x_n}$ sind stochastisch unabhängig.
2. $\tilde{D}_{x_i}, X_i$ sind für $i = 1, \ldots, n$ stochastisch unabhängig.[13]

[10] Bruchlos 2015, S. 90 f., Lemma 2 (ii). Vgl. Kendall und Stuart 1979, S. 407 f.; Schach und Schäfer 1978, S. 152 f.

[11] Bruchlos 2015, S. 90 f., Lemma 2 (iii).

[12] Die spezielle Wahl der Varianz der Verteilung, $\sigma_\delta^2/2$ garantiert die Symmetrie des strukturellen Modells.

[13] Dies wird gefordert, damit W_{x_i} entsprechend normalverteilt ist. Die Forderung der stochastischen Unabhängigkeit entspricht dem nicht systematischen , dem zufälligen Fehler. Vgl. Fuller 1987, S. 3, (1.1.3).

3. $\tilde{D}_{x_i} \sim N(0, \sigma_\delta^2/2)$ für $i = 1, \ldots, n$.[14]

Sei nun noch $W_{x_i} := X_i + \tilde{D}_{x_i}$. Es gilt $W_{x_i} \sim N(x_i, \sigma_\delta^2)$.[15]

Definition 2.2.4:[16] Das Modell

$$
\begin{array}{rl}
(X, Y) & \text{zweidimensionale Zufallsvariable} \\
x_1, \ldots, x_n & \text{fehlerfreie Werte der Zufallsvariable } X \\
y_1, \ldots, y_n & \text{beobachtete Realisationen von } Y_{x_1}, \ldots, Y_{x_n} \\
W_{x_i} := X_i + \tilde{D}_{x_i} & \text{Zufallsvariable beobachteter Werte von } X \\
w_1, \ldots, w_n & \text{beobachtete Realisationen von } W_{x_1}, \ldots, W_{x_n} \\
Y_{x_i} = \alpha + \beta \cdot x_i + E_{x_i} & \text{Regressionsgleichungen} \\
y = \hat{\alpha} + \hat{\beta} \cdot w & \text{Regressionsfunktion} \\
\eta = \mathbb{E}(Y|X = x) = \alpha + \beta \cdot x & \text{„wahrer“ Zusammenhang} \\
Y_{x_i} \sim N(\alpha + \beta \cdot x_i, \sigma_\varepsilon^2) & \\
X_i \sim N(x_i, \sigma_\delta^2/2) & \\
E_{x_i} \sim N(0, \sigma_\varepsilon^2) & \\
\tilde{D}_{x_i} \sim N(0, \sigma_\delta^2/2) & \\
\tilde{D}_{x_1}, \ldots, \tilde{D}_{x_n}, E_{x_1}, \ldots, E_{x_n} & \text{stochastisch unabhängig} \\
\tilde{D}_{x_1}, \ldots, \tilde{D}_{x_n}, X_1, \ldots, X_n & \text{stochastisch unabhängig}
\end{array}
$$

heißt **ultra-strukturelles Modell**.[17] Gilt $X_i \sim N(\mu, \sigma_\delta^2/2),\ \mu \in \mathbb{R}$ für $i = 1, \ldots, n$, dann liegt das **strukturelle Modell** vor.[18]

Bemerkung 2.2.5: (i) Für die Namen der Modelle bietet sich folgende Eselsbrücke an: Strukturell wie stochastische, funktional wie fest.[19]

(ii) $w_1, \ldots, w_n$ entspricht im Modell der linearen Regressionsrechnung $\xi_1, \ldots, \xi_n$.

(iii) Die Darstellung des strukturellen Modells ändert sich gegenüber der des ultra-strukturellen an drei Stellen:

$$
\begin{array}{rl}
Y_{x_i} = \alpha + \beta \cdot \mu + E_{x_i} & \text{Regressionsgleichungen} \\
Y_{x_i} \sim N(\alpha + \beta \cdot \mu, \sigma_\varepsilon^2) & \\
X_i \sim N(\mu, \sigma_\delta^2/2) &
\end{array}
$$

[14] Die spezielle Wahl der Varianz der Verteilung, $\sigma_\delta^2/2$ garantiert die Symmetrie des strukturellen Modells.

[15] Pestman 1998, S. 34, Theorem I.6.6.

[16] Vgl. Kendall und Stuart 1979, S. 400 f., S. 403; Schach und Schäfer 1978, S. 152 f., S. 155.

[17] Dolby 1976, S. 39 f. Vgl. Kendall und Stuart 1979, S. 407.

[18] Vgl. Kendall und Stuart 1979, S. 400 f., S. 403; Schach und Schäfer 1978, S. 152 f., S. 155; Fuller 1987, S. 1 ff.,insbesondere (1.1.1), (1.1.2) und (1.1.3).

[19] Vgl. Fuller 1987, S. 2.

(iv) Die Regressionsfunktion im funktionalen und ultra-strukturellen Modell, also $y = \hat{\alpha} + \hat{\beta} \cdot v$ und $y = \hat{\alpha} + \hat{\beta} \cdot w$ unterscheiden sich von der Regressionsfunktion $y = a + b \cdot x$ im gewöhnlichen Fall in drei Punkten: Wir haben noch keine Schätzer vorliegen, d.h. die Vorgehensweise ist genau umgekehrt: Es gibt keine in der Praxis bewährten Schätzer, die auf ihre wahrscheinlichkeitstheoretischen Aussagen hin untersucht werden. Die Regressionfunktion unterscheidet sich jetzt in beiden Variablen von dem „wahren" Zusammenhang $\eta = \alpha + \beta \cdot x$ und die Regressionfunktion hat eine andere unabhängige Variable als die Regressionsgleichungen $Y_{x_i} = \alpha + \beta \cdot x_i + E_{x_i}$. Hier deutet sich die Schwierigkeit der Schätzung an.

(v) Eine übliche Darstellung des FVM, hier für das stukturelle Modell ist:[20]

$$\begin{aligned} Y_{x_i} &= \alpha + \beta \cdot x_i + E_{x_i} && W_{x_i} = X_i + \tilde{D}_{x_i} \\ X_i &\sim N(\mu, \sigma_x^2) && \\ E_{x_i} &\sim N(0, \sigma_\varepsilon^2) && \tilde{D}_{x_i} \sim N(0, \sigma_\delta^2) \\ \tilde{D}_{x_i}, E_{x_i}, X_i & && \text{stochastisch unabhängig} \end{aligned}$$

Das Grundmodell wird nicht erwähnt, die Zufallsvariablen der Fehler werden nicht hergeleitet, die Regressionsfunktion nicht angegeben. Insbesondere fallen dadurch Zusammenhänge der Varianzen der Fehler nicht auf und die unterschiedliche Variable der Regressionsfunktion und der Regressionsgleichungen. Die Darstellung des funktionalen Modells fällt noch spartanischer aus.[21]

Ein wesentlicher Vorteil der beiden Fehler-in-den-Variablen-Modelle ist, dass die Regressiongerade grundsätzlich symmetrisch geschätzt werden kann, d.h. das bei der Regression von y auf v bzw. w die entsprechende Regressionsgerade (Regressionsfunktion) vorliegt wie bei der Regression von v bzw. w auf y.[22] Dies liegt daran, dass beide Variablen nicht fehlerfrei beobachtet werden. Bei entsprechenden Schätzern kommt es also nicht mehr darauf an, ob v bzw. w die unabhängige Variable ist und y die abhängige oder umgekehrt. Welche Eigenschaft müssen die Schätzer dafür erfüllen?

Analog dem linearen Zusammenhang $\eta = \alpha + \beta \cdot x$ in der Regressionsrechnung gilt im FVM

$$\eta = \alpha + \beta \cdot x$$

im x-η-Koordinatensystem. Für die Symmetrie muss im η-x-Koordinatensystem

$$x = -\frac{\alpha}{\beta} + \frac{1}{\beta} \cdot \eta$$

[20] Fuller 1987, S. 13 oder S. 30. Vgl. Schach und Schäfer 1978, S. 152, S. 155; Kendall und Stuart 1979, S. 400, S.403.

[21] Vgl. Schach und Schäfer 1978, S. 152, S. 163; Kendall und Stuart 1979, S. 407 f.

[22] Vgl. Fuller 1987, S. 30.

sein. Sind $\tilde{a}, \tilde{b}$ die Schätzer für α, β und $\breve{a}, \breve{b}$ die Schätzer für $-\frac{\alpha}{\beta}, \frac{1}{\beta}$, dann ist das FVM bezüglich dieser Schätzer genau dann **symmetrisch**, wenn

$$\breve{b} = \frac{1}{\tilde{b}} \quad \text{und} \quad \breve{a} = -\frac{\tilde{a}}{\tilde{b}}$$

gilt.[23] Diese Eigenschaft ist bei den Schätzern zu überprüfen.

Ist $\tilde{a} := \bar{y} - \tilde{b} \cdot \bar{w}$, dann gilt wegen

$$\breve{a} = -\frac{\tilde{a}}{\tilde{b}} = -\frac{\bar{y} - \tilde{b} \cdot \bar{w}}{\tilde{b}} = \bar{w} - \frac{1}{\tilde{b}} \cdot \bar{y}$$

insbesondere das

Lemma 2.2.6: *Ist $\tilde{a} := \bar{y} - \tilde{b} \cdot \bar{w}$ bzw. $\tilde{a} := \bar{y} - \tilde{b} \cdot \bar{v}$, dann ist das FVM genau dann symmetrisch, wenn $\breve{b} = \frac{1}{\tilde{b}}$ ist.*

Im Gegensatz zum hiesigen Vorgehen werden oft das funktionale und das strukturelle Modell zunächst ohne Verteilungsannahme definiert, um zu zeigen, dass die KQ-Schätzer ohne jede Verteilungsannahme nicht konsistent sind.[24] Für alle weiteren Aussagen ist dann aber die Normalverteilungsannahme von Nöten.

Ziel der beiden nächsten Abschnitte ist die Bestimmung der Regressionsfunktion

$$y = \hat{\alpha} + \hat{\beta} \cdot v \quad \text{bzw.} \quad y = \hat{\alpha} + \hat{\beta} \cdot w \ ,$$

also die Schätzung von α und β mit Hilfe der beobachteten Werte $(v_1, y_1), \ldots, (v_n, y_n)$ im funktionalen Modell bzw. $(w_1, y_1), \ldots, (w_n, y_n)$ im (ultra-)strukturellen Modell.

2.2.1 KQ-Schätzer

Für die Schätzung von α und β bietet sich zunächst der KQ-Schätzer der gewöhnlichen Regressionsrechnung an, vergleiche Abschnitt 1.1. Mit Blick auf Bemerkung 1.3.2 wird nicht die Erwartungstreue betrachtet, sondern die Konsistenz:

Satz 2.2.7:[25] *Im strukturellen Modell konvergiert der KQ-Schätzer b für $n \to \infty$ P-fast sicher gegen*

$$\beta \cdot \frac{\sigma_\delta^2/2}{\sigma_\delta^2/2 + \sigma_\delta^2/2} = \frac{\beta}{2} \ .$$

Die KQ-Schätzer b und a sind also keine konsistenten Schätzer für β und α.

[23] Vgl. Schach und Schäfer 1978, S. 159 f., Fußnote +).

[24] Vgl. Schach und Schäfer 1978, S. 155; Kendall und Stuart 1979, S. 402 f.;Miller 1996, S. 205.

[25] Madansky 1959, S. 177; Gujarati 1988, S. 418, (13.6.10); Schach und Schäfer 1978, S. 155; Miller 1996, S. 205; Fuller 1987, S. 3, (1.1.6). Zur Aussage bezüglich des KQ-Schätzers a beachte Georgii 2004, S. 210, Satz 7.19. a und b werden mit (w_i, y_i) berechnet.

Bemerkung 2.2.8: Die Darstellung der Konvergenz in Satz 2.2.7 ist in der Literatur allgemeiner:

$$\beta \cdot \frac{\sigma_x^2}{\sigma_x^2 + \sigma_d^2}$$

Hierbei ist $\sigma_x^2 := \mathbb{V}(X_i)$ und $\sigma_d^2 := \mathbb{V}(\tilde{D}_{x_i})$. Die Darstellung in Satz 2.2.7 ist eine Folge der Konstruktion von W_{x_i}.

Korollar 2.2.9:[26] *Im funktionalen Modell sind die KQ-Schätzer b und a keine konsistenten Schätzer für β und α.*

Bemerkung 2.2.10: Die Schätzer a und b liefern natürlich weiterhin kein symmetrisches FVM - siehe den Schluss von Abschnitt 1.1.

2.2.2 ML-Schätzer

Satz 1.3.8 legt es nahe, als nächste Schätzmethode Maximum-Likelihood zu betrachten. Für die ML-Methode benötigen wir einen Verteilungstyp. Ebenfalls mit Blick auf Satz 1.3.8 bietet sich als Verteilungstyp die Normalverteilung an. Entsprechend sind das strukturelle und das funktionale Modell definiert. Als Qualitätsmerkmal versuchen wir Konsistenz zu zeigen und nicht Erwartungstreue - vergleiche Bemerkung 1.3.2.

Die folgende Aussage schränkt die Modellauswahl deutlich ein:

Satz 2.2.11:[27] *Im ultra-strukturellen Modell sind die ML-Schätzer für α und β nicht konsistent. Dies ändert sich auch nicht, wenn σ_δ^2 bekannt ist.*

Ein konsistenter Schätzer ist derzeit nur für den Spezialfall, das strukturelle Modell bekannt. Die Bedingung $X_i \sim N(\mu, \sigma_\delta),\ \mu \in \mathbb{R}$ für $i = 1, \ldots, n$ bedeutet, dass es nur einen „wahren“ Wert gibt, μ und die Messwerte um ihn herum streuen. Dies trifft nur für sehr spezielle Beispiel zu.[28] Trotzdem wird in der Regel nur das strukturelle Modell in der Literatur betrachtet.

Im funktionalen wie strukturellen Modell existiert der ML-Schätzer für α oder β im Allgemeinen nicht. Es bedarf in beiden Modellen einer weiteren Information, um den ML-Schätzer berechnen zu können, etwa der Information, dass der Quotient der Fehlervarianzen λ bzw. $\tilde{\lambda}$ bekannt ist.

2.2.2.1 Strukturelles Modell

Bei der Berechnung der ML-Schätzer für α und β tritt das Problem auf, dass fünf Gleichungen mit sechs Unbekannten eindeutig gelöst werden sollen. Deshalb gilt

[26] a und b werden mit (v_i, y_i) berechnet.
[27] Dolby 1976, S. 43.
[28] Vgl. Fuller 1987, S. 34; Madansky 1959, S. 198. Beachte Schach und Schäfer 1978, S. 156 f.

Satz 2.2.12:[29] *Im strukturellen Modell existieren die ML-Schätzer für α und β im Allgemeinen nicht.*

Setzen wir voraus, dass

$$\tilde{\lambda} := \frac{2 \cdot \sigma_\varepsilon^2}{\sigma_\delta^2}$$

bekannt ist, so gibt es zunächst eine Einschränkung (überbestimmtes Gleichungssystem):

Satz 2.2.13:[30] *Ist im strukturellen Modell $\tilde{\lambda}$ bekannt und $\alpha = 0$, so existiert der ML-Schätzer für β nicht.*

Liegt aber keine Ursprungsgerade vor, dann gilt

Satz 2.2.14:[31] *Ist im strukturellen Modell $\tilde{\lambda}$ bekannt und $\alpha \neq 0$, so sind*

$$\tilde{b} := \theta + \sqrt{\theta^2 + \tilde{\lambda}}$$

mit

$$\theta := \frac{\sum\limits_{i=1}^{n}(y_i - \bar{y})^2 - \tilde{\lambda} \cdot \sum\limits_{i=1}^{n}(w_i - \bar{w})^2}{2 \cdot \sum\limits_{i=1}^{n}(y_i - \bar{y}) \cdot (w_i - \bar{w})}$$

und

$$\tilde{a} := \bar{y} - \tilde{b} \cdot \bar{w}$$

die konsistenten ML-Schätzer für β und α, wenn der Schätzer der Kovarianz von W und Y, also die empirische Kovarianz $s_{W,Y}$ ungleich Null ist. Hierbei ist $\bar{y}$ bzw. $\bar{w}$ das arithmetische Mittel von $y_1, \ldots, y_n$ bzw. $w_1, \ldots, w_n$. Gilt $s_{W,Y} = 0$, so ist $\tilde{b} = 0$. $\tilde{a}$ und $\tilde{b}$ sind auch die Momentmethodenschätzer für α und β.[32] *Mit diesen beiden Schätzern für α und β ist das strukturelle Modell symmetrisch.*[33]

Bemerkung 2.2.15: (i) Es gibt verschiedene Darstellungen des ML-Schätzers für β aus Satz 2.2.14. Allerdings gibt es auch fehlerhafte Formeln, wobei bei Miller 1996, S. 208 ein uralter Fehler wieder auftaucht.[34]

(ii) Die Regressionsgerade, die mit den Schätzern aus Satz 2.2.14 berechnet wird und auch als **strukturelle Gerade** bezeichnet, liegt zwischen den mit KQ-Schätzern bestimmten Regressionsgeraden von w auf y und von y auf w. Der Schnittpunkt aller drei Gerade ist $(\bar{w}, \bar{y})$:[35]

[29] Miller 1996, S. 206 f.; Schach und Schäfer 1978, S. 157. Vgl. Kendall und Stuart 1979, S. 404.

[30] Creasy 1956, S. 69.

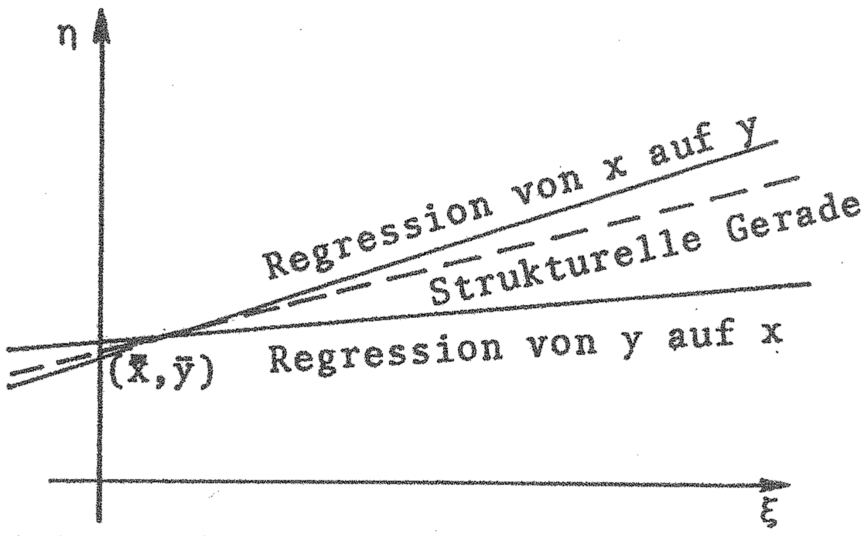
[31] Kendall und Stuart 1979, S. 405, S. 410 f.; Schach und Schäfer 1978, S. 158 f. Vgl. Hartung [u.a.] 2009, S. 603; Fuller 1987, S. 31 f. Zur Darstellung vergleiche Miller 1996, S. 208

[32] Fuller 1987, S. 31; Miller 1996, S. 209..

[33] Vgl. Schach und Schäfer 1978, S. 160, Fußnote +). Beachte Lemma 2.2.6.

[34] Vgl. Kendall und Stuart 1979, S. 405; Schach und Schäfer 1978, S. 158, Fußnote +).

[35] Schach und Schäfer 1978, S. 159ff. In der Notation von Schach, Schäfer entspricht x



Die Regressionsgeraden von w auf y und von y auf w sind obere und untere Schranke für die strukturelle Gerade. Schach und Schäfer 1978 gibt auf Seite 160 eine Abschätzung an, wie weit die Geraden aus einander liegen.

Für Ursprungsgeraden gilt

Satz 2.2.16:[36] *Ist im strukturellen Modell $\alpha = 0$, so ist der ML-Schätzer für β*

$$\tilde{b} := \frac{\bar{y}}{\bar{w}} \ .$$

Mit diesem Schätzer ist das strukturelle Modell symmetrisch.

Satz 2.2.17:[37] *Ist im strukturellen Modell σ_ε^2 bekannt und $\alpha \neq 0$, so sind*

$$\tilde{b} := \frac{\sum\limits_{i=1}^{n} y_i^2 - n \cdot \bar{y}^2 - (n-1) \cdot \sigma_\varepsilon^2}{\sum\limits_{i=1}^{n} w_i \cdot y_i - n \cdot \bar{w} \cdot \bar{y}}$$

und

$$\tilde{a} := \bar{y} - \tilde{b} \cdot \bar{w}$$

die konsistenten ML-Schätzer für α und β.

hier w und ξ hier x. Vgl. Motulsky und Christopoulos 2004, S. 50. Vgl. auch Casella und Berger 2002, S. 583 unter Beachtung von Bemerkung 2.3.4.

[36] Miller 1996, S. 211. Beachte Kendall und Stuart 1979, S. 403, (29.19)

[37] Kendall und Stuart 1979, S. 405, S. 410f; Fuller 1987, S. 14 f. Vgl. Hartung [u.a.] 2009, S. 602 f.

Satz 2.2.18:[38] *Ist im strukturellen Modell σ_δ^2 bekannt und $\alpha \neq 0$, so sind*

$$\tilde{b} := \frac{\sum\limits_{i=1}^{n} w_i \cdot y_i - n \cdot \bar{w} \cdot \bar{y}}{\sum\limits_{i=1}^{n} w_i^2 - n \cdot \bar{w}^2 - (n-1) \cdot \sigma_\delta^2/2}$$

und

$$\tilde{a} := \bar{y} - \tilde{b} \cdot \bar{w}$$

die konsistenten ML-Schätzer für α und β.

Bemerkung 2.2.19: (i) Ist im strukturellen Modell entweder σ_ε^2 oder σ_δ^2 bekannt, so stellt sich die Frage, ob es überhaupt sinnvoll ist, die Symmetrie zu betrachten, da nur von einer Variablen die Fehlervarianz bekannt ist. Fragen wir trotzdem nach der Symmetrie, dann zeigt das folgende Beispiel für Satz 2.2.17, dass die ML-Schätzer für α und β asymmetrisch sind:
Wir betrachten den „Idealgewicht-Index“[39]

$$m = 0,9 \cdot l - 100 \ ,$$

wobei das Idealgewicht m in Kilogramm angegeben wird und die Körpergröße l in Zentimetern. Ohne Messfehler bedeutet dies („die Wahrheit“):

Körpergröße x_i	195	163	178
Idealgewicht η_i	75,5	46,7	60,2

Die Schätzer sind natürlich $\tilde{a} = -100$ und $\tilde{b} = 0,9$ sowie $\check{a} = 111,\bar{1}$ und $\check{b} = 1,\bar{1}$, womit Symmetrie vorliegt . – Für die mit Fehlern gemessenen Werte

Körpergröße w_i	196	162	180
Idealgewicht y_i	76,3	46,6	59,6

wobei $\sigma_\varepsilon^2 = 0,71$ gerundet ist, ergeben sich die gerundeten Schätzer $\tilde{b} = 0,877$ und $\check{b} = 1,146$. Da $\check{b}^{-1} = 0,872$ ist, sind die Regressionsgeraden nicht symmetrisch.
Ein entsprechendes Beispiel zeigt, dass die ML-Schätzer für α und β des Satzes 2.2.18 auch asymmetrisch sind.

(ii) Sind σ_ε^2 und σ_δ^2 bekannt, so können ML-Schätzer für α und β unter bestimmten Bedingungen vorliegen. Vergleiche Kendall und Stuart 1979, S. 405 f.

[38]Kendall und Stuart 1979, S. 405, S. 410f. Vgl. Hartung [u.a.] 2009, S. 601.
[39]Dieser Index ist angelehnt an den Broca-Index unter Berücksichtigung der unteren Grenze des Bereiches des Normalgewichtes von 18,5 des Body-Mass-Indexes.

2.2.2.2 Funktionales Modell

Beim strukturellen Modell tritt bei der Berechnung der ML-Schätzer für α und β das Problem des unterbestimmten Gleichungssystems auf. Dieses Problem konnte durch zusätzliche Annahmen behoben werden. Ein ähnliches Problem tritt jetzt bei der Berechnung der ML-Schätzer im funktionalen Modell auf: Sie lassen sich ohne eine zusätzliche Annahme nicht berechnen, da die Likelihood-Funktion kein lokales Maximum, sondern einen Sattelpunkt hat.[40] In der Literatur wird nun nur die zusätzliche Annahme, dass

$$\lambda := \frac{\sigma_\varepsilon^2}{\sigma_\delta^2}$$

bekannt ist, angeführt. Ist dies die einzige Möglichkeit, die ML-Schätzer im funktionalen Modell zu berechnen?

Satz 2.2.20:[41] *Ist im funktionalen Modell λ bekannt, so sind*

$$\tilde{b} := \theta + \sqrt{\theta^2 + \lambda}$$

mit

$$\theta := \frac{\sum\limits_{i=1}^{n}(y_i - \bar{y})^2 - \lambda \cdot \sum\limits_{i=1}^{n}(v_i - \bar{v})^2}{2 \cdot \sum\limits_{i=1}^{n}(y_i - \bar{y}) \cdot (v_i - \bar{v})}$$

und

$$\tilde{a} := \bar{y} - \tilde{b} \cdot \bar{v}$$

die konsistenten ML-Schätzer für β und α, wenn der Schätzer der Kovarianz von V und Y, also die empirische Kovarianz $s_{V,Y}$ ungleich Null ist. Hierbei ist $\bar{y}$ bzw. $\bar{v}$ das arithmetische Mittel von $y_1, \ldots, y_n$ bzw. $v_1, \ldots, v_n$ ist. Mit diesen beiden Schätzern für α und β ist das funktionale Modell symmetrisch.[42]

Bemerkung 2.2.21: (i) Das funktionale Modell wird in der Literatur selten behandelt. In dem Buch von Fuller wird der Satz 2.2.20 nicht erwähnt, es wird nur festgestellt, dass im Mehrdimensionalen die ML-Schätzung nicht zum Ziel führt, S. 103 f. Selbst im Hartung wird der funktionale Fall nicht betrachtet. Bemerkenswert ist hierbei, dass im Abschnitt „Regression bei Fehlern in den Variablen“ im Gegensatz zur gewöhnlichen Regression kein wahrscheinlichkeitstheoretisches Modell angegeben wird. Damit wird auch nicht die schwere Einschränkung im strukturellen Modell, also $X_i \sim N(\mu, \sigma_\delta)$, $\mu \in \mathbb{R}$ für $i = 1, \ldots, n$ angeführt.

[40] Kendall und Stuart 1979, S. 407 f.; Dolby 1976, S. 43; Schach und Schäfer 1978, S. 165.
[41] Kendall und Stuart 1979, S. 410 f.; Schach und Schäfer 1978, S. 167 f.
[42] Vgl. Schach und Schäfer 1978, S. 160, Fußnote +). Beachte Lemma 2.2.6.

(ii) Da bei gegebener Stichprobe die Schätzer von Satz 2.2.14 und von Satz 2.2.20 identisch sind, liegt die **funktionale Gerade** natürlich auch zwischen den mit KQ-Schätzern bestimmten Regressionsgeraden von v auf y und von y auf v.

2.2.3 Bewertung der Modelle

Wie Lemma 2.2.3 zu entnehmen ist, ist das funktionale Modell die Verallgemeinerung des gewöhnlichen Regressionsmodells und schon aus diesem Grund dem strukturellen vorzuziehen. Das strukturelle Modell kann auch von der ursprünglichen Optimierungsaufgabe her keine Verallgemeinerung des gewöhnlichen Regressionsmodells sein, da der Definitionsbereich der Regressionsfunktion, also die Werte für X, nur eine Zahl, nämlich μ beinhaltet. Im Falle des gewöhnlichen Regressionsmodells sowie des funktionalen Modells besteht der Definitionsbereich aus einem reellen Intervall.

Wie sehen Anwendungen zum strukturellen Modell aus? Es darf nur eine „wahre“ Merkmalsausprägung für X geben. Madansky führt das Beispiel der Durchschlagskraft von Artilleriegranaten an,[43] Fuller das Beispiel des Bestandes an Fasanhennen in Iowa,[44] mithin sehr spezielle Anwendungen. Die Anzahl der Einsatzmöglichkeiten des strukturellen Modells ist sehr klein. Zu bevorzugen wäre das ultra-strukturelle Modell, zu dem aber keine Schätzer bekannt sind, vergleiche Satz 2.2.11.

Im funktionalen wie strukturellen Modell liegen nur dann Schätzer vor, wenn eine weitere Eigenschaft des Modells bekannt ist. Im strukturellen Modell gibt es Schätzer bei folgenden Zusatzinformationen:

1. Der Quotient der Fehlervarianzen $\tilde{\lambda}$ ist bekannt, vergleiche Satz 2.2.14.
2. Es handelt sich um eine Ursprungsgerade, d.h. es ist $\alpha = 0$, vergleiche Satz 2.2.16.
3. Die Fehlervarianz σ_ε^2 ist bekannt, vergleiche Satz 2.2.17.
4. Die Fehlervarianz σ_δ^2 ist bekannt, vergleiche Satz 2.2.18.

Im funktionalen Modell gibt es Schätzer bei folgender Zusatzinformation:

1. Der Quotient der Fehlervarianzen λ ist bekannt, vergleiche Satz 2.2.20.

Ob es sich bei einer konkreten Anwendung um eine Ursprungsgerade handelt, wird in aller Regel bekannt sein. Oft wird auch der Quotient der Fehlervarianzen λ bzw. $\tilde{\lambda}$ vorliegen[45] und zwar mit dem Wert 1. Dieser Fall

[43] Madansky 1959, S. 198.
[44] Fuller 1987, S. 34.
[45] Kendall und Stuart 1979 schreiben dazu auf Seite 405: *This is the classical method of resolving the identifiability problem.*

tritt immer dann auf, wenn beide Merkmale auf die gleiche Art erhoben, gemessen werden. Aber woher soll die Fehlervarianz σ_δ^2 oder σ_ε^2 der unabhängigen oder abhängigen Variable bekannt sein? Mir liegt dafür kein Beispiel vor. Fuller schlägt die Schätzung der Fehlervarianz — dies ist natürlich nicht die Kenntnis derselbigen — aus einer großen Zahl unabhängig wiederholter Messungen vor.[46]

Das FVM ermöglicht grundsätzlich die symmetrische Schätzung der Regressionsgeraden. Warum ist dies von Vorteil? Es kommt immer wieder vor, dass zwischen zwei Merkmalen X und Y ein linearer Zusammenhang besteht oder vermutet wird, aber keine Wirkungsweise des einen Merkmals auf das andere vorliegt, d.h. nicht von einer erklärenden (unabhängigen) und einer erklärten (abhängigen) Variable gesprochen werden kann. Bei den beiden Merkmalen „Düngemittelmenge“ und „Weizenernteertrag“ liegt eine eindeutige Wirkungsweise vor: Die Menge an Düngemittel beeinflusst die geerntete Menge an Weizen. Bei den Merkmalen „Körpergewicht“ und „Körpergröße“ des Menschen ist keine Wirkungsweise gegeben.

Liegt keine Wirkungsweise zwischen den Merkmalen X und Y vor und besteht nicht die Möglichkeit einer symmetrischen Schätzung, dann hängt das Ergebnis der Regressionsanalyse davon ab, welches Merkmal als unabhängige Variable herangezogen wird. Somit liefert die Regressionsanalyse kein befriedigendes Ergebnis.

Wie oft der Fall, dass keine Wirkungsweise gegeben ist, im Verhältnis zum Fall, dass die Wirkungsweise vorliegt, auftritt, ist schwer zu beurteilen. Es gibt diverse Beispiele, in denen keine Wirkungsweise vorliegt, womit die symmetrische Schätzung von Interesse ist und nicht vernachlässigt werden darf.

2.3 Regression nach Kummell

Die Schätzer für α und β der Sätze 2.2.14 und 2.2.20 hat Kummell 1879 bereits vorgeschlagen. [47] Allerdings erfolgt die Herleitung auf eine ganz andere Art, nämlich mit Hilfe des Taylorpolynoms.[48] Bezeichnender Weise ergibt sich auch bei dieser Vorgehensweise ein unterbestimmtes Gleichungssystem.[49]

Die Regression nach Kummell wird in einigen Naturwissenschaften als **Deming-Regression** bezeichnet,[50] obwohl dies wenig zutreffend ist. Denn Deming 1948 beschäftigt sich mit der Ausgleichsrechnung meist unter Ver-

[46] Fuller 1987, S. 13 f.

[47] Kummell 1879, S. 101, Gleichung (f) ist der Schätzer für α, Gleichung (h) und folgende Ausführungen der Schätzer für β.

[48] Kummell 1879, S. 98. Beachte Deming 1948, S. 38, Gleichung (4).

[49] Kummell 1879, S. 100: „Unless we make a certain assumption with regard to the weights of observed quantities, we cannot solve, by a direct process, even these eq'ns.“

[50] Vgl. Motulsky und Christopoulos 2004, S. 50; Linnet 1998.

wendung der Methode der kleinsten Quadrate[51] und führt die Regression nach Kummell nur nebenbei mit Verweis auf ihn an.[52]

Warum die Forschung der 1940er Jahre die Ergebnisse von Kummell nicht berücksichtigt hat, auf die ja Deming hinweist, ist mir unverständlich. Vor allem beim funktionalen Modell wären schneller Ergebnisse erzielt worden.

Die orthogonale Regression ist ein Spezialfall[53] der Regression nach Kummell:

Definition 2.3.1:[54] Sei das Modell der linearen Regressionsrechnung mit der Regressionsfunktion $y = \hat{\alpha} + \hat{\beta} \cdot \xi$ gegeben. Verwenden wir zur Bestimmung der Regressionsfunktion den quadrierten euklidischen Abstand, also[55]

$$\sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta} \cdot x_i)^2 + (\xi_i - x_i)^2 \to \min! \ ,$$

dann liegt die **Methode der orthogonalen Regression** vor.

Bemerkung 2.3.2: (i) Der euklidische Abstand eines Punktes zu einer Geraden ist der kürzester Abstand dieses Punktes zur Geraden. Die Strecke, die durch den kürzesten Abstand festgelegt ist, steht senkrecht (orthogonal) auf der Regressionsgeraden.[56]

(ii) Worin besteht der Unterschied zwischen der KQ-Methode und der Methode der orthogonalen Regression? Die KQ-Methode minmiert den Abstand nur bezüglich der y-Achse, die Methode der orthogonalen Regression bezüglich beider Achsen.

Satz 2.3.3:[57] *Die Methode der orthogonalen Regression führt zu den Schätzern*

$$\hat{b} := \theta + \sqrt{\theta^2 + 1}$$

mit

$$\theta := \frac{\sum\limits_{i=1}^{n} (y_i - \bar{y})^2 - \sum\limits_{i=1}^{n} (\xi_i - \bar{\xi})^2}{2 \cdot \sum\limits_{i=1}^{n} (y_i - \bar{y}) \cdot (\xi_i - \bar{\xi})}$$

und

$$\hat{a} := \bar{y} - \tilde{b} \cdot \bar{v}$$

[51] Deming 1948, S. iii.
[52] Deming 1948, S. 184.
[53] Beachte Bemerkung 2.3.4.
[54] Cramér 1946, S. 275. Vgl. Sachs 2004, S. 508.
[55] Madansky 1959, S. 37, (1.3.14); Casella und Berger 2002, S. 582.
[56] Casella und Berger 2002, S. 581.
[57] Fuller 1987, S. 38, (1.3.21), (1.3.27). Bei der Angabe der Bedingungen zu Gleichung (1.3.27) liegt ein Fehler vor. Es muss $\sigma_{ee} = \sigma_{uu}$ heißen. Casella und Berger 2002, S. 582. Vgl. Sachs 2004, S. 508.

Bemerkung 2.3.4: Die Schätzer der orthogonalen Regression entsprechen denen im Satz 2.2.14 (strukturelles Modell) mit $\tilde{\lambda} = 1$ und denen im Satz 2.2.20 (funktionales Modell) mit $\lambda = 1$. Diesen Spezialfall führt bereits Kummell mit Verweis auf Adcock an.[58]

Die orthogonale Regression ist der Ausgangspunkt, um auf eine dritte Weise die Schätzer der Sätze 2.2.14 und 2.2.20 (für beliebige $\tilde{\lambda}, \lambda$) herzuleiten. Glaister 2001 zeigt dies auf den Seiten 105 f.[59]

[58] Kummell 1879, S. 101 f.
[59] Beachte, dass bei Glaister 2001 $\mu = 1/\lambda$ ist.

3 Skalierungsunabhängigkeit

Betrachten wir das Merkmal X Körpergröße und das Merkmal Y Körpergewicht beim Menschen und gehen einmal davon aus, dass es zwischen diesen beiden Merkmalen den linearen Zusammenhang

$$\eta = f(x) = \alpha + \beta \cdot x$$

gibt, dann wäre es für die Schätzung der Regressionsgeraden fatal, wenn die Schätzung davon abhängt, ob Körpergröße in Zentimetern oder Metern angegeben wird beziehungsweise Gewicht in Kilogramm oder Gramm.

Wie wirkt sich eine Änderung der Skalierung überhaupt auf einen linearen Zusammenhang aus? Nehmen wir an, die Körpergröße wird nicht mehr in Metern, sondern in Zentimetern angegeben. Dies führt zu der Geraden

$$\check{f}(x) = \alpha + \check{\beta} \cdot x \ ,$$

die denselben Ordinatenabschnitt wie $f(x)$ hat, aber um das 100fache flacher verläuft. Damit der lineare Zusammenhang unverändert bleibt, muss also

$$\check{\beta} = \frac{1}{100} \cdot \beta$$

gelten. – Wird andererseits das Körpergewicht nicht in Kilogramm, sondern in Gramm angegeben, so führt dies zu

$$1000 \cdot f(x) = 1000 \cdot (\alpha + \beta \cdot x) \ .$$

Das heißt, dass die Gerade $f(x)$ nur an die Änderung der Skalierung der Ordinatenachse angepasst wird, entsprechend verschoben wird, an sich aber erhalten bleibt.

Die Änderung der Skalierung ist also auf folgende Weise zu berücksichtigen, damit der gleiche lineare Zusammenhang vorliegt: Sei für

$$f(x) = \alpha + \beta \cdot x$$

x in der Einheit Λ angegeben und $f(x)$ in der Einheit Θ. Wird nun $\check{x}$ in der Einheit $p \cdot \Lambda$ angegeben und $\check{f}(\check{x})$ in der Einheit $q \cdot \Theta$, wobei $p, q \in \mathbb{R}\backslash\{0\}$ sind, dann gilt

$$\check{f}(\check{x}) = q \cdot \alpha + \frac{q}{p} \cdot \beta \cdot \check{x} \ .$$

Ändern die Schätzer sich entsprechend, dann spielt die Skalierung einer Einheit bei der Schätzung keine Rolle:

Definition 3.0.1: Sei die Regressionsfunktion $y = f(\xi) = \hat{\alpha} + \hat{\beta} \cdot \xi$ gegeben[1], wobei ξ in der Einheit Λ und $f(\xi)$ in der Einheit Θ vorliegt.

(i) $f_p(\check{\xi}) = \hat{\alpha}_p + \hat{\beta}_p \cdot \check{\xi}$ sei die Regressionsfunktion, bei der $\check{\xi}$ in der Einheit $p \cdot \Lambda,\ p \in \mathbb{R}\backslash\{0\}$ angegeben ist. Die Regressionsfunktion $f(\xi)$ heißt **unabhängig von der Skalierung bezüglich** ξ, wenn $f_p(\xi) = \hat{\alpha} + \frac{1}{p} \cdot \hat{\beta} \cdot \check{\xi}$ gilt.

(ii) $f_q(\xi) = \hat{\alpha}_q + \hat{\beta}_q \cdot \xi$ ist die Regressionsfunktion, bei der $f_q(\xi)$ in der Einheit $q \cdot \Theta,\ q \in \mathbb{R}\backslash\{0\}$ angegeben ist. Die Regressionsfunktion $f(\xi)$ heißt **unabhängig von der Skalierung bezüglich** $f(\xi)$, wenn $q \cdot f(\xi) = f_q(\xi)$ gilt.

Bemerkung 3.0.2: (i) Die Schätzer $\hat{\alpha}, \hat{\beta}, \hat{\alpha}_p, \hat{\beta}_p, \hat{\alpha}_q, \hat{\beta}_q$ in Definition 3.0.1 werden auf dieselbe Art berechnet.

(ii) Die „Einheit“ ist die Dimension der Skala, in der das Merkmal gemessen wird. Das können physikalische Einheiten sein wie Meter oder Milliampere, aber auch ökonomische wie Arbeitslosenzahl in Tausend oder Eurocent. Das Merkmal kann in aller Regel in verschiedenen Einheiten gemessen werden, die ineinander umgerechnet werden können. Betrachten wir zum Beispiel das Merkmal „Zeitspanne“ , dann kann dieses in den Einheiten Jahr oder Sekunde gemessen werden.

(iii) Was bedeutet eine Änderung der Skalierung der Einheit? Hat ξ die Einheit $p \cdot \Lambda$, so hat $\frac{1}{p} \cdot \xi$ die Einheit Λ. Hat beispielsweise ξ die Einheit Zentimeter ($= 100 \cdot$ Meter), so hat $\frac{1}{100} \cdot \xi$ die Einheit Meter. Entsprechendes gilt für $f(\xi)$.

Lemma 3.0.3: *Die Regressionsfunktion $y = a + b \cdot x$ bzw. $y = a + b \cdot \xi$ der Regressionrechnung mit KQ-Schätzern ist unabhängig von der Skalierung bezüglich x bzw. ξ und y.*

Beweis: Sei $\check{x} := p \cdot x, \check{y} := q \cdot y$ und $\check{y} = \check{a} + \check{b} \cdot \check{x}$ mit $p, q \in \mathbb{R}\backslash\{0\}$.
Zu zeigen: $\check{y} = q \cdot a + \frac{q}{p} \cdot b \cdot \check{x}$
Es gilt

$$\check{b} = \frac{\sum_{i=1}^{n}(p \cdot x_i - p \cdot \bar{x})(q \cdot y_i - q \cdot \bar{y})}{\sum_{i=1}^{n}(p \cdot x_i - p \cdot \bar{x})^2} = \frac{p \cdot q \cdot \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{p^2 \cdot \sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{q}{p} \cdot b$$

und

[1] Diese Regressionsfunktion steht hier stellvertretend für alle in dieser Abhandlung angeführten Regressionsfunktionen.

$$\breve{a} = q \cdot \bar{y} - \breve{b} \cdot \overline{(p \cdot x)} = q \cdot \bar{y} - \frac{q}{p} \cdot b \cdot p \cdot \bar{x} = q \cdot \bar{y} - q \cdot b \cdot \bar{x} = q \cdot a \ .$$

□

Satz 3.0.4: *Die Regressionsfunktion $y = \tilde{a} + \tilde{b} \cdot w$ des strukturellen Modells bei bekannten $\tilde{\lambda}$ mit den Schätzern aus Satz 2.2.14 sowie die Regressionsfunktion $y = \tilde{a} + \tilde{b} \cdot v$ des funktionalen Modells bei bekannten λ mit den Schätzern aus Satz 2.2.20 ist unabhängig von der Skalierung bezüglich w bzw. v und y.*

Beweis: Sei $\breve{w} := p \cdot w, \breve{y} := q \cdot y$ und $\breve{y} = \breve{a} + \breve{b} \cdot \breve{w}$ mit $p, q \in \mathbb{R} \backslash \{0\}$.
Zu zeigen: $\breve{y} = q \cdot \tilde{a} + \frac{q}{p} \cdot \tilde{b} \cdot \breve{w}$
Zunächst gilt[2]

$$\breve{\lambda} = \frac{q^2}{p^2} \cdot \tilde{\lambda} \ .$$

Hieraus folgt

$$\breve{\theta} = \frac{q^2 \cdot \sum_{i=1}^{n} (y_i - \bar{y})^2 - \frac{q^2}{p^2} \cdot \tilde{\lambda} \cdot p^2 \cdot \sum_{i=1}^{n} (w_i - \bar{w})^2}{2 \cdot p \cdot q \cdot \sum_{i=1}^{n} (y_i - \bar{y}) \cdot (w_i - \bar{w})} = \frac{q}{p} \cdot \theta$$

und somit ist

$$\breve{b} = \breve{\theta} + \sqrt{\breve{\theta}^2 + \breve{\lambda}} = \frac{q}{p} \cdot \theta + \sqrt{\left(\frac{q}{p} \cdot \theta\right)^2 + \frac{q^2}{p^2} \cdot \tilde{\lambda}} = \frac{q}{p} \cdot \tilde{b} \ .$$

Nun gilt

$$\breve{a} = q \cdot \bar{y} - \breve{b} \cdot \overline{(p \cdot w)} = q \cdot \bar{y} - \frac{q}{p} \cdot \tilde{b} \cdot p \cdot \bar{w} = q \cdot \tilde{a} \ .$$

□

Bemerkung 3.0.5: Der Unterschied zwischen den Schätzern des strukturellen Modells bzw. des funktionalen Modells auf der einen Seite und denen der orthogonalen Regression auf der anderen Seite ist, dass $\tilde{\lambda}$ bzw. λ bei der orthogonalen Regression 1 sind – vergleiche Bemerkung 2.3.4. Für den Beweis von Satz 3.0.4 ist es aber essentiell, dass $\tilde{\lambda}$ bzw. λ der Quotient der Fehlervarianzen ist. Deshalb ist die orthogonalen Regression weder bezüglich ξ noch y unabhängig von der Skalierung. Darauf weist schon Wald 1940 auf Seite 284 hin.

[2]Georgii 2004, S. 107, Satz 4.23, a).

4 Fachliche Fragen

Im folgenden werden Fragen angeführt, zu denen ich die Antwort nicht kenne:

1. Gibt es im strukturellen oder funktionalen Modell neben dem Schätzer aus Satz 2.2.14 weitere Schätzer nach der Momentmethode?

2. Die Schätzer in Satz 2.2.14 (strukturelles Modell) sind auch Momentmethodenschätzer. Gilt dies auch für den korrespondierenden Satz 2.2.20 (funktionales Modell)?

3. Im funktionalen Modell liegt der ML-Schätzer bei gegebenem λ vor. Gibt es ML-Schätzer unter anderen Voraussetzungen, etwa mit der Kenntnis von σ_ε^2 oder σ_δ^2? Können andere Schätzer konstruiert werden? Welche Eigenschaften haben diese?

4. Gilt die Eigenschaft, dass die strukturelle Gerade zwischen den beiden KQ-Regressionsgeraden liegt — vergleiche Bemerkung 2.2.15 —, auch unter anderen Bedingungen? Welche Aussagen bezüglich der funktionalen Geraden sind im funktionalen Modell möglich — vergleiche Bemerkung 2.2.21 —, wenn λ nicht bekannt ist?

5. Satz 2.2.14 gibt einen Schätzer für den Fall $s_{W,Y} = 0$ an. Ist eine entsprechende Aussage im parallelen Fall des Satzes 2.2.20 für $s_{V,Y} = 0$ möglich?

6. Gibt es Möglichkeiten der Konstruktion eines Schätzers im ultra-strukturellen Modell?

7. Im funktionalen Modell — auch im (ultra-)strukturellen — haben die Regressionsgleichungen $Y_{x_i} = \alpha + \beta \cdot x_i + E_{x_i}$ eine andere unabhängige Variable als die Regressionsfunktion $y = \hat{\alpha} + \hat{\beta} \cdot v$. Kann das funktionale Modell so konstruiert werden, dass $Y_{x_i} = \alpha + \beta \cdot v_i + E_{x_i}$ gilt? Beachte Bemerkung 2.2.5, (ii).

Literaturverzeichnis

Bruchlos, Kai. „A test for the slope in the functional measurement error model.“ *Acta et Commentationes Universitatis Tartuensis de Mathematica.* 19.2 (2015): 87 - 95.

Casella, George und Roger L. Berger. *Statistical Inference.* 2. Aufl. Pacific Grove: Duxbury / Thomson Learning, 2002.

Cramér, Harald.*Mathematical methods of statistics.* 1. Aufl. Princeton: Princeton University Press, 1946.

Creasy, Monica A. „Confidence Limits for the Gradient in the Linear Functional Relationship.“ *Journal of the Royal Statistical Society, Series B.* 18 (1956): 65 - 69.

Deming, W. Edwards. *Statistical Adjustment of Data.* 4. Aufl. New York: John Wiley& Sons, 1948.

Dolby, G. R. „The ultrastructural relation: A synthesis of the functional and structural relations.“ *Biometrika.* 63.1 (1976): 39 - 50.

Fahrmeier, Ludwig, Rita Künstler, Iris Pigeot, Gerhard Tutz. *Statistik. Der Weg zur Datenanalyse.* 7. Aufl. Korr. Nachdruck. Heidelberg [u.a.]: Springer, 2011.

Fisz, Marek. *Wahrscheinlichkeitsrechnung und mathematische Statistik.* 11. Aufl. Berlin: Deutscher Verlag der Wissenschaft, 1989.

Fuller, Wayne A. *Measurement Error Models.* 1. Aufl. New York: John Wiley& Sons, 1987.

Georgii, Hans-Otto. *Stochastik.* 2. bearb. Aufl. Berlin, New York: Walter de Gruyter, 2004.

Glaister, P. „Least squares revisited.“ *The Mathematical Gazette.* 85.502 (2001): 104 - 107.

Gujarati, Damodar N. *Basic Econometrics.* 2. Aufl. New York [u.a.]: McGraw-Hill, 1988.

Hartung, Joachim, Bärbel Elpelt und Karl-Heinz Klösener. *Statistik: Lehr- und Handbuch der angewandten Statistik.* 15., überarb. u. wesentl. erw. Aufl. München: Oldenbourg, 2009.

Kendall, Maurice und Alan Stuart, *The Advanced Theory of Statistics. Volume 2: Inference and Relationship.* 4. Aufl. London: Griffin, 1979.

Kummell, Chas. H. „Reduction of Observation Equations Which Contain More Than One Observed Quantity.“ *The Analyst.* 6.4 (1879): 97 - 105.

Lexikon der Statistik: Siehe Walz 2004.

Lindley, Dennis V. „Regression lines and the linear functional relationship.“ *Journal of the Royal Statistical Society, Supplement.* 9 (1947): 218 - 244.

Linnet, Kristian. „Performance of Deming regression analysis in case of misspecified analytical error ratio in method comparison studies.“ *Clinical Chemistry.* 44.5 (1998): 1024 - 1031.

Madansky, Albert. „The Fitting of Straight Lines When Both Variables Are Subject To Error.“ *Journal of the American Statistical Association.* 54 (1959): 173 - 205.

Miller, Rupert G. *Grundlagen der Angewandten Statistik.* Übs. Boris Schlittgen. 1. Aufl. München, Wien: Oldenbourg, 1996.

Motulsky, Harvey und Arthur Christopoulos. *Fitting Models to Biological Data Using Linear and Nonlinear Regression.* 1. Aufl. Oxford: Oxford University Press, 2004.

Pestman, Wiebe R. *Mathematical Statistics.* 1. Aufl. Berlin, New York: de Gruyter, 1998.

Sachs, Lothar. *Angewandte Statistik.* 11. überarb. u. aktu. Aufl. Berlin [u.a.]: Springer, 2004.

Schach, Siegfried und Thomas Schäfer. *Regressions- und Varianzanalyse.* 1. Aufl. Berlin [u.a.]: Springer, 1978.

Schönfeld, Peter. *Methoden der Ökonometrie II.* 1. Aufl. München: Vahlen, 1971.

Schwarze, Jochen. *Grundlagen der Statistik I.* 10. Aufl. Herne, Berlin: Neue Wirtschafts-Briefe, 2005.

Wald, Abraham. „The fitting of straight lines if both variables are subject to error.“ *The Annals of Mathematical Statistics.* 11.3 (1940): 284 - 300.

Walz, Guido, Hg. *Lexikon der Statistik.* 1. Aufl. München: Elsevier, 2004.

Technische Hochschule Mittelhessen
University of Applied Sciences

Wiesenstraße 14
D-35390 Gießen

www.thm.de