

Kapitel 2

Lineares und nichtlineares Regressionsmodell und LTS-Schätzungen

Ein wichtiger Bestandteil dieser Arbeit ist die Regressionsanalyse, auf deren Basis die Optimierung des Bauteils erfolgt. Der Fokus bei diesen Untersuchungen liegt im Vergleich der Ergebnisse aus der Regressionsanalyse basierend auf linearen und nichtlinearen Regressionsmodellen. Aus diesem Grund wird in diesem Kapitel die grundlegende Theorie der linearen Regression dargestellt und auf nichtlineare Regression erweitert. Zudem werden getrimmte Schätzungen bei den linearen und nichtlinearen Modellen untersucht, da die Ausreißer-Robustheit bei den Schätzungen in dieser Arbeit eine große Rolle spielt.

2.1 Lineares Regressionsmodell

Die Grundidee der Regressionsanalyse ist, einen funktionalen Zusammenhang zwischen den einzelnen Zielgrößen¹ und den Einflussgrößen² des betrachteten Systems über eine Regressionsfunktion nachzubilden (KLEPPMANN (2008), KUEHL (2000), PETERSEN (1991)). Dieser Zusammenhang wird durch ein Modell beschrieben, das in der Regel ein Polynom zweiter oder dritter Ordnung ist (TAYLOR-Ansatz). Die Charakterisierung des Zusammenhanges erfolgt durch die Modellparameter (unbekannten Koeffizienten), durch die das Optimum bestimmt werden kann.

Wird der Ansatz des Modells als allgemeines lineares Regressionsmodell dargestellt, so folgt

$$\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} x(t_1)^T \beta \\ \vdots \\ x(t_N)^T \beta \end{pmatrix} + \begin{pmatrix} z_1 \\ \vdots \\ z_N \end{pmatrix}. \quad (2.1)$$

¹Für die Fragestellung wichtige Kenngrößen, die aus den Versuchen bzw. den Simulationen gewonnen werden; sie können Messgrößen sein, oder aus einer oder mehreren Messgrößen errechnet werden.

²Parameter, die die Zielgrößen und damit die Versuchsergebnisse beeinflussen können.

Dabei ist $y = (y_1, \dots, y_N)^T \in \mathbb{R}^N$ der Beobachtungsvektor, der die einzelnen Zielgrößen enthält. Damit sind $y_1, \dots, y_N \in \mathbb{R}$ die abhängigen Beobachtungen, wobei mit N die Anzahl der Versuche angegeben wird. Mit $X_D = (x(t_1), \dots, x(t_N))^T \in \mathbb{R}^{N \times q}$ wird die Planungsmatrix bezeichnet, die die einzelnen Einflussgrößen beinhaltet, wobei $x : \mathbb{R}^q \rightarrow \mathbb{R}^r (t \rightarrow x(t))$ eine bekannte Regressorfunktion ist. Die unabhängigen Variablen $t = (t_1, \dots, t_N) \in \mathbb{R}^q$ werden auch Versuchsbedingungen genannt, da sie meist vom Experimentator vorgegeben werden. Der unbekannte Parametervektor ist $\beta \in \mathbb{R}$ und der Fehlervektor $z = (z_1, \dots, z_N)^T \in \mathbb{R}^N$. In dem Fehlervektor sind zufällige Schwankungen und nicht berücksichtigte Einflüsse enthalten. Durch diesen Fehlerterm wird die Abweichung des Modells von der Messung beschrieben. Oft wird für z vorausgesetzt, dass der Erwartungswert $E(z) = 0$ und die Varianz $Var(z) = \sigma^2$ sind. Somit liegt ein allgemeines lineares Regressionsmodell vor, wenn es eine bekannte Regressorfunktion und einen unbekanntem Regressionsparameter gibt, so dass durch die Regressionsfunktion $f : \mathbb{R}^q \rightarrow \mathbb{R}$ mit $f(t, \beta) = x(t)^T \beta$ der Zusammenhang zwischen den Beobachtungen (Zielgrößen) und den Versuchsbedingungen (Einflussgrößen) gegeben ist. Es existiert seit Jahrzehnten eine sehr gut ausgebaute Theorie zu der linearen Regression, die in zahlreicher Literatur zu finden ist. An dieser Stelle seien beispielhaft CHATTERJEE und HADI (1988), CHATTERJEE und PRICE (1995), DANIEL und WOOD (1980), DRAPER und SMITH (1981), FAHRMEIR et al. (2004), FOX (1984), GRAYBILL (1976), GUNST und MASON (1980), JOHN (1971), MCCULLOCH und SEARLE (2008), RAO (1973) sowie SEARLE (1997) genannt.

2.1.1 Schätzungen im linearen Regressionsmodell

Im allgemeinen linearen Regressionsmodell $f(t, \beta) = x(t)^T \beta$ sind die Beobachtungen y_n bei t_n durch $y_n = x(t_n)^T \beta + z_n$ gegeben. Wird die Gleichung (2.1) in Vektorform dargestellt, so folgt

$$y = X_D \beta + z. \quad (2.2)$$

In der Regel weichen die Funktionswerte des Regressionsmodells an einigen Stellen von den tatsächlich ermittelten Werten y_n ab. Dieses kann zum einen durch die Schwankungen der Ausprägungen des Merkmals erklärt werden. Zum anderen kann dieser Effekt auch durch Mess- und Rechenungenauigkeit oder Fehler des Anwenders zu Stande kommen.

Die unbekanntem Parameter β müssen aus den Beobachtungen möglichst gut geschätzt werden. Die Regressionsschätzung für β hängt demnach von y und X_D ab und kann z.B. mit der Methode der Kleinsten-Quadrate bestimmt werden. Die damit erhaltenen Schätzungen für β heißen Kleinst-Quadrat-Summen-Schätzungen. In einschlägiger Literatur ist diese weitverbreitete Schätzmethode zu finden. Beispielsweise beschreibt DRAPER und SMITH (1981) die historische Entwicklung von der Kleinsten-Quadrat-Summen-Schätzung. In BATES und WATTS (1988) wird auf die geometrische Darstellung der Methode der Kleinsten-Quadrate eingegangen. Eine ausführliche Beschreibung dieser Schätzmethode sowie deren Eigenschaften sind z.B. in BARD (1974), CHATTERJEE und HADI (1988), CHATTERJEE und PRICE (1995), GUNST und MASON (1980), FOX (1984), RAO

(1973), SEARLE (1997) sowie SEBER und LEE (2003) aufgeführt. Demnach ist die Kleinste-Quadrat-Summen-Schätzung wie folgt definiert:

Definition 2.1 (Kleinste-Quadrat-Summen-Schätzung)

Falls

$$\hat{\beta}(y, X_D) \in \arg \min_{\beta \in \mathbb{R}^r} \sum_{n=1}^N (y_n - x(t_n)^T \beta)^2 = \arg \min_{\beta \in \mathbb{R}^r} (y - X_D \beta)^T (y - X_D \beta) \quad (2.3)$$

erfüllt, so heißt $\hat{\beta}(y, X_D)$ die *Kleinste-Quadrat-Summen-Schätzung (KQSS)* für β bei y und X_D .

Die gesuchten Parameter werden geschätzt, indem die Summe der quadratischen Abweichungen der Modellkurve von den beobachteten Punkten (z.B. Messwerten) minimiert wird. Es kann auch eine explizite Darstellung der Kleinsten-Quadrat-Summen-Schätzung angegeben werden. Dazu wird die *g-Inverse* (verallgemeinerte Inverse) einer Matrix definiert (RAO (1973), SEARLE (1997)).

Definition 2.2 (*g-Inverse*)

$A^- \in \mathbb{R}^{m \times n}$ heißt eine *g-Inverse* von $A \in \mathbb{R}^{n \times m}$, wenn

$$A A^- A = A \quad (2.4)$$

gilt.

Die *g-Inverse* von A ist nicht eindeutig, wenn A keine invertierbare Matrix ist. In diesem Fall gibt es mehrere Matrizen A^- , die die Bedingung (2.4) erfüllen.

Satz 2.1

Die Schätzung $\hat{\beta}(y, X_D)$ ist *Kleinste-Quadrat-Summen-Schätzung* für β bei y und X_D genau dann, wenn

$$\hat{\beta}(y, X_D) = (X_D^T X_D)^- X_D^T y \quad (2.5)$$

gilt.

Damit ist die Gleichung (2.5) äquivalent zur Gleichung (2.3) (DRAPER und SMITH (1981), HASTIE et al. (2001), KOK und STANDER (1999), MÜLLER (1997), REDHE und NILSSON (2006), ROUX et al. (1998), SCHOOFs (1992)).

Wenn die Planungsmatrix X_D von vollem Rang ist, dann ist $X_D^T X_D$ regulär (invertierbar) und damit deren Inverse eindeutig. In diesem Fall ist die KQSS eindeutig. Ist dagegen die Planungsmatrix X_D nicht von vollem Rang, so ist die KQSS nicht eindeutig. Es gibt dann einen affinen Unterraum von \mathbb{R}^r , so dass jedes Element aus diesem Unterraum eine KQSS ist.

Um diese Aussage zu verdeutlichen, wird nachfolgend ein Beispiel mit einer kubischen Regressionsfunktion aufgeführt.

Beispiel 1:

Der funktionale Zusammenhang einer kubischen Funktion zwischen y und t ist durch $f(t, \beta) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3$ gegeben, das ist gleichbedeutend mit $f(t, \beta) = x(t)^T \beta$, wobei $x(t) = (1, t, t^2, t^3)^T \in \mathbb{R}^4$ bekannt und $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)^T \in \mathbb{R}^4$ unbekannt ist. Der Versuchsplan D sei gegeben durch $D = (t_1, \dots, t_6)^T = (0, 0, 0, 1, 1, 1)^T$, d.h., es werden drei Beobachtungen bei $t = 0$ und drei Beobachtungen bei $t = 1$ gemacht. Damit ist die Planungsmatrix $X_D \in \mathbb{R}^{4 \times 6}$ vom Rang 2 und somit nicht von vollem Rang. In diesem Fall ist die KQSS für $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)^T$ nicht eindeutig, egal wie die Beobachtungen $y = (y_1, \dots, y_6)^T$ aussehen. Denn eine kubische Funktion kann nicht durch Punkte an zwei unterschiedlichen Stellen, hier bei $t = 0$ und $t = 1$, eindeutig bestimmt werden.

Ist man aber nur an dem konstanten Term β_0 von dem unbekanntem Parametervektor β der quadratischen Funktion interessiert, so wird dieser durch den Versuchsplan D vollständig bestimmt. Denn die Beobachtung bei 0 gibt den konstanten Term β_0 wieder.

Es kann also sein, dass bestimmte Teile (Aspekte) des Parametervektors β eindeutig bestimmbar sind, auch wenn die KQSS nicht eindeutig ist. Für weitere Untersuchungen ist der lineare Aspekt sowie die damit folgende lineare Identifizierbarkeit von Bedeutung (MÜLLER (1997)).

Definition 2.3 (Linearer Aspekt, Lineare Identifizierbarkeit)

a) Der lineare Aspekt von $\beta \in \mathbb{R}^r$ ist definiert als

$$\varphi(\beta) = L \beta,$$

wobei $L \in \mathbb{R}^{s \times r}$ vom Rang s ist.

b) Wenn für alle $\beta \in \mathbb{R}^r$

$$X_D \beta = 0 \Rightarrow L \beta = 0$$

gilt, dann heißt der lineare Aspekt $\varphi(\beta) = L \beta$ linear identifizierbar bei D .

Beispiel 2:

Es wird das obere Beispiel 1 fortgesetzt und überprüft, ob der lineare Aspekt $\varphi(\beta) = \beta$ linear identifizierbar bei D ist. Dafür müssen die zwei Bedingungen $X_D \beta = 0$ und $L \beta = 0$ erfüllt sein.

Ist $X_D \beta = 0$?

Nach dem Auflösen dieser Gleichung folgt, dass $\beta_0 = 0$ und $\beta_0 + \beta_1 + \beta_2 + \beta_3 = 0 \Rightarrow \beta_1 = -\beta_2 - \beta_3$. Also ist in diesem Fall $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)^T = (0, -\beta_2 - \beta_3, \beta_2, \beta_3)^T$.

Ist $L \beta = 0$?

Da der lineare Aspekt $\varphi(\beta) = \beta$ betrachtet wird, folgt, dass $L = I_{4 \times 4}$ ist. Damit ergibt $L \beta \neq 0$. D.h., in diesem Beispiel gibt es einen $\beta \in \mathbb{R}^r$, wobei die Bedingung $X_D \beta = 0$ erfüllt wird, aber

$L\beta \neq 0$ ist. Demnach ist der lineare Aspekt $\varphi(\beta)$ mit der Matrix $L = I_{4 \times 4}$ nicht linear identifizierbar bei D .

Wird aber der lineare Aspekt $\varphi(\beta) = \beta_0$ untersucht, so ist dieser bei D identifizierbar. Gilt $X_D\beta = 0$, so folgt $x(t_n)^T\beta = x(0)^T\beta = 0$. Dieses ergibt $(1, 0, 0, 0) \cdot (\beta_0, \beta_1, \beta_2, \beta_3)^T = 0 \Rightarrow \beta_0 = 0$. Aufgrund von $\varphi(\beta) = \beta_0$ folgt, dass $L = (1, 0, 0, 0) \in \mathbb{R}^{1 \times 4}$ ist. D.h., die zweite Bedingung $L\beta = 0$ ist somit erfüllt. Also ist der lineare Aspekt $\varphi(\beta) = \beta_0$ linear identifizierbar bei D .

Zusammenfassend kann man sagen, wenn es einen $\beta \in \mathbb{R}^r$ gibt für den

$$X_D \beta = 0 \text{ und } L \beta \neq 0,$$

gilt, dann ist der lineare Aspekt $\varphi(\beta) = L\beta$ nicht linear identifizierbar bei D .

Die Kleinste-Quadrat-Summen-Schätzung ist als Verallgemeinerung des arithmetischen Mittels sehr ausreißer-empfindlich. Im Folgenden werden alternative Regressionsschätzungen vorgestellt, die weniger ausreißer-empfindlich sind. Als erstes wird die L1-Schätzung aufgeführt (RAO und TOUTENBURG³ (1999)).

Definition 2.4 (L1-Schätzung)

Die L1-Schätzung $\hat{\beta}(y, X_D)$ basierend auf y und X_D ist definiert durch

$$\hat{\beta}(y, X_D) \in \arg \min_{\beta \in \mathbb{R}^r} \sum_{n=1}^N |y_n - x(t_n)^T \beta|.$$

Bemerkt sei, dass wenn die Betragsfunktion $|\cdot|$ durch eine Funktion $\rho(\cdot)$ ersetzt wird, führt dieses zu einer Verallgemeinerung der L1-Schätzung, nämlich der Regressions-M-Schätzung (HUBER (1964)). Die L1-Schätzung verallgemeinert die Charakterisierung des Medians, die durch den folgenden Satz gegeben ist.

Satz 2.2

Es gilt für den Median $med(y) = \tilde{y}_{0.5}$

$$med(y) \in \arg \min_{\mu \in \mathbb{R}} \sum_{n=1}^N |y_n - \mu|.$$

Interpretation: Die Funktion $g := g(\mu) = \sum_{n=1}^N |y_n - \mu|$ nimmt ihr Minimum bei $\mu = med(y) = \tilde{y}_{0.5}$ an. Es sind noch weitere Verallgemeinerungen des Medians möglich, die auf andere Regressionsschätzungen führen (siehe ROUSSEEUW und HUBERT (1999)).

Bei den bisher betrachteten Methoden KQSS und L1-Schätzung fließen alle Versuchspunkte in die Regressionsschätzung ein. Die Grundidee der LTS-Schätzung (engl.: least trimmed squares, Abk.: LTS) besteht darin, eine feste Anzahl h von Versuchspunkten aus der Schätzung herauszunehmen,

³RAO und TOUTENBURG bezeichnen L1-Schätzung als LAD-Schätzung (engl.: least absolute deviation).

also zu trimmen. Dabei ist das Ziel, dass die herauszunehmenden Versuchspunkte gerade die Ausreißer (siehe Abschnitt 1.2) sind. Somit stellt die LTS-Schätzung eine spezielle ausreißer-robuste Schätzung dar. Die Schwierigkeit bei diesem Verfahren ist die Festlegung der nicht zu betrachtenden Versuchspunkte. Die in dieser Arbeit verwendete Definition der LTS-Schätzung lehnt sich an die von ROUSSEEUW (1984) sowie ROUSSEEUW und LEROY (1987) vorgeschlagene Definition an.

Definition 2.5 (LTS-Schätzung)

$\hat{\beta}_h(y, X_D)$ ist die LTS-Schätzung bezüglich h basierend auf y und X_D und ist definiert durch

$$\hat{\beta}_h(y, X_D) \in \arg \min_{\beta \in \mathbb{R}^r} \sum_{n=1}^{N-h} r_{(n)}(y, X_D, \beta)^2.$$

Dabei ist $r_n(y, X_D, \beta) = |y_n - x(t_n)^T \beta|$ das absolute Residuum der n -ten Beobachtung mit $n = 1, \dots, N$ und mit $r_{(1)}(y, X_D, \beta) \leq r_{(2)}(y, X_D, \beta) \leq \dots \leq r_{(N)}(y, X_D, \beta)$ werden die geordneten absoluten Residuen bezeichnet.

Sowohl die L1-Schätzung als auch die LTS-Schätzung erhalten als Spezialfall die Kleinste-Quadrat-Summen-Schätzung und zwar für $|z| = z^2$ bzw. für $h = 0$.

Die LTS-Schätzung ist in dieser Arbeit von Bedeutung und wird im Abschnitt 2.2.2 auf nichtlineare Modelle übertragen und ausführlicher dargestellt. Im Abschnitt 4.2 wird auf die Berechnung der LTS-Schätzung eingegangen und deren Anwendung in den Abschnitten 4.3, 4.4 sowie 7.2.2 dargestellt.

2.1.2 Ausreißer-Robustheit im linearen Regressionsmodell

Mit dem Bruchpunkt (engl.: breakdown point) wird die Ausreißer-Robustheit von Regressionsschätzungen gemessen. Dabei ist der Bruchpunkt der kleinste Anteil von Ausreißern, der eine Schätzung beliebig verfälschen kann (DONOHO und HUBER (1983)). Der Bruchpunkt wird auf unterschiedliche Weise definiert, je nachdem ob die Regressoren (erklärende Variablen) Ausreißer enthalten können oder nicht. In diesem Fall wird der Bruchpunkt betrachtet bei dem die Versuchsbedingungen keine Ausreißer enthalten, da diese vom Experimentator vorgegeben werden. MÜLLER (1995, 1997) sowie MIZERA und MÜLLER (1999) haben den Bruchpunkt $\epsilon^*(\hat{\varphi}, y, X_D)$ wie folgt definiert.

Definition 2.6 (Regressions-Bruchpunkt)

Die Menge aller Beobachtungsvektoren \tilde{y} , die sich in höchstens M Einzelwerten vom Beobachtungsvektor y unterscheiden, sei $\mathcal{Y}_F^M(y) = \{\tilde{y} \in \mathbb{R}^N; \text{card} \{n \mid y_n \neq \tilde{y}_n\} \leq M\}$. Dann ist der Bruchpunkt $\epsilon^*(\hat{\varphi}, y, X_D)$ einer Regressionsschätzfunktion $\hat{\varphi}$ für $\varphi(\beta) = L\beta$ bei y und X_D definiert als

$$\epsilon^*(\hat{\varphi}, y, X_D) = \frac{1}{N} \min \left\{ M; \sup_{\tilde{y} \in \mathcal{Y}_F^M(y)} \|\hat{\varphi}(\tilde{y}, X_D)\| = \infty \right\}.$$

Ist der lineare Aspekt $\varphi(\beta) = L\beta$ bei D nicht identifizierbar, dann ist der Bruchpunkt $\epsilon^*(\hat{\varphi}, y, X_D)$ der Kleinsten-Quadrat-Summen-Schätzung für den linearen Aspekt gleich Null. Damit der Bruchpunkt größer Null erreicht werden kann, muss zumindest $\varphi(\beta) = L\beta$ bei D identifizierbar sein. MÜLLER (1997) zeigt, dass der Bruchpunkt tatsächlich vom Identifizierungsparameter $\mathcal{N}_\varphi(D)$ abhängt, der nachfolgend definiert wird.

Definition 2.7 (Identifizierungsparameter)

Der Identifizierungsparameter $\mathcal{N}_\varphi(D)$ ist gegeben durch

$$\mathcal{N}_\varphi(D) = \max \left\{ \sum_{n=1}^N 1_{\mathcal{D}}(t_n); \mathcal{D} \subset \{t_1, \dots, t_N\} \text{ und } \varphi(\beta) = L\beta \right. \\ \left. \text{ist nicht identifizierbar bei } \mathcal{D} \right\}.$$

Im Spezialfall ist der Identifizierungsparameter $\mathcal{N}_\beta(D)$ in linearen und verallgemeinerten linearen Modellen, wo $f(t_n, \beta) = x(t_n)^T \beta$ gilt, wie folgt definiert:

$$\mathcal{N}_\beta(D) = \max_{0 \neq \beta \in \mathbb{R}^p} \text{card} \left\{ n \in \{1, \dots, N\}; x(t_n)^T \beta = 0 \right\}.$$

Dabei gibt $\mathcal{N}_\beta(D)$ die maximale Anzahl von Versuchspunkten an, die in einem Unterraum liegen. Wenn die Beobachtungen an „beliebigen“ Stellen vorgenommen werden, ist oft $\mathcal{N}_\beta(D) = p - 1$, wobei das der minimale Wert für $\mathcal{N}_\beta(D)$ ist. In anderen Fällen ist $\mathcal{N}_\beta(D)$ größer. Diese anderen Fälle tauchen in der Regel auf, wenn die Beobachtungen nicht zufällig, sondern fest vom Experimentator vorgegeben sind.

Im Folgenden wird ein Beispiel für lineare Identifizierbarkeit und Bestimmung des Identifizierungsparameters (mit Anwendung der Definition 2.7) bei einer quadratischen Regressionsfunktion aufgeführt.

Beispiel 3:

Wie bereits erwähnt sei das allgemeine lineare Regressionsmodell durch $y_n = x(t_n)^T \beta + z_n$ gegeben. Für die quadratische Regressionsfunktion wird das Modell $y_n = \beta_0 + \beta_1 t_n + \beta_2 t_n^2 + z_n$ betrachtet. Dabei ist $x(t_n)^T = (1, t_n, t_n^2)^T$ bekannt und $\beta = (\beta_0, \beta_1, \beta_2)^T$ unbekannt. Der Versuchsplan D ist durch

t	-1	-0.5	0	0.5	1
$N(t)$	4	2	5	2	3
$x(t)$	$\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 1 \\ -0.5 \\ 0.25 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0.5 \\ 0.25 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

gegeben, wobei $N(t)$ angibt, wie oft die Versuchsbedingung t wiederholt wird. In der folgenden Abbildung ist eine mögliche Darstellung der Versuchsbedingungen und der Beobachtungen für diesen Versuchsplan aufgezeigt.

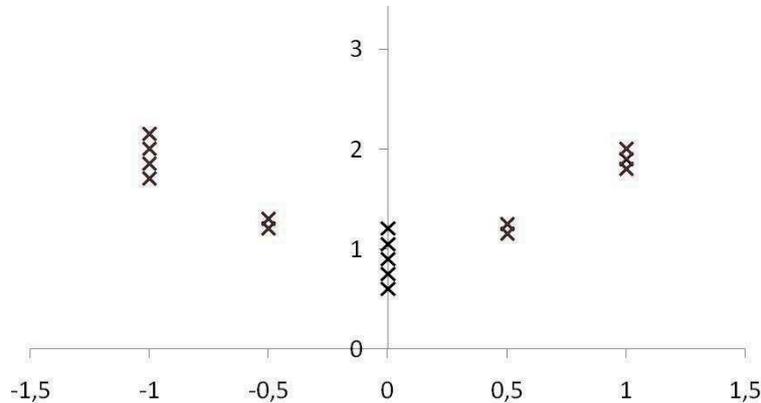


Abbildung 2.1: Mögliche Darstellung der Versuchsbedingungen

Bestimmung des Identifizierungsparameters

Fall 1

Wird der lineare Aspekt $\varphi(\beta) = \beta$ betrachtet, so folgt, dass φ nicht linear identifizierbar ist, wenn nur Beobachtungen an zwei unterschiedlichen Stellen gemacht werden. Denn zwei Beobachtungen reichen nicht aus, um eine eindeutige Lösung einer quadratischen Funktion mit drei Unbekannten zu erzielen. Zwei Punkte mit den meisten Wiederholungen sind bei $t = 0$ mit $N(t) = 5$ und $t = -1$ mit $N(t) = 4$.

Nun wird geprüft, ob die Bedingungen für die Nicht-Identifizierbarkeit bei den ausgewählten Versuchspunkten erfüllt werden. Wird der Versuchsplan $D = (0, -1)^T$ in die Gleichung $X_D \beta = 0$ eingesetzt, so ergibt sich $\beta = (0, \beta_1, \beta_1)^T$. Da hier der lineare Aspekt $\varphi(\beta) = \beta$ betrachtet wird, ist $L = I_{3 \times 3}$. Somit gilt $L \beta = (0, \beta_1, \beta_1)^T \neq 0$. Wird beispielsweise angenommen, dass $\beta_1 = 1$ ist, folgt $\beta = (0, 1, 1)^T$. Es gilt $X_D \beta = 0 \Leftrightarrow x(t_n)^T \beta = 0$ für $t_n \in D$ und damit ist $x(0)^T \beta = 0$ und $x(-1)^T \beta = 0$. Wird das ausgewählte β in die zweite Bedingung eingesetzt, folgt $L \beta = (0, 1, 1)^T \neq 0$. Somit ist der lineare Aspekt $\varphi(\beta)$ nicht linear identifizierbar, da es ein $\beta \in \mathbb{R}^r$ gibt mit $X_D \beta = 0$ und $L \beta \neq 0$.

Da die Nicht-Identifizierbarkeit bei $t = 0$ und $t = -1$ gegeben ist, folgt mit der Definition 2.7:

$$\mathcal{N}_\varphi(D) = \sum_{n=1}^N 1_{\{0\}}(t_n) + 1_{\{-1\}}(t_n) = 5 + 4 = 9.$$

Fall 2

Wird dagegen der lineare Aspekt $\varphi(\beta) = \beta_0$ betrachtet, so ist φ linear identifizierbar, wenn Beobachtungen bei Null gemacht werden. Denn durch die Beobachtung bei Null ist nur der konstante

Term einer quadratischen Funktion eindeutig bestimmbar. Bei dem linearen Aspekt $\varphi(\beta) = \beta_0$ ist $L = (1, 0, 0) \in \mathbb{R}^{1 \times 3}$. Die lineare Identifizierbarkeit ist erfüllt, da sowohl $x(0)^T \beta = 0 = x(-1)^T \beta$ als auch $L \beta = 0$ gilt.

Allerdings ist φ nicht identifizierbar bei D , wenn Beobachtungen nur an zwei von Null verschiedenen Versuchsbedingungen gemacht werden. Zwei solche Punkte mit den meisten Wiederholungen sind bei $t = -1$ mit $N(t) = 4$ und $t = 1$ mit $N(t) = 3$. Die Nicht-Identifizierbarkeit der beiden Versuchspunkte wird im Folgenden überprüft. Wird die Gleichung $X_D \beta = 0$ mit dem Versuchsplan $D = (-1, 1)^T$ aufgelöst, folgt $\beta = (-\beta_2, 0, \beta_2)^T$. Bei dem hier betrachteten linearen Aspekt $\varphi(\beta) = \beta_0$ ist $L = (1, 0, 0)^T$, also $L \beta = \beta_0 \neq 0$. Mit $\beta_2 = -1$ ist $\beta = (1, 0, -1)^T$. Für die beiden ausgewählten Versuchsbedingungen folgt $x(1)^T \beta = 0 = x(-1)^T \beta$ und $L \beta = 1 \neq 0$. Damit gibt es einen $\beta \in \mathbb{R}^r$ mit $X_D \beta = 0$ und $L \beta \neq 0$, so dass der lineare Aspekt $\varphi(\beta)$ bei D nicht linear identifizierbar ist.

Da die Nicht-Identifizierbarkeit bei $t = -1$ und $t = 1$ gegeben ist, folgt:

$$\mathcal{N}_\varphi(D) = \sum_{n=1}^N 1_{\{-1\}}(t_n) + 1_{\{1\}}(t_n) = 4 + 3 = 7.$$

MÜLLER (1995, 1997) hat den folgenden Satz aufgestellt, mit dem eine obere Schranke des Bruchpunktes angegeben wird.

Satz 2.3

Falls $\hat{\varphi}$ eine regressions-äquivalente Schätzfunktion für den linearen Aspekt $\varphi(\beta) = L \beta$ ist, gilt

$$\epsilon^*(\hat{\varphi}, y, X_D) \leq \frac{1}{N} \left\lfloor \frac{N - \mathcal{N}_\varphi(D) + 1}{2} \right\rfloor$$

für alle $y \in \mathbb{R}^N$ und alle $X_D \in \mathbb{R}^{N \times r}$.

Den Beweis zu diesem Satz sowie eine ausführliche Beschreibung zu regressions-äquivalenten Schätzern findet der Leser bei MÜLLER (1997) im Kapitel 4.

Abschließend wird der Bruchpunkt mit Anwendung vom Satz 2.3 für die beiden aufgeführten Fälle vom Beispiel 3 bestimmt.

Beispiel 4:

Der Bruchpunkt für **Fall 1** mit $\mathcal{N}_\varphi(D) = 9$ und $N=16$ ist

$$\left\lfloor \frac{N - \mathcal{N}_\varphi(D) + 1}{2} \right\rfloor = \left\lfloor \frac{16 - 9 + 1}{2} \right\rfloor = 4.$$

D.h., es reicht schon vier Beobachtungen abzuändern, damit ein Zusammenbruch erfolgen kann. In der Abbildung 2.2 a) ist beispielhaft ein möglicher Zusammenbruch der Funktion dargestellt. Die schwarze Kurve stellt eine geschätzte Funktion mit allen Beobachtungen dar. Laut der Berechnung reicht es vier Beobachtungen abzuändern, damit ein Zusammenbruch stattfinden kann. Da die Nicht-Identifizierbarkeit bei $t = -1$ und $t = 0$ gegeben ist, müssen vier Beobachtungen von den restlichen

drei Versuchsbedingungen ($t = -0.5$, $t = 0.5$ oder $t = 1$) abgeändert werden. Werden beispielsweise drei Beobachtungen bei $t = 1$ und eine Beobachtung bei $t = 0.5$ abgeändert, so entsteht eine neue Funktion (rote Kurve), die aber einen völlig anderen qualitativen Verlauf aufweist, als die geschätzte Funktion.

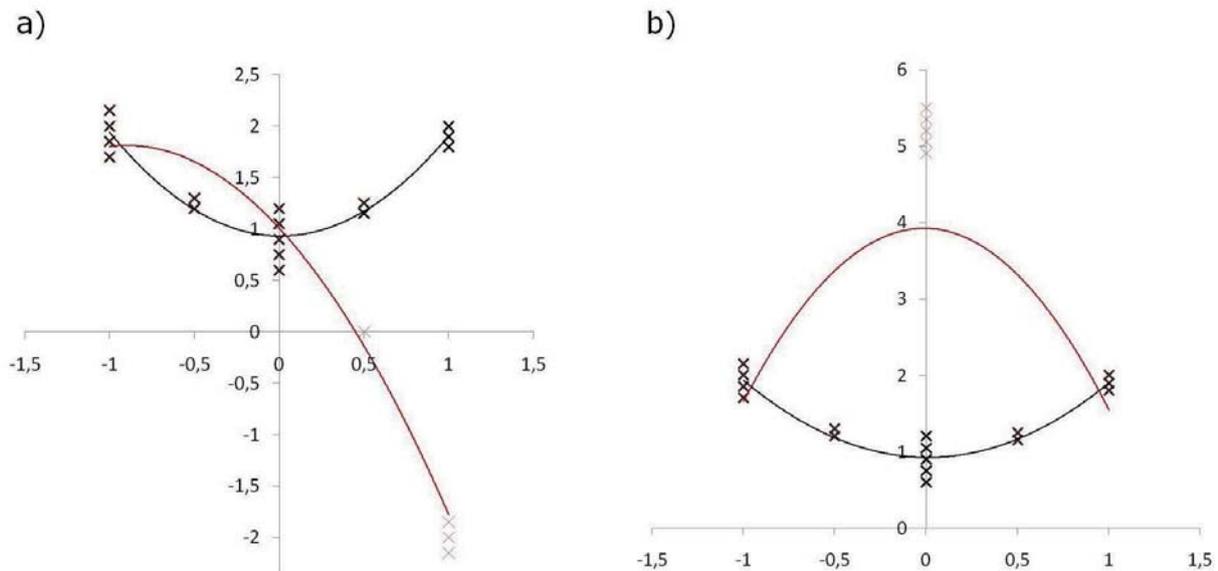


Abbildung 2.2: Zusammenbruch bei Änderungen von 4 Beobachtungen (a) bzw. 5 Beobachtungen (b)

Der Bruchpunkt für **Fall 2** mit $\mathcal{N}_\varphi(D) = 7$ und $N=16$ ist

$$\left\lfloor \frac{N - \mathcal{N}_\varphi(D) + 1}{2} \right\rfloor = \left\lfloor \frac{16 - 7 + 1}{2} \right\rfloor = 5.$$

Hier müssen fünf Beobachtungen abgeändert werden, damit ein Zusammenbruch stattfinden kann. Die Abbildung 2.2 b) zeigt einen möglichen Zusammenbruch bei Änderung von fünf Beobachtungen der geschätzten quadratischen Funktion (schwarze Kurve). Die Nicht-Identifizierbarkeit ist bei $t = -1$ und $t = 1$ gegeben, so dass fünf Beobachtungen bei den restlichen drei Versuchsbedingungen ($t = -0.5$, $t = 0$ oder $t = 0.5$) abgeändert werden müssen. Durch die Änderung von z.B. fünf Beobachtungen bei $t = 0$ entsteht eine neue quadratische Funktion (rote Kurve), die im Gegensatz zu der geschätzten Funktion nach unten geöffnet ist und somit einen ganz anderen qualitativen Verlauf darstellt.

Es existiert eine Vielzahl an Literatur zur Ausreißer-Robustheit und zum Bruchpunkt in linearen Modellen. An dieser Stelle seien stellvertretend BARNETT und LEWIS (1994), DONOHO und HUBER (1983), HAMPEL et al. (1986), HAWKINS (1980), HUBER (1981) sowie ROUSSEUW und LEROY (1987) genannt. Weitere Literaturverweise sind im Abschnitt 2.2.3 zu finden, wo die Ausreißer-Robustheit übertragen auf nichtlineare Modelle ausführlicher dargestellt wird.